

Како новинариџе да ги искористаат податоциџе за квалитетно известување



Прирачник за новинарство засновано на податоци

Приреден од Џонатан Греј, Лилијана Бунегру
и Луси Чејмберс

Како новинарите да ги искористат податоците за квалитетно известување

Прирачник за новинарство засновано на податоци

Приреден од Џонатан Греј, Лилијана Бунегру и Луси Чејмберс

КАКО НОВИНАРИТЕ ДА ГИ ИСКОРИСТАТ ПОДАТОЦИТЕ ЗА КВАЛИТЕТНО ИЗВЕСТУВАЊЕ

Прирачник за новинарство засновано на податоци

Приреден од Џонатан Греј, Лилијана Бунегру и Луси Чејмберс

Издавач

Центар за развој на медиуми
Ул. Македонија 11/1-5
1000 Скопје
www.mdc.org.mk

Превод од англиски јазик:
Лектура:

Дејан Георгиевски
Ема Маркоска – Милчин

Наслов на оригиналот:

Data Journalism Handbook, edited by
Jonathan Gray, Liliana Bounegru,
and Lucy Chambers



**ЦЕНТАР за
РАЗВОЈ на
МЕДИУМИ**



FOUNDATION ФОНДАЦИЈА
OPEN ОТВОРЕНО
SOCIETY ОПШТЕСТВО
MACEDONIA МАКЕДОНИЈА

Печатењето на оваа книга е поддржано од
Фондација Отворено општество Македонија
(Foundations Open Society Macedonia)



„Прирачникот за новинарство засновано на податоци“ може слободно да се користи, копира и редистрибуира под условите наведени во лиценцата „Кријтив комонс Наведи Извор-СподелиПодИстиУслови“. Авторските права над индивидуалните текстови опфатени во „Прирачникот за новинарство засновано на податоци“ им припаѓаат на нивните автори, кои беа љубезни да ни ги отстапат правата под условите на горенаведената лиценца.

CIP - Каталогизација во публикација

Национална и универзитетска библиотека "Св. Климент Охридски", Скопје
070.18:070.489(035)

ПРИРАЧНИК за новинарство засновано на податоци : како новинарите да ги искористат податоците за квалитетно известување / приреден од Џонатан Греј, Лилијана Бунегру и Луси Чејмберс ; [превод од англиски јазик Дејан Георгиевски]. - Скопје : Центар за развој на медиуми, 2014. - 288 стр. : илустр. ; 24 см

Превод на делото: Data journalism handbook / edited by Jonathan Gray, Liliana Bounegru, and Lucy Chambers
ISBN 978-9989-2512-4-5

а) Дата новинарство - Посебни видови на новинарство - Прирачници
COBISS.MK-ID 95469322

Содржина

НАМЕСТО ПРЕДГОВОР КОН МАКЕДОНСКОТО ИЗДАНИЕ	9
ПРЕДГОВОР	11
ЗА ОНИЕ ШТО ОСТАНУВААТ НЕИМЕНУВАНИ	11
Соработници	12
За што се работи (а за што не) во оваа книга.....	14
Издавачки конвенции користени во оваа книга.....	15
Safari® Books Online	15
Информации за контакт	16
 Глава 1	
ВОВЕД	19
ШТО Е НОВИНАРСТВО ЗАСНОВАНО НА ПОДАТОЦИ?	20
ЗОШТО НОВИНАРИТЕ ТРЕБА ДА КОРИСТАТ ПОДАТОЦИ	22
Зошто е значајно новинарството	
засновано на податоци?	25
Филтрирање на протоколот на податоците.....	25
Нови приоди кон раскажувањето приказни.....	25
Слично е на фото-новинарството, но се врши со лаптоп	25
Иднината е во дата-новинарството	26
Аритметиката се среќава со пишувањето	26
Унапредување на вештините што ги поседувате	26
Лек за информатичката асиметрија	27
Одговор на работата на одделите за	
односи со јавноста засновани на податоци.....	27
Независно толкување на официјалните информации.....	28
Справување со поплавата од податоци	28
Нашите животи се збир од податоци	29
Заштеда на време	29
Суштински елемент во комплетот новинарски алатки.....	30
Приспособување на промените во информатичката околина.....	30
Воочување на нештата што инаку не би ги виделе	31
Можност за збогатување на приказните	31
НЕКОЛУ ОМИЛЕНИ ПРИМЕРИ	32
Серијата „Не нанесувај штета“ во „Лас Вегас сан“	32
База на податоци со платите на владините службеници.....	33
Визуелизација на ирачките воени дневници, Асошијетед прес.....	34
Нерешени убиства	36
Машина за пораки.....	37
„Чартбол“ (Chartball)	38
ДАТА-НОВИНАРСТВОТО ВО ПЕРСПЕКТИВА	39
Известување со помош на компјутер и прецизно новинарство	40
Дата-новинарството и известувањето со помош на компјутер	43
Дата-новинарството е поврзано со масовната информатичка писменост	44
 Глава 2	
ВО РЕДАКЦИЈАТА	45
АКТИВНОСТИТЕ НА ЕЈ-БИ-СИ НА ПОЛЕТО НА ДАТА-НОВИНАРСТВОТО	46
Нашiot тим.....	49
Од каде ги добивме податоците?.....	49
Што научивме?	50
Големата слика: Неколку идеи	50

ДАТА-НОВИНАРСТВОТО ВО БИ-БИ-СИ	51
Дадете му лична нота.....	52
Едноставни алатки.....	52
„Копање“ по податоците.....	54
Разбирање на проблемот.....	54
Преглед на тимот.....	55
КАКО ФУНКЦИОНИРА ТИМОТ ЗА НОВИНСКИ АПЛИКАЦИИ НА „ЧИКАГО ТРИБЈУН“	56
ЗАД СЦЕНАТА ВО „ДАТАБЛОГОТ НА ГАРДИЈАН“	58
ДАТА-НОВИНАРСТВОТО ВО ОНЛАЈН-ИЗДАНИЕТО НА „ЦАЈТ“ (ZEIT ONLINE)	62
КАКО СЕ НАЈМУВА ХАКЕР	66
Неколку идеи:.....	68
КОРИСТЕЊЕ НАДВОРЕШНА ЕКСПЕРТИЗА ПРЕКУ „ХАКАТОНИ“	71
СЛЕДЕЊЕ НА ПАРИТЕ: ДАТА-НОВИНАРСТВОТО И ПРЕКУГРАНИЧНАТА СОРАБОТКА	75
НАШИТЕ СТОРИИ СЕ ВО ФОРМА НА КОМПЈУТЕРСКИ КОД	79
КАС И МУЛВАД: ПОЛУГОТОВИ СОДРЖИНИ ЗА ИНТЕРЕСНИТЕ ГРУПИ	83
Некои од нивните проекти во областа на дата-новинарството.....	85
Процеси: Иновативни информатички технологии плус анализа.....	86
Создадена вредност: Лични и компаниски брендови и приходи.....	86
Клучни сознанија од овој пример.....	87
БИЗНИС МОДЕЛИ НА ДАТА-НОВИНАРСТВОТО	87
Глава 3	
СТУДИИ НА СЛУЧАИ	91
РАЗЛИКИ ВО МОЖНОСТИТЕ (THE OPPORTUNITY GAP)	92
ДЕВЕТМЕСЕЧНА ИСТРАГА НА ЕВРОПСКИТЕ СТРУКТУРНИ ФОНДОВИ	94
1. Идентификување кој ги има податоците и како се чуваат.....	95
2. Снимање и подготовка на податоците.....	96
3. Креирање база на податоци.....	96
4. Проверка и анализа.....	96
КРИЗАТА НА „ЕВРОЗОНАТА“	97
ИЗВЕСТУВАЊЕ ЗА ЈАВНИТЕ БЏЕТИ СО OPENSPEENDING.ORG	103
ПАРЛАМЕНТАРНИТЕ ИЗБОРИ И ФИНАНСИРАЊЕТО НА КАМПАЊИТЕ ВО ФИНСКА	108
1. Пронаоѓање податоци и девелопери.....	109
2. Разгледување идеи.....	109
3. Реализација на идејата на хартија и на интернет.....	110
4. Објавување на податоците.....	110
ХАКИРАЊЕ НА ИЗБОРИ ВО РЕАЛНО ВРЕМЕ (ХЕКС/ХАКЕРС БУЕНОС АИРЕС)	111
Какви податоци користевме?.....	112
Како ја развиевме апликацијата?.....	112
Позитивни страни.....	113
Негативни страни.....	113
Импликации.....	114
ПОДАТОЦИТЕ СЕ ВЕСТ: ВИКИЛИКС	114
ХАКАТОНОТ „МАПА76“	118
ИЗВЕСТУВАЊЕТО НА „ДАТАБЛОГОТ НА ГАРДИЈАН“ ЗА НЕМИРИТЕ ВО ВЕЛИКА БРИТАНИЈА	121
Фаза 1: Немирите како што се одвиваа.....	122
Фаза два: Читање на немирите.....	123
СВИДЕТЕЛСТВА ЗА УСПЕХОТ НА УЧИЛИШТАТА ВО ИЛИНОИС	124
БОЛНИЧКИ СМЕТКИ	126
Криза во домовите за нега на стари лица.....	129
ТЕЛЕФОНОТ НЕ ЧУВА ТАЈНИ	130
КОЈ МОДЕЛ НА АВТОМОБИЛ? СТАПКАТА НА ПАЃАЊЕ НА ТЕХНИЧКИ ПРЕГЛЕД	132
СУБВЕНЦИИ ЗА ЈАВНИОТ АВТОБУСКИ ПРЕВОЗ ВО АРГЕНТИНА	134
Кој го спроведуваше проектот?.....	138
Кои алатки ги користевме?.....	138
ГРАЃАНСКИ ДАТА-НОВИНАРИ	140
ГОЛЕМАТА ТАБЛА СО ИЗБОРНИ РЕЗУЛТАТИ	143
„КРАУДСОРСИНГ“ НА ЦЕНИТЕ НА ВОДАТА	145

Глава 4

СОБИРАЊЕ НА ПОДАТОЦИТЕ	149
КРАТОК ПРЕГЛЕД НА ТЕМАТА НА РАБОТА	150
Стеснете го пребарувањето	150
Прегледување на онлајн-сервисите за чување и давање податоци	152
Поставете прашање на некој форум	154
Прашајте преку „мејлинг-листа“	154
Придружете и се на групата „Хекс/хакерс“	154
Прашајте експерт	154
Дознајте повеќе за информатичките технологии што ги користи владата	155
Повторете го пребарувањето	155
Напишете барање за слободен пристап до информации	156
ВАШЕТО ПРАВО НА ПРИСТАП ДО ПОДАТОЦИТЕ	157
ПРИСТАПОТ ДО ИНФОРМАЦИИ ФУНКЦИОНИРА. КОРИСТЕТЕ ГО!	163
Студија на случај 1: Субвенции за земјоделците	163
Студија на случај 2: Несакани дејства	165
Студија на случај 3: Шверцување на смртта	166
СОБИРАЊЕ ПОДАТОЦИ ОД ИНТЕРНЕТ	168
Што се машински читливи податоци?	168
„Стругање“ на веб-страници: со каква цел?	169
Што може, а што не може да се „струга“	169
Алатки што ќе ви помогнат во „стругањето“	170
Како функционираат „веб-стругачите“?	171
Анатомијата на веб-страниците	171
Пример: „Стругање“ на податоците за нуклеарните инциденти со помош на програмскиот јазик „Python“	172
ИНТЕРНЕТ КАКО ИЗВОР НА ПОДАТОЦИ	177
Веб-алатки	177
Веб-страници, илустрации и видео	180
Електронска пошта	182
Трендови	183
„КРАУДСОРСИНГ“ НА ПОДАТОЦИ ЗА „ДАТАБЛОГОТ НА ГАРДИЈАН“	184
КАКО „ДАТАБЛОГОТ“ ГО КОРИСТЕШЕ „КРАУДСОРСИНГОТ“	
ВО ИЗВЕСТУВАЊЕТО ЗА ПРОДАЖБАТА НА БИЛЕТИ ЗА ОЛИМПИЈАДАТА	187
КОРИСТЕЊЕ И РАЗМЕНА НА ПОДАТОЦИ:	
ШТО ВЕЛАТ ПРОПИСИТЕ, ШТО Е НАПИШАНО СО МАЛИ БУКВИ И КАКВА Е РЕАЛНОСТА	190

Глава 5

РАЗБИРАЊЕ НА ПОДАТОЦИТЕ	195
ПОДОБРЕТЕ ГИ ПОЗНАВАЊАТА НА ПОЛЕТО НА РАБОТА СО ПОДАТОЦИ ВО ТРИ ЛЕСНИ ЧЕКОРИ	196
1. Како се собрани податоците?	197
2. Што ново може да се научи?	199
3. Колку веродостојни се информациите?	200
СОВЕТИ ЗА КОРИСТЕЊЕ БРОЈКИ ВО ВЕСТИТЕ	201
Основни чекори во работата со податоци	202
Знајте кои прашања сакате да ги одговорите	203
Чистење на збркани податоци	204
Податоците можат да имаат својства што не се документирани	205
ВЕКНА ЛЕБ ОД 32 ФУНТИ	207
ОД ПОДАТОЦИ ДО СТОРИЈА	209
ПРИКАЗНИТЕ ВО ПОДАТОЦИТЕ	210
НОВИНАРИТЕ И НИВНИТЕ ОМИЛЕНИ АЛАТКИ	212
КОРИСТЕЊЕ ВИЗУЕЛИЗАЦИИ ЗА ПОДОБАР УВИД ВО ПОДАТОЦИТЕ	218
Користење на визуелизацијата за нови сознанија	219
Кои алатки да ги користиме	224
Пример: Разбирање на податоците за донациите на изборните кампањи во САД	224
Што можеме да научиме од овој пример?	231
Преземете го изворниот код	231

ИСПОРАКА НА ПОДАТОЦИТЕ	233
ПРЕСТАВУВАЊЕ НА ПОДАТОЦИТЕ ВО ЈАВНОСТА	234
Да се визуелизира или не?	234
Користење анимирани графички решенија	234
Нека слушне цел свет	235
Објавување на податоците	236
Давање пристап до податоците	236
Основање платформа за отворени податоци	236
Давање човечки лик на податоците	237
Отворени податоци, отворен изворен код, отворени вести	238
Додадете линк за снимање на податоците	238
Познавајте ги своите можности	239
КАКО ДА ИЗРАБОТИТЕ НОВИНСКА АПЛИКАЦИЈА	239
Која е мајката публика и какви се нејзините потреби?	240
Колку време да му посветам на проектот?	241
Како да достигнам ново, повисоко ниво?	242
Пакување	242
НОВИНСКИТЕ АПЛИКАЦИИ НА „ПРОПУБЛИКА“	243
ВИЗУЕЛИЗАЦИЈАТА КАКО МОТОР ЗА ДАТА-НОВИНАРСТВОТО	245
Совет 1: Користете мали пресметки за брза ориентација во некоја голема група податоци	246
Совет 2: Погледнете ги податоците свртени наопаку и од страна	248
Совет 3: Не претпоставувајте ништо	249
Совет 4: Не бидете опседнати со прецизноста	249
Совет 5: Подгответе хронологиите на случаите и настаните	250
Совет 6: Составувајте се со луѓето од графичкото одделение што почесто	250
Совети за објавувањето	250
КОРИСТЕЊЕ ВИЗУЕЛИЗАЦИИ ВО РАСКАЖУВАЊЕТО ПРИКАЗНИ	251
Нов поглед на познатото	252
Прикажување на промените што настанале со текот на времето	253
Споредување вредности	255
Прикажување на меѓусебните врски и тековите на податоците	256
Дизајнирање со помош на податоци	257
Прикажување на хиерархиските структури	258
Прегледување големи бази на податоци	259
Замислување на поинакви резултати	261
Кога не треба да се користи визуелизација	262
РАЗЛИЧНИ ГРАФИКОНИ РАСКАЖУВААТ РАЗЛИЧНИ ПРИКАЗНИ	263
„НАПРАВИ САМ“ ВИЗУЕЛИЗАЦИЈА:	
НАЈДОБРИТЕ АЛАТКИ СПОРЕД НОВИНАРИТЕ НА „ГАРДИЈАН“	269
Табелите на „Гугл Фјужн“ (Google Fusion Tables)	269
„Табло паблик“ (Tableau Public)	271
Дијаграмите на „Гугл спредшит“ (Google Spreadsheet Charts)	273
„Датамаркет“ (Datamarket)	274
„Мени ајс“ (Many Eyes)	274
„Готвач за бои“ (Color Brewer)	275
Уште неколку алатки	275
КАКО „ВЕРДЕНС ГАНГ“ ГИ ПРИКАЖУВА ПОДАТОЦИТЕ	275
Броеви	276
Мрежи	276
Мапи	277
Анализа на текст	278
Заклучоци	278
ЈАВНИТЕ ПОДАТОЦИ СТАНУВААТ ОПШТЕСТВЕНО ПРАШАЊЕ	279
КАКО ДА ГИ ЗАИНТЕРЕСИРАТЕ ЛУЃЕТО ЗА ВАШИТЕ ПОДАТОЦИ	282
БЕЛЕШКА ЗА УРЕДНИЦИТЕ	285
БЕЛЕШКА ЗА КООРДИНАТОРИТЕ НА ПРОЕКТОТ	285

Наместо предговор кон македонското издание

Почитувани читатели,

Ќе ни дозволите најнапред да ѝ се заблагодариме на Фондацијата отворено општество Македонија (ФООМ), за помошта што ни ја даде во подготовката на овој Прирачник за дата-новинарство. Без финансиската поддршка на ФООМ, ова издание, сигурно, немаше да види бел ден.

Ние, во Центарот за развој на медиуми, подолго време баравме начин како овој Прирачник, подготвен од „Европскиот центар за новинарство“ (European Journalism Centre) и Фондацијата „Отворено знаење“ (Open Knowledge Foundation), да го издадеме и на македонски јазик.

Дата-новинарството е значаен дел од нашите активности на полето на подобрување на професионалните стандарди на новите медиуми. Сметаме дека се побрзиот развој на медиумите, пред сè, на оние на Интернет, бара од новинарите, што побрзо и поинтензивно да почнат да го практикуваат новинарството засновано на податоци.

Оваа книга е наменета, пред сè, за новинарите од онлајн-медиумите (без оглед дали објавуваат само на Интернет или пак, нивните веб-страници се само продолжение на печатените и радиодиузните изданија) кои сакаат на својата публика да ѝ ги понудат сториите што се кријат зад суровите податоци, но и податоците што послужиле како основа за нивните стории.

Неколку зборови и за терминологијата, односно за проблемот со кој се соочивме во подготовката на ова издание - како да го преведеме англискиот поим „data-journalism“. Можностите што ни стојеа на располагање – „статистичко новинарство“, „аналитичко новинарство“, „податочно новинарство“ – или се многу пошироки, или се многу потесни од темата, или пак, звучат чудно и неприродно. Најсоодветниот превод би бил опи-сен, и би гласел „новинарство засновано на податоци“, решение што го искористивме и во насловот на македонското издание на прирачниците. Во потрагата по поконцизен и пократок израз, решивме да погледнеме со која терминологија се служи новинарската професија. Бидејќи се работи за релативно нова новинарска дисциплина, бројот на референци не е многу голем. Сепак, дури и површното пребарување на Интернет покажува дека македонските новинари, а и нивните колеги кои пишуваат и известуваат на некој од јужнословенските јазици, претежно го користат терминот „дата-новинарство“.

Имајќи предвид дека овој прирачник е наменет за новинарите кои се занимаваат, или сакаат да се занимаваат со „новинарство врз основа на податоци“, решивме да се приклониме кон нивниот избор и во овој прирачник да користиме „дата-новинарство“ за означување на она што на англиски јазик се нарекува „data journalism“.

Македонското издание на Прирачникот за новинарство засновано на податоци му го посветуваме на Роберто Беличанец, основачот на Центарот за развој на медиуми и неуморниот борец за слободно и професионално новинарство, кој нè напушти прерано. Затоа, се надеваме дека прирачниците ќе успее макар малку да придонесе кон натамошниот слободен и професионален развој на медиумите во Македонија, што беше негова животна и професионална мисија.

Со почит,

Центар за развој на медиуми

Предговор

За оние што остануваат неименувани

Идејата за подготовка на Прирачникот за новинарство засновано на податоци се роди на една 48-часовна работилница организирана од Европскиот центар за новинарство и Фондацијата „Отворено знаење“ во рамките на фестивалот „МозФест2011“ (MozFest2011) во Лондон. Веднаш по фестивалот, идејата се преточи во меѓународен колаборативен потфат во кој се вклучија десетици водечки застапници на новинарството засновано на сурови податоци и најдобрите практичари на полето на дата-новинарството.



Во шестте месеци што поминаа од идејата за настанување на оваа книга до првото целосно издание, на еден или на друг начин, свој придонес кон нејзината содржина дадоа стотици луѓе. И покрај нашите обиди да ги бележиме сите дополнувања и индивидуални придонеси, сепак, бројот на оние што своите дополнувања, предлози и сугестии ги направиле анонимно или со користење псевдоними останува прилично голем, а има и такви за кои не можеме со сигурност да го посочиме точниот извор.

Имаме две нешта да им порачаеме на сите оние што дадоа свој придонес, а не се спомнати поименично на листата на соработници. Прво, сакаме да ви се заблагодариме за помошта. Второ, ве молиме да ни кажете кои сте и од каде сте, за да можеме да ги спомнеме и вашите заслуги во создавањето на прирачникот.

Соработници

Следните луѓе работеа на нацртот или на друг начин дадоа директен придонес кон текстот на сегашната верзија на книгата (илустрациите се на графичката дизајнерка Кејт Хадсон):

- Грегор Ајш, Фондација „Отворено знаење“ (Gregor Aisch, Open Knowledge Foundation)
- Бригите Алфтер (Brigitte Alfter), Journalismfund.eu
- Дејвид Андертон, слободен новинар (David Anderton, Freelance Journalist)
- Џејмс Бол, „Гардијан“ (James Ball, the Guardian)
- Каелин Бар, „Ситивајр“ (Caelainn Barr, Citywire)
- Маријана Беруецо, „Хекс/хакерс Буенос Аирес“ (Mariana Berruezo, Hacks/Hackers Buenos Aires)
- Мајкл Бластленд, слободен новинар (Michael Blastland, Freelance Journalist)
- Маријано Блехман, „Хекс/хакерс Буенос Аирес“ (Mariano Blejman, Hacks/Hackers Buenos Aires)
- Џон Боунс, „Верденс генг“ (John Bones, Verdens Gang)
- Маријане Бушар, „Блумберг њус“ (Marianne Bouchart, Bloomberg News)
- Лилијана Бунегру, „Европски центар за новинарство“ (Liliana Bounegru, European Journalism Centre)
- Брајан Бојер, „Чикаго трибјун“ (Brian Boyer, Chicago Tribune)
- Пол Бредшо, Градски универзитет на Бирмингем (Paul Bradshaw, Birmingham City University)
- Венди Карлајл, Еј-Би-Си – Австралиска радиодифузна корпорација (Wendy Carlisle, Australian Broadcasting Corporation)
- Луси Чејмберс, Фондација „Отворено знаење“ (Lucy Chambers, Open Knowledge Foundation)
- Сара Коен, Универзитет „Дјук“ (Sarah Cohen, Duke University)
- Алистер Дент, „Гардијан“ (Alastair Dant, the Guardian)
- Хелен Дарбишер (Helen Darbishire), Access Info Europe
- Чејс Дејвис, Центар за истражувачко новинарство (Chase Davis, Center for Investigative Reporting)
- Стив Дојг, Школа за новинарство „Волтер Кронкајт“, Државен универзитет на Аризона (Steve Doig, Walter Cronkite School of Journalism, Arizona State University)
- Лиза Еванс, „Гардијан“ (Lisa Evans, the Guardian)
- Том Фрис, Фондација „Бертелсман“ (Tom Fries, Bertelsmann Stiftung)
- Данкан Гир (Duncan Geere), Wired UK
- Џек Гилам, „Асошијетед прес“ (Jack Gillum, Associated Press)
- Џонатан Греј, Фондација „Отворено знаење“ (Jonathan Gray, Open Knowledge Foundation)
- Алекс Хауард, „О’Рајли медија“ (Alex Howard, O’Reilly Media)
- Бела Харел, Би-Би-Си (Bella Hurrell, BBC)
- Николас Кајзер-Брил (Nicolas Kayser-Bril), „Журнализам++“ (Journalism++)
- Џон Киф (John Keefe), „WNYC“ Радио
- Скот Клајн, „ПроПублика“ (Scott Klein, ProPublica)
- Александар Лешне, „Ле Монд“ (Alexandre Léchenet, Le Monde)

- Марк Ли Хантер, „ИНСЕАД“ (Mark Lee Hunter, INSEAD)
- Ендрју Лајмдорфер, Би-Би-Си (Andrew Leimdorfer, BBC)
- Фридрих Линденберг, Фондација „Отворено знаење“ (Friedrich Lindenberg, Open Knowledge Foundation)
- Мајк Линксвејер (Mike Linksvayer), „Криејтив комонс“ (Creative Commons)
- Мирко Лоренц, „Дојче веле“ (Mirko Lorenz, Deutsche Welle)
- Еса Макинен, „Хелсингин саномат“ (Esa Mäkinen, Helsingin Sanomat)
- Педро Маркун, „Транспаренција хакер“ (Pedro Markun, Transparència Hacker)
- Исао Мацунами, „Токио шимбун“ (Isao Matsunami, Tokyo Shimbun)
- Лоренц Мацат, „Опендасити“ (Lorenz Matzat, OpenDataCity)
- Геоф МекГи, Универзитет „Стенфорд“ (Geoff McGhee, Stanford University)
- Филип Мејер, почесен професор, Универзитет на Северна Каролина во Чепел Хил (Philip Meyer, Professor Emeritus, University of North Carolina at Chapel Hill)
- Клер Милер, „Велс онлајн“ (Claire Miller, WalesOnline)
- Синтија О’Мурку, „Фајненшал тајмс“ (Cynthia O’Murchu, Financial Times)
- Олусеун Онигбинде (Oluseun Onigbinde), „БаџАјТи“ (BudgIT)
- Ѓорѓе Падејски, Програма за новинарство „Најт“ на Универзитетот „Стенфорд“ (Djordje Padejski, Knight Journalism Fellow, Stanford University)
- Џејн Парк (Jane Park), „Криејтив комонс“ (Creative Commons)
- Анхелика Пералта Рамос, „Ла Насион“, Аргентина (Angélica Peralta Ramos, La Nacion, Argentina)
- Черил Филипс, „Сиетл тајмс“ (Cheryl Phillips, The Seattle Times)
- Арон Пилхофер, „Њујорк тајмс“ (Aron Pilhofer, New York Times)
- Лулу Пини, дизајнерка на инфографички решенија (Lulu Pinney, Freelance Infographic Designer)
- Паул Раду, Проект за известување за организираниот криминал и корупцијата (Paul Radu, Organised Crime and Corruption Reporting Project)
- Сајмон Роџерс, „Гардијан“ (Simon Rogers, the Guardian)
- Мартин Розенбаум, Би-Би-Си (Martin Rosenbaum, BBC)
- Аманда Роси, „Пријателите на Јануарија“ (Amanda Rossi, Friends of Januária)
- Мартин Сарсале, „Хекс/хакерс Буенос Аирес“ (Martin Sarsale, Hacks/Hackers Buenos Aires)
- Фабрицио Скролини, Лондонска школа за економија и политички науки (Fabrizio Scrollini, London School of Economics and Political Science)
- Сара Слобин, „Волстрит журнал“ (Sarah Slobin, Wall Street Journal)
- Серхио Сорин, „Хекс/хакерс Буенос Аирес“ (Sergio Sorin, Hacks/Hackers Buenos Aires)
- Џонатан Стреј, Проект „Овервју“ (Jonathan Stray, The Overview Project)
- Брајан Суда (Brian Suda), optional.is
- Крис Тагарт, „ОпенКорпорејтс“ (Chris Taggart, OpenCorporates)
- Џер Торп, Група за развој и истражување на „Њујорк тајмс“ (Jer Thorp, The New York Times R&D Group)
- Енди Тоу, „Хекс/хакерс Буенос Аирес“ (Andy Tow, Hacks/Hackers Buenos Aires)
- Лук Н. ван Васенхоф, „ИНСЕАД“ (Luk N. Van Wassenhove, INSEAD)
- Саша Фенор, Онлајн издание на „Цајт“ (Sascha Venohr, Zeit Online)
- Џери Ферманен (Jerry Vermanen), NU.nl
- Сезар Вијана, Универзитетот на Гојас (César Viana, University of Goiás)
- Фарида Вис, Универзитетот на Лестер (Farida Vis, University of Leicester)
- Пит Ворден, независен аналитичар на податоци и девелопер (Pete Warden, Independent Data Analyst and Developer)
- Крис Ву, „Хекс/Хакерс“ (Chrys Wu, Hacks/Hackers)

За што се работи (а за што не) во оваа книга

Целта на оваа книга е да послужи како корисен извор на информации за сите оние што мислат дека би сакале да се занимаваат со дата-новинарство – или барем да се испробаат на тоа поле.

Во пишувањето на книгата свој придонес дадоа многумина, а ние се обидовме, приредувајќи ја, да обезбедиме текстот да ги отсликува нивните индивидуални ставови и размислувања. Се надеваме дека читателите ќе ја примат книгата како еден опсежен и информативен разговор за тоа што претставува, какво е значењето и како изгледа занимавањето со новинарството засновано на податоци. Оваа книга, за жал, не нуди сеопфатен репертоар на знаења и вештини што се потребни за некој да може да се занимава со дата-новинарство. За такво нешто би била потребна огромна библиотека во која работат стотици експерти на разни полиња што би можеле да одговорат на стотици прашања од исто толку теми и области. Среќна околност е што таква библиотека веќе постои. Ја нарекуваме Интернет. Ние се надеваме дека оваа книга ќе ви помогне да дознаете како оние од вас што сакаат да продолжат да се занимаваат со дата-новинарство да ги направат првите чекори и каде да ги побараат потребните информации. Примерите и вежбите понудени во книгата се само илустрација, а не некаков конечен и дефинитивен суд за определено прашање.

Благодарни сме им на сите наши соработници за времето, енергијата и трпението што ни ги подарија и направивме сè што е во наша моќ тоа време, енергија и трпение да ги искористиме на најдобар можен начин. Се надеваме дека – освен што ќе послужи како корисен извор на референтни информации – оваа книга во определена мера е документ што ги бележи и страста и ентузијазмот, визијата и енергијата на едно движење во повој. Книгата е обид да ви доловиме што се случува зад сцената, да ви ги понудиме приказните што се одвиваат зад статиите.

Прирачникот за дата-новинарство не е завршен и исцрпен труд, и работата на негово подобрување тече секојдневно. Ако сметате дека нешто треба да се дополни, или дека постои некаква празнина што боде в очи, известете нè за да ја пополниме за следното издание. Книгата е достапна без надомест, под лиценца „Криејтив комонс Наведи извор-СподелиПодИстиУслови“ (Creative Commons Attribution-ShareAlike), на интернет-адресата www.datajournalismhandbook.org, а ние можеме само да ве охрабриме да ја споделите со сите оние за кои сметате дека би биле заинтересирани да ја прочитаат.

– Џонашан Греј, Фондација „Опворено знаење“ (@jwug), Лилијана Бунеџру, Европски центар за новинарство (@bb_liliana), и Луси Чејмберс, Фондација „Опворено знаење“ (@lucyfedra), во март 2012 година

Издавачки конвенции користени во оваа книга

Во книгава се користат следните издавачки и печатарски конвенции:

Курзив

Укажува на нов термин, интернет-адреса (URL), е-маил адреса, имиња на документи и екстензии за означување на форматот на документите.



Оваа слика означува корисен совет, сугестија или општа забелешка.



Оваа слика означува предупредување.

Safari® Books Online



Safari Books Online е дигитална библиотека чиј фонд на книги е достапен по барање, што нуди стручни содржини, во форма на текст и видео, од водечките автори на полето на технологијата и деловното работење во светот.

Професионалци на полето на технологијата, програмери и девелопери на софтверски решенија, веб-дизајнери и професионалци во областа на деловното работење и креативните области ја користат дигиталната библиотека „Safari Books Online“ како основен извор на информации и материјали за истражување, решавање проблеми, учење и обука за стекнување професионални сертификати.

„Safari Books Online“ нуди широк опсег на комплекти од производи и програми по повластени цени за организации, владини агенции и индивидуални лица. Претплатниците добиваат пристап до илјадници книги, видеа за обука и необјавени ракописи, излистани во база на податоци што може да се претражува. Материјалите се изданија на O'Reilly Media, Prentice Hall Professional, Addison-Wesley Professional, Microsoft Press, Sams, Que, Peachpit Press, Focal Press, Cisco Press, John Wiley & Sons, Syngress, Morgan Kaufmann, IBM Redbooks, Packt, Adobe Press, FT Press, Apress, Manning, New Riders, McGraw-Hill, Jones & Bartlett, Course Technology и десетици други издавачи. За повеќе информации за „Safari Books Online“, посетете ја нашата интернет-страница (<http://www.safaribooksonline.com/>).

Информации за контакт

Ве молиме, сите прашања и коментари што се однесуваат на оваа книга да ги испратите до издавачот:

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
800-998-9938 (in the United States or Canada)
707-829-0515 (international or local)
707-829-0104 (fax)

Оваа книга има и своја интернет-страница на која се наведени одредени исправки, дополнителни примери и други дополнителни информации. Интернет-страницата е достапна на адресата:

<http://oreil.ly/data-journalism-handbook>

Вашите коментари или технички прашања за книгата испратете ги по електронска пошта на:

bookquestions@oreilly.com

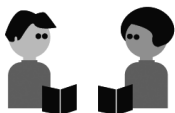
За повеќе информации за нашите изданија, курсеви, конференции, како и за најновите вести, посетете ја нашата интернет-страница на адресата

<http://www.oreilly.com>.

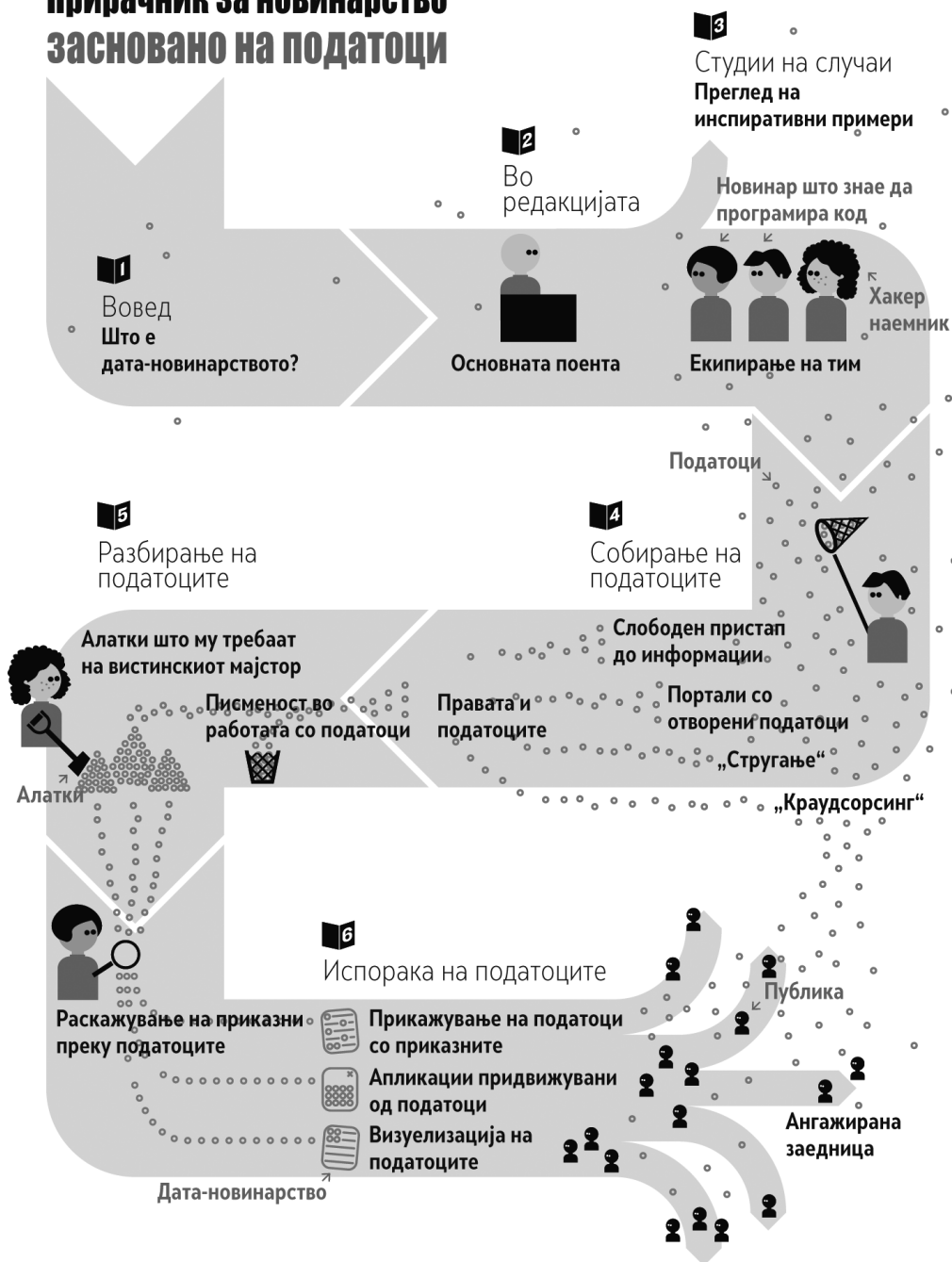
Можете да нè пронајдете и на Facebook: *<http://facebook.com/oreilly>*

Следете нè на Twitter: *<http://twitter.com/oreillymedia>*

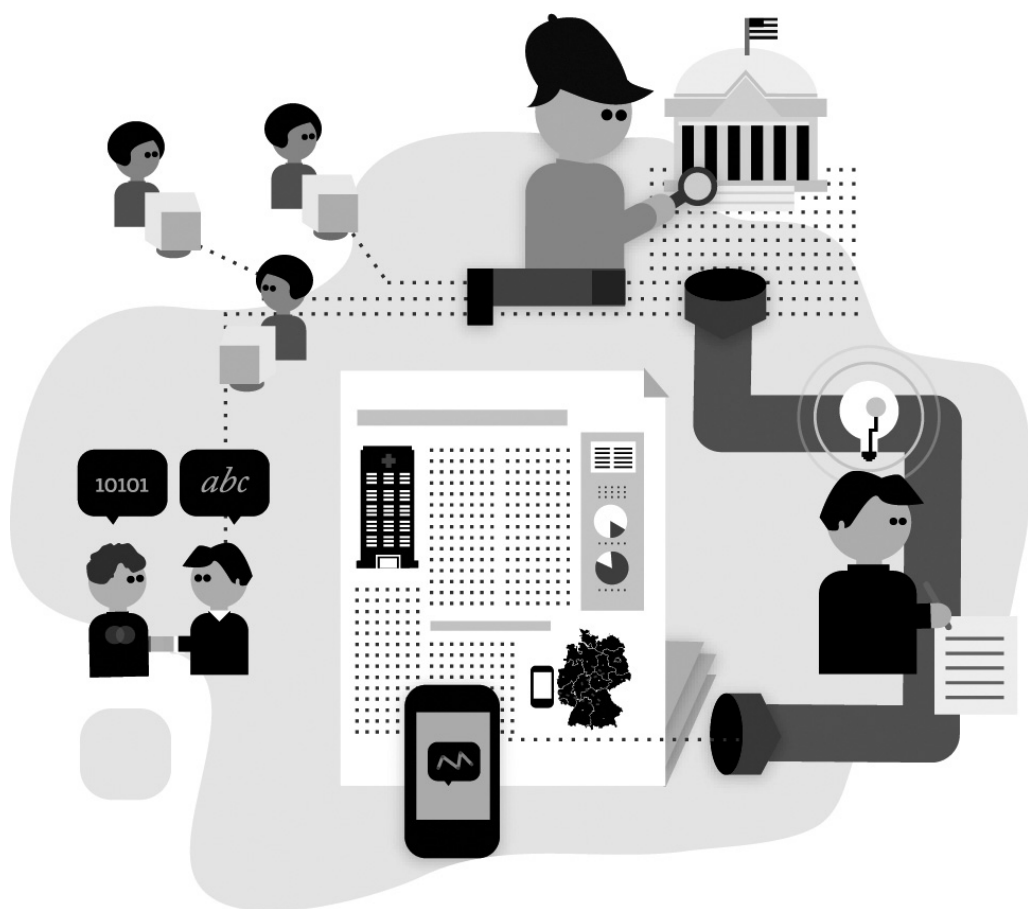
Погледајте ги нашите видеа на каналот на „О’Рајли Медија“ на YouTube: *<http://www.youtube.com/oreillymedia>*



Прирачник за новинарство засновано на податоци



Вовед



Што е новинарството засновано на податоци? Какви се неговите можности и потенцијали? Кои се ограничувањата со кои се соочува? Од каде потекнува? Во ова поглавје ќе видиме што е дата-новинарството и какво значење може да има тоа за новинските организации. Пол Бредшо од Градскиот универзитет на Бирмингем и Мирко Лоренц од „Дојче веле“ даваат краток осврт на основните карактеристики на новинарството засновано на податоци. Неколку водечки дата-новинари пишуваат за она што сметаат дека е значајно и ни ги нудат нивните омилен примери за дата-новинарство. На крајот, Лилијана Бунегру од Европскиот центар за новинарство дава осврт на новинарството засновано на податоци во еден поширок историски контекст.

Што е новинарство засновано на податоци?

За што зборуваме кога зборуваме за дата-новинарство? Наједноставно би било да се каже дека тоа е новинарство што се занимава со податоци. Сепак, не мислам дека тоа би ни било од особена помош.

Двата термини, „податоци“ и „новинарство“, се прилично проблематични. Некои луѓе, кога зборуваат за „податоци“, под тоа подразбираат каков било збир на броеви, најчесто собрани во сметководствени табели. Точно е дека пред дваесетина години тоа, помалку или повеќе, беше единствениот вид на податоци со кои се занимаваа новинарите. Денес, од друга страна, живееме во дигитален свет, свет во кој сè може да се опише и е опишано со бројки.

Вашата кариера, 300.000 доверливи документи, сите ваши пријатели; сето тоа може да се опише (и е опишано) со помош на две бројки: нули и единици. Фотографии, видео и аудио снимки се опишуваат со користење на истите две бројки: нули и единици. Убиства, болести, политички избори, корупција, лаги: сите тие се само нули и единици.

Каква е разликата помеѓу дата-новинарството и останатите видови на новинарство? Можеби разликата е во новите можности што тоа ги нуди за комбинација на традиционалната „способност да се намириса добра вест“ и способноста да се раскаже убедлива приказна преку огромниот обем и дијапазон на дигитални информации што денес се достапни.

Таквите можности се јавуваат во кој било стадиум на новинарскиот процес: користењето компјутерски програми за автоматизирање на процесот на собирање и комбинирање на информациите добиени од локалните власти, полицијата и други граѓански извори, нешто што Адријан Холовати (Adrian Holovaty) го прави на интернет-страниците „ЧикагоКрајм“ (Chicago-Crime) и „Евриблок“ (EveryBlock).

Или користењето софтвер за да се пронајдат врските меѓу стотици илјади документи, пристап што „Телеграф“ (The Telegraph) го примени на трошоците на членовите на Парламентот на Велика Британија (<http://tgr.ph/mps-expenses>).

Дата-новинарството може да му помогне на новинарот да раскаже една комплексна приказна со користење уверливи и атрактивни инфографики. На пример, спектакуларните говори на Ханс Рослинг (Hans Rosling) за визуелизацијата на сиромаштијата во светот со користење „Гепмајндер“ (Gepminder) привлекоа милиони гледачи во целиот свет. Исто така, популарните потфати на Дејвид Мекендлес (David McCandless) на прочистувањето големи броеви – ставањето на јавната потрошувачка во поширок контекст, загадувањето што ерупцијата на исландскиот вулкан истовремено го предизвика и го спречи – се добра илустрација за значењето на јасниот и едноставен дизајн што го користи студиото „Информациите се прекрасни“ (Information is Beautiful).

Дата-новинарството може да ни ја расветли врската помеѓу сторијата и некое лице, нешто што „Фајненшал тајмс“ и Би-Би-Си денес рутински го прават со нивните интерактивни прикази на буџетите (каде што можете да видите како буџетот влијае на вас лично, а не само на просечниот граѓанин „Петар Петровски“). Дата-новинарството може да го отвори и процесот на собирање вести и информации, нешто што „Гардијан“ со успех го прави во споделувањето податоци, различни контексти и прашања на својот специјализиран блогерски сајт за работа со податоци „Датаблог“ (Datablog), за учество на сите граѓани.

Податоците можат да бидат извор за дата-новинарство или да послужат како алатка за раскажување некоја приказна – или и двете нешта одеднаш. Како и секој друг извор, треба да им се пристапи со доза на скепса; а како и кај секоја друга алатка, треба да бидеме свесни за начините на кои можат да ги обликуваат или ограничат приказните што ги создаваме со нивна помош.

– Пол Бредшо, Градски универзитет на Бирмингем

Investigate your MP's expenses

Join us in digging through the documents of MPs' expenses to identify individual claims, or documents that you think merit further investigation. You can work through your own MP's expenses, or just hit the button below to start reviewing. (Update, Fri pm: we now have a virtually complete set of expenses documents so you should be able to find your MP's) Already created an account? [Log in here.](#)

We have **458,832** pages of documents. **32,755** of you have reviewed **225,443** of them. Only **233,389** to go...

Start reviewing

Please read our [privacy policy](#) to find out how we use your data. You must also read our [terms of service](#). By reviewing pages, you are agreeing that you have read the terms of service, and that you agree to them.

Слика 1-1. Илустрациите ги покажуваат на вашиот пратеник во Парламентот („Гардијан“)

Зошто новинарите треба да користат податоци

Новинарството денес е „сардисано“ од сите страни. Во минатото, ние од новинската индустрија, се потпиравме на тоа дека сме единствени што ја користат технологијата за да ги мултиплицираат и дистрибуираат настаните што се случиле во текот на денот. Печатарската преса беше порта кон светот. Тој што сакаше неговата порака следното утро да дојде до жителите на некој град или регион, мораше да се обрати кај весниците. Таа ера одамна е завршена.

Денес новинарските приказни течат паралелно со настаните, доаѓаат од повеќе извори, сведоци и блогови, а настаните се филтрираат низ една огромна мрежа од општествени врски што ги рангира и коментира, а многу често едноставно ги игнорира.

Токму тука се крие вистинското значење на новинарството засновано на податоци. Вредноста на собирањето, филтрирањето и визуелизацијата на она што се случува зад тоа што е видливо со обично око постојано расте. Сокот од портокали што го пиете за појадок, кафето што го варите: во денешната глобална економија постојат многу невидливи врски меѓу тие производи, другите луѓе и вас. Таа мрежа зборува со јазикот на податоците: мали точки што носат информации кои често изгледаат неважни, но добиваат огромно значење ако ги погледнеме од вистинскиот агол.

Во моментот додека ги пишуваме овие редови, неколку новинари прават пионерски обиди да покажат како податоците можат да се искористат за да дадат подлабок увид во она што се случува околу нас и како тоа може да влијае врз нас.

Анализата на податоците може да ги открие „контурите на една историја“ (Сара Коен) или да ни даде „нов објектив низ кој ќе ја набљудуваме приказната“ (Дејвид Мекендлес). Користењето податоци го поместува фокусот на новинарската работа од обидот да се биде прв што ќе објави нешто кон потребата да се објасни значењето на тоа што се случило. Дијапазонот на теми може да биде мошне широк. На пример, наредната финансиска криза која во моментот е во процес на настанување. Економските процеси зад производите што ги користиме. Злоупотребите на фондовите или политичките грешки, претставени во форма на атрактивна визуелизација што остава малку место за сомневање или оспорување.

Тоа се причините поради кои новинарите треба да гледаат на податоците како на нова можност. Податоците можат, на пример, да откријат како некоја апстрактна закана (како што е невработеноста) влијае врз луѓето врз основа на нивната возраст, род или образование. Користењето податоци претвора нешто што е апстрактно во нешто што сите можат да го разберат и спрема кое имаат некаков однос.

Тие можат да создадат персонализирани калкулатори што ќе ни помогнат во носењето одлуки, сеедно дали се работи за купување автомобил, кука, за

продолжување на образованието или изборот на професијата, или за проверка на трошоците во обид да се избегне задолженоста.

Можат да ја анализираат динамиката на некоја комплексна ситуација, какви што се протестите или политичката дебата, да укажат на логичките грешки и да помогнат во идентификувањето можни решенија за сложените проблеми.

Стекнувањето знаења на полето на пребарувањето, прочистувањето и визуелизирањето на податоците ја преобразува и професијата на собирање информации. Новинарите што ќе го усовршат тоа знаење ќе можат да се уверат дека градењето на нивните статии врз факти и наоди претставува олеснување. Помалку нагаѓање, помалку барање изјави и цитати; наместо тоа, новинарот може да изгради цврста позиција поткрепена со податоци, а тоа може сериозно да влијае на улогата што ја има новинарството воопшто.

Дополнително, влезот во областа на дата-новинарството нуди и перспективи за иднината. Денес, во услови на намалување на редакциите, најголемиот број новинари се надеваат дека ќе можат да преминат во областа на односите со јавноста. Дата-новинарите и стручњаците за обработка на податоци веќе се мошне баран кадар, и тоа не само во медиумите. Компаниите и институциите во целиот свет бараат „осмислувачи“ и професионалци што знаат како да „копаат“ низ податоците и да ги трансформираат во нешто поопипливо.

Податоците имаат потенцијал што ветува, а тоа ги раздвижува редакциите и ги тера да бараат нов вид репортери. За слободните новинари, познавањето на работата со податоците е добар начин да стигнат до нови понуди за ангажман и до постабилен извор на приходи. Гледајте на тоа на следниов начин: наместо плаќање новинари што брзо ќе ги пополнуваат страниците на весниците и интернет-страниците со содржини со мала вредност, користењето на податоците може да создаде побарувачка од интерактивни пакети, каде што работењето цела седмица на решавање на едно единствено прашање е единствениот начин да се заврши работата. Таа промена е добредојдена во многу сегменти од медиумите.

Сепак, има една бариера што ги спречува новинарите да го користат тој потенцијал: потребата да се обучат за сите чекори во работењето со податоци – од почетното прашање до големата истражувачка приказна заснована на податоци.

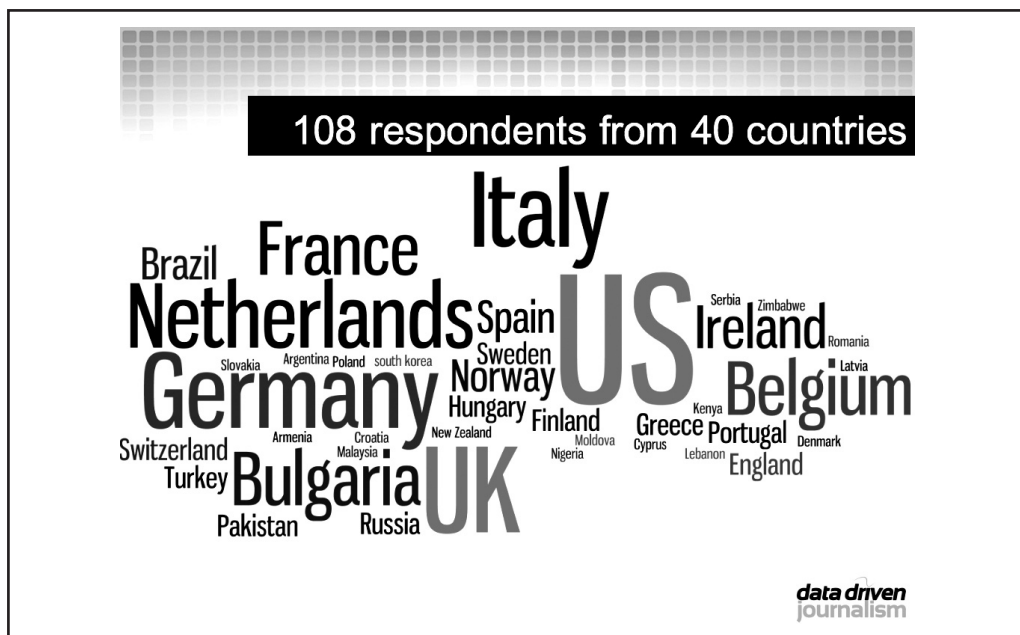
Работењето со податоци прилега на зачекорување на нов, огромен и неистражен континент. На прв поглед, необработените податоци се збунувачки. Необработените податоци изгледаат несмасно и збркано. Тешко е правилно да се обликуваат за да може да се визуелизираат. За тоа се потребни искусни новинари што имаат доволно кондиција за прегледување, мошне често, збркани или здодевни податоци и во нив да „видат“ скриени приказни.

– Мирко Лоренц, „Дојче веле“

Истражувањето на потребите за обука

Европскиот центар за новинарство спроведе истражување за да утврди на кои полиња им е потребна обука на новинарите. Откривме дека кај нив постои подготвеност да излезат од зоната на традиционалното новинарство, во која се чувствуваат најудобно и најбезбедно, и да го вложат своето време во усвојување на нови вештини. Резултатите од истражувањето покажуваат дека новинарите се свесни за можностите, но дека им треба поддршка во надминувањето на почетните проблеми што ги спречуваат да се зафатат со работа со податоците. Откривме дека постои високо ниво на убеденост дека, ако дата-новинарството стане општоприфатена пракса, брзо ќе дојде до подобрувања во начинот на работа, алатките и резултатите. Пионерите во таа област, како што се „Гардијан“, „Њујорк тајмс“, „Тексас трибјун“ и „Цајт“ продолжуваат да отвараат нови граници во своите приказни засновани на податоци.

Дали дата-новинарството ќе остане „забран“ на неколку пионери, или наскоро сите новински организации ќе имаат свои тимови посветени на дата-новинарството? Се надеваме дека овој прирачник ќе поттикне повеќе новинари и редакции да ги искористат можностите што ги нуди ова ново поле на новинарската работа.



Слика 1-2. Истражувањето на потребите за обука на Европскиот центар за новинарство

Зошто е значајно новинарството засновано на податоци?

Побаравме од неколку водечки практичари и промотори на новинарството засновано на податоци да ни кажат зошто сметаат дека станува збор за значајна новина. Ова се нивните одговори.

Филтрирање на протокот на податоците

Порано, кога информациите беа скудни, најголемиот дел од времето го минувавме во „лов и собирање плодови“. Денес, кога постои изобилство од информации, позначајно е како ќе ги обработиме. Обработката ја вршиме на две нивоа: 1) анализа која треба да внесе смисла и структура во непресушниот проток на податоци и 2) презентација која треба на потрошувачите да им покаже што е тоа што е важно и релевантно.

Слично на научните дисциплини, дата-новинарството не ги крие методите што ги користи и ги претставува своите наоди на начин што овозможува да бидат потврдени со повторување.

— *Филиј Мејер, њочесен ѡрофесор на Универзитетот на Северна Каролина во Чейел Хил*

Нови приоди кон раскажувањето приказни

Дата-новинарство е кровен термин што, барем така мислам, го покрива растечкиот збир од алатки, техники и приоди кон раскажувањето приказни. Тоа вклучува сè, од традиционалното известување помогнато од компјутерската технологија (користење на податоците како „извор“) до најсовремените и најнапредни апликации за визуелизација на податоци и пренесување вести. Крајната цел е новинарска по природа: обезбедување информации и анализи што треба да ни помогнат да бидеме информирани за значајните прашања што нè опкружуваат.

— *Арон Пилхофер, „Њујорк тајмс“*

Слично е на фото-новинарството, но се врши со лаптоп

„Дата-новинарството“ се разликува од „новинарството со зборови“ само по различните алатки што ги користи. Ние сите душкаме наоколу, известуваме и ги пренесуваме приказните за да заработиме за живот. Слично е на „фото-новинарството“, само заменете ја камерата со преносен компјутер.

— *Брајан Бојер, „Чикаго трибјун“*

Иднината е во дата-новинарството

Иднината лежи во новинарството засновано на податоци. Новинарите мора да бидат вешти во користењето податоци. Порано баравме приказни разговарајќи со луѓето во кафеани, и не велам дека некои од вас сè уште не го прават тоа од време на време. Но денес сè повеќе се работи за прегледување податоци и опремување со алатки за нивна анализа и одвојување на она што е интересно. Исто така, и нивно ставање во перспектива за да им помогнеме на луѓето да разберат како сето тоа се вклопува во определена слика и што се случува во земјата.

— Тим Бернерс Ли (Tim Berners-Lee), основач на „Глобалната мрежа“ (World Wide Web)

Аритметиката се среќава со пишувањето

Новинарството засновано на податоци го премостува јазот помеѓу статистичарите и пишувачите. Се работи за определување на координатните точки и идентификување на трендовите што не се значајни само од статистичка гледна точка, туку и во расчленувањето на инхерентно комплексниот свет во кој живееме денес.

— Дејвид Андерџон (David Anderton), слободен новинар

Унапредување на вештините што ги поседувате

Новинарството засновано на податоци претставува сосема нов комплет од вештини за пребарување, разбирање и визуелизирање на дигиталните извори во време во кое основните вештини на традиционалното новинарство едноставно не се доволни. Тоа не е замена, туку дополнување на традиционалното новинарство.

Во време кога изворите се селат во дигиталниот свет, новинарите можат и мораат да останат во близина на тие извори. Интернетот отвара можности кои ја надминуваат способноста за разбирање што во моментот ја поседуваме. Дата-новинарството е само почетен стадиум на еволуцијата на практиките од минатото и нивното приспособување за работа во онлајн-доменот.

Дата-новинарството им помага на новинските организации во постигнувањето две значајни цели: пронаоѓање уникатни стории (кои не потекнуваат од информативните агенции) и вршење на функцијата на контролор на власта. Тоа се значајни цели за весниците, особено во време на финансиска несигурност.

Од позиција на еден регионален дневен весник, улогата на дата-новинарството е клучна. Постои една изрека дека „расклиманата плоча на тротоарот

пред вашата влезна врата е поважна од немирите во некоја далечна земја“. Таа плоча ве засега и влијае врз вашиот живот на многу подиректен начин. Во исто време, светот забрзано се дигитализира. Поради директното влијание што локалните весници го имаат во својата заедница, а изворите истовремено се дигитализираат, новинарот мора да знае како во податоците да пронајде, анализира и визуелизира некоја приказна.

– Џери Ферманен, *NU.nl*

Лек за информатичката асиметрија

Информатичката асиметрија – не се работи за недостиг на информации туку за неможност тие да се прифатат и процесираат поради брзината и количеството кои доаѓаат до нас – е еден од најзначајните проблеми со кои граѓаните се соочуваат во одлучувањето на кој начин ќе го живеат својот живот. Информациите што ги добиваат од печатените и аудиовизуелните медиуми влијаат врз нивниот избор и активности. Доброто дата-новинарство помага во борбата против информатичката асиметрија.

– Том Фрис, Фондација „Берџелсман“

Одговор на работата на одделите за односи со јавноста засновани на податоци

Достапноста на алатки за мерење и нивната сè пониска цена – во самоодржлива комбинација фокусирана на изведбата и ефикасноста во сите аспекти на општеството – ги принудува носителите на одлуки да го квантифицираат напредокот на нивните политики, да ги следат трендовите и да ги идентификуваат можностите.

Компаниите постојано пронаоѓаат нови начини за мерење на успехот на нивната работа. Политичарите љубат да се фалат со бројките за намалувањето на невработеноста и порастот на БДП. Недостигот на новинарски увид во аферите „Енрон“ (Enron), „Ворлдком“ (Worldcom), „Медоф“ (Madoff) или „Солиндра“ (Solyndra) е доказ дека многу новинари не поседуваат способност за разбирање на броевите. Голема е веројатноста бројките да бидат прифатени такви какви што се многу повеќе од другите факти, затоа што поседуваат аура на сериозност, дури и тогаш кога се целосно измислени.

Познавањето на работата со податоци ќе им помогне на новинарите да го изострат својот критички поглед кога разгледуваат броеви и, искрено се надеваме, ќе им помогне повторно да освојат дел од територијата изгубена во „војната“ со одделенијата за односи со јавноста.

– Николас Кајзер-Брил, „Журнализам++“ (*Journalism++*)

Независно толкување на официјалните информации

Разурнувачкиот земјотрес и инцидентот во атомската централа во Фукушима во 2011 година ѝ го потврдија на медиумската заедница во Јапонија, земја која генерално заостанува на полето на дигиталното новинарство, значењето на дата-новинарството.

Бевме целосно збунети и изгубени кога властите и експертите не можеа да понудат веродостојни податоци за претрпената штета. Кога официјалните лица од јавноста ги скрија податоците од „СПЕЕДИ“ (SPEEDI), систем за предвидување на распространувањето на радиоактивните материјали, ние не бевме подготвени да ги дешифрираме тие податоци, дури и да успеевме да дојдеме до нив. Граѓаните доброволно почнаа да ги собираат податоците за нивото на радијација со сопствени уреди, но ние не го имавме потребното знаење на полето на статистиката, интерполацијата на податоци, визуелизацијата и така натаму. Новинарите треба да имаат пристап до сурови податоци, но и да знаат дека не можат да се потпрат на официјалните интерпретации на податоците.

— Исао Мацунами, „Токио Шимбун“

Справување со поплавата од податоци

Предизвиците и можностите што со себе ги носи дигиталната револуција продолжуваат да го потресуваат новинарството. Во време на изобилство на информации, и на новинарите и на граѓаните им се потребни подобри алатки, без оглед дали се грижат за „самиздатите“ на 21-от век на Блискиот Исток, дали обработуваат новообјавени податоци до длабоко во ноќта или бараат најдобар начин за визуелизација на квалитетот на водата за една потрошувачка нација. Додека се обидуваме да се фатиме во костец со проблемите на потрошувачката што ги предизвикува таквиот порој од податоци, новите платформи за објавување им овозможуваат на сите луѓе податоците да ги собираат и разменуваат дигитално, трансформирајќи ги во информации. Ако порано новинарите и уредниците беа традиционалните вектори за собирање и ширење информации, во нивелираната информатичка средина во 2012 година, новите вести попрво настануваат на интернет, а не на редакцискиот деск.

Насекаде во светот таа врска помеѓу податоците и новинарството станува сè поцврста. Во ера на големи количества податоци, растечкото значење на дата-новинарството лежи во способноста на оние што го практикуват да обезбедат контекст, јасност и – можеби најважно – откривање на вистината во постојано растечкото количество дигитални содржини. Тоа не значи дека денешните интегрирани медиумски организации повеќе ја немаат клучната улога. Ни случајно. Во информатичката ера, новинарите ни се потребни повеќе од кога било порано да ги чуваат, верификуваат, анализираат и синтетизираат податоците. Токму тој контекст му го дава на новинарството засновано на податоци огромното значење што тоа го има во општеството.

Разбирањето на големите количества податоци, особено ако не се структурирани, денес се поставува како главна цел за оние што работат со податоци на научна основа, без оглед дали работат за некоја редакција, во финансискиот сектор на Волстрит или во Силиконската долина. Секако, во постигнувањето на таа цел ќе помогне еден комплет од алатки кој постојано се збогатува, независно од тоа дали таквите алатки ги користат владините стручњаци за технологија што информатички го „отвораат“ градот Чикаго, оние што се занимаваат со технологијата на здравствените услуги или девелоперите што работат во новински редакции.

— Алекс Хауард, *О’Рајли Мегија*

Нашите животи се збир од податоци

Доброто дата-новинарство е тешка и напорна работа, затоа што секое добро новинарство е тешка и напорна работа. Тоа значи дека треба да се смисли како да се добијат податоците, како тие податоци да бидат разбрани и како да се пронајде приказната што ја кријат. Понекогаш ќе нè одведат во кор-сокак, а понекогаш во нив нема никаква голема приказна. Најпосле, ако работата е само да се притисне вистинското копче, тоа не е новинарство. А токму таму лежи вредноста на дата-новинарството и му дава суштинско значење за постоењето слободно и праведно општество во свет во кој и нашите животи сè повеќе се сведуваат на збир од податоци.

— Крис Таџарџи, *„ОјенКорјорејџс“*

Заштеда на време

Новинарите не можат да губат време на рачна транскрипција за да ги извадат податоците од „ПДФ“ документ (PDF), така што од огромно значење е да знаат како да напишат неколку линии програмерски код (или да знаат каде да ги побараат луѓето што можат да им помогнат).

Еден новинар од „Фоља де Сао Паоло“ (Folha de São Paulo) што работеше на обработка на локален буџет ме побара да ни заблагодари што сме ги објавиле сметките на општината Сао Паоло на интернет (два дена работа за еден единствен хакер!). Новинарот ни кажа дека претходните три месеци се обидуваа сам, на рака, да ги транскрибира тие податоци во обид да ја изгради својата приказна. Помнам и како го решивме „проблемот со ПДФ документите“ за „Конџас аберџас“ (Contas Abertas, *„Ојворени сметки“*), новинска организација што ја следи работата на парламентот: 15 минути работа и 15 линии компјутерски код заменија цел месец работа на рачно внесување на податоците.

— Педро Маркун, *„Трансџаренција хакер“*

Суштински елемент во комплетот новинарски алатки

Сметам дека е значајно да се стави акцент на „новинарството“ или на известувачките аспекти на „дата-новинарството“. Анализата и визуелизацијата на податоците не треба да се спроведуваат сами за себе, туку за да се искористат за да се дојде до вистината за тоа што се случува во светот. На способноста за анализа и толкување на податоците гледам како на основен и суштински елемент на современиот комплет на алатки со кои се служат новинарите, а не како на одвоена дисциплина. Најпосле, поентата е во тоа дали се работи за добро новинарство или не, и дали приказните се пренесуваат на најдобриот и најсоодветен начин.

Дата-новинарството е нов начин за следење на тоа што се случува во светот и за повикување на одговорност на оние што ја имаат моќта. Како што се зголемува количеството податоци што ни се достапни, сè позначајно е новинарите да бидат свесни за техниките што ги воведува новинарството засновано на податоци. Сите новинари треба да го имаат вклучено во својот комплет на алатки, без оглед дали учат како самите да работат со податоци или соработуваат со некој што знае како се прави тоа.

Вистинската моќ на дата-новинарството се крие во помошта што ви ја дава во пронаоѓањето информации кои инаку би биле многу тешки за откривање или докажување. Добар пример за тоа е приказната на Стив Дојг која дава анализа на штетите што ги предизвика ураганот „Ендрју“. Дојг спои две различни групи податоци: првата го мапира нивото на уништување предизвикано од ураганот, а втората ја прикажува промената во брзината на ветерот. Тоа му помогна да ги пронајде оние области во кои поблагата регулатива во градежништвото и пониските стандарди на градба придонеле да настане поголема штета. За својата приказна, Дојг ја доби „Пулицеровата награда“ за 1993 година (<http://www.pulitzer.org/awards/1993>), а приказната и понатаму служи како одличен пример за тоа што може да се постигне со дата-новинарството.

Во идеални услови, податоците ги користите за пронаоѓање на граничните вредности и координатите, областите од посебен интерес или оние нешта што отскокнуваат и претставуваат изненадување. Во таа смисла, податоците можат да послужат како трага што ќе ја следите или како замена за доверливата информација што сте ја добиле од информаторот. Иако бројките можат да бидат интересни, не е доволно само да пишувате за податоците. Сè уште ви преостанува задачата да објасните, преку класично известување, кое е значењето на тие податоци.

— Синџија О’Мурку, Фајненшиал ѿајмс

Приспособување на промените во информатичката околина

Новите дигитални технологии со себе носат нови начини на производство и ширење на знаењето во општеството. На дата-новинарството можеме да гледаме и како на обид на медиумите да се приспособат и да одговорат на предизвиците во ова информатичко опкружување, вклучувајќи го и интерактивното, повеќедимензионално раскажување што им овозможува на читателите да ги истражат изворите на кои се заснова веста и да ги охрабри да се вклучат во процесот на создавање и оценување на приказните.

— Сезар Вијана, Универзитетот во Гојас

Воочување на нештата што инаку не би ги виделе

Некои приказни можат да бидат разбрани и објаснети само преку анализа – а понекогаш и визуелизација – на податоците. Врските помеѓу оние лица или субјекти што ја имаат моќта би останале неоткриени, бројот на смртни случаи што ги предизвикале политиките на борба против дрогата би останале скриени, еколошките политики што ѝ наштетуваат на животната средина би останале недопрени. А сепак, сите нешта спомнати во претходната реченица беа променети благодарение на податоци собрани, анализирани и понудени на читателите од страна на новинари. Тие податоци можат да бидат едноставни, проста сметководствена табела или запис на телефонските разговори направени преку мобилен телефон, или комплексни, на пример, резултати од училишните тестирања или податоци за инфекциите што настанале во една болница. Сепак, во сите нив се кријат приказни што вреди да бидат раскажани.

– *Черил Филијс, „Сиетл ѿјмс“*

Можност за богатување на приказните

Дигиталните траги што ги оставаме зад себе можат да исцртаат слика на нашите животи. Од она што го купуваме и прегледуваме, каде и кога патуваме, нашите музички вкусови, први љубови, значајните постигнувања на нашите деца, дури и нашите последни желби – сето тоа може да се следи, дигитализира, да се чува во облак („cloud“) и да се шири наоколу. Тој универзум од податоци може да се извади на површината за да се раскажуваат приказни, да се одговараат прашања и да се пренесе некое сфаќање за животот на начини што во моментов далеку ги надминуваат дури и најдоследните и највнимателно реконструирани анегдоти.

– *Сара Слобин, „Волсѿриѿ журнал“*

Не ви требаат нови податоци за нова приказна

Понекогаш податоците што ви се потребни веќе постојат, но никој не се потрудил да ги разгледа повнимателно. Во случајот на извештајот на „Асошиетед прес“ (Associated Press) што се однесуваше на 4,500 страници декласифицирани документи за активностите и работата на приватните компании за обезбедување за време на војната во Ирак, материјалите веќе беа собрани од еден независен новинар преку барања за пристап до информации поднесени до Стејт департментот на САД. Новинарите на „Асошиетед прес“ ги скенираа добиените документи и ги поставија на „ДокументКлауд“ (DocumentCloud), што нам ни овозможи да ја спроведеме нашата детална и сеопфатна анализа.

– *Џонаѿан Сѿреј, Проектѿ „Овервју“*

Неколку омилен примери

Побаравме од неколку од нашите соработници да ни ги посочат своите омилен примери на новинарство засновано на податоци и да ни кажат што најмногу ни се допаѓа кај нив. Ова се нивните примери.

Серијата „Не нанесувај штета“ во „Лас Вегас сан“

Мојот омилен пример е серијата статии за болничката нега „Не нанесувај штета“ на весникот „Лас Вегас сан“ (<http://www.lasvegassun.com/hospital-care/>). „Лас Вегас сан“ анализираше повеќе од 2,9 милиони сметки за болничко лекување, при што откри повеќе од 3,600 повреди, инфекции и хируршки грешки што можеле да бидат спречени. Податоците беа добиени преку барање за пристап до јавни документи и преку нив се идентификувани повеќе од 300 случаи на пациенти кои починале поради грешки што можеле да се спречат. Серијата содржи различни елементи, вклучувајќи и интерактивен графикон што им овозможува на читателите да видат (прикажано по индивидуални болници) каде имало повеќе хируршки грешки од очекуваното; мапа со хронолошки приказ на ширењето на инфекциите по болници; како и интерактивно графичко решение што им овозможува на корисниците да ги подредуваат податоците според повредите што можеле да се спречат, или според болницата каде настанале, за да видат во кои болници луѓето трпат повреди. Мене ми се допаѓа затоа што е лесна за разбирање и навигација. Корисниците можат да ги прегледаат податоците на начин што е мошне интуитивен.

Сторијата имаше реален ефект: законодавните власти во државата Невада одговорија со усвојување шест законски прописи. Инволвираните новинари работеа напорно да ги добијат и прочистат податоците. Еден од нив, Алекс Ричардс (Alex Richards), неколку десетини пати испраќаше податоци назад во болниците со барање за исправка на грешките.

— Анхелика Пералџа Рамос, Ла Насион (Арџентина)



Слика 1-4. Платиите на владините службеници („Тексас трибјун“)

Визуелизација на ирачките воени дневници, Асошијетед прес

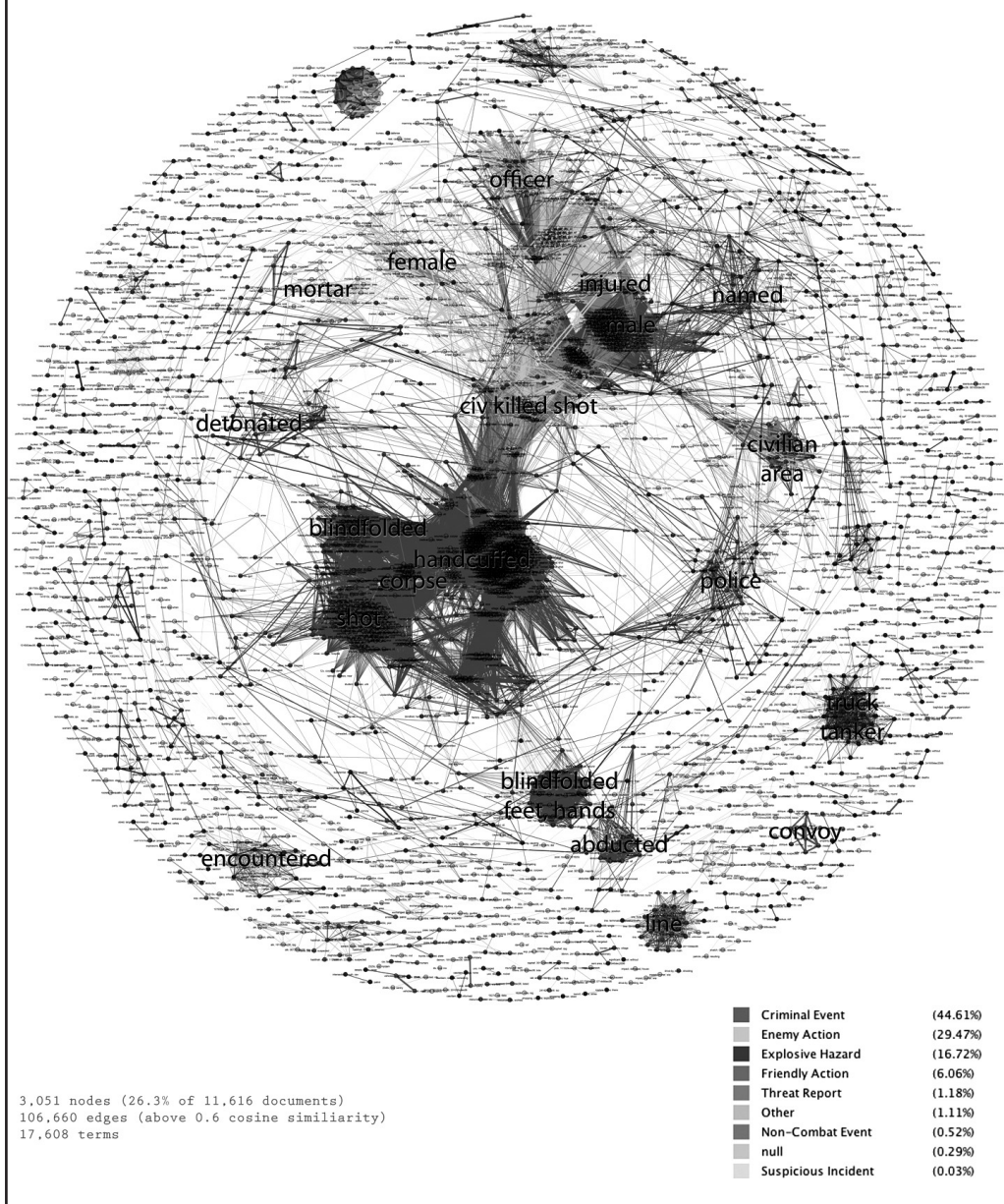
Работата со ирачките воени дневници на Џонатан Стреј и Џулијан Бардис е инспиративен чекор на полето на анализа и визуелизација на текстови со користење техники што овозможуваат да се добие увид во темите што се наоѓаат во еден голем комплет текстуални податоци и кои вреди понатаму да се истражуваат.

Со помош на техники и алгоритми за анализа на текст, Џонатан и Џулијан создадоа метод за прикажување на гроздовите и концентрациите на клучни зборови во илјадници извештаи на владата на САД за војната во Ирак, објавени од Викиликс (WikiLeaks) во визуелен формат.

Иако методот си има свои ограничувања и е експериментален, сепак се работи за свеж и иновативен пристап. Наместо читање на сите документи или прегледување на воените дневници со една однапред усвоена замисла што може да се пронајде во нив со внесување одредени клучни зборови и преглед на добиените резултати, оваа техника ги пресметува и визуелизира темите/клучните зборови со определено и специфично значење.

Пронаоѓањето начини да се одредат клучните области на интерес ќе биде се позначајно како што расте количината на текстуални (електронска пошта, извештаи и сл.) и нумерички податоци што пристигаат во јавниот домен. Се работи за возбудливо поле внатре во областа на дата-новинарството.

– Синџија О’Мурку, „Фајненшал тајмс“

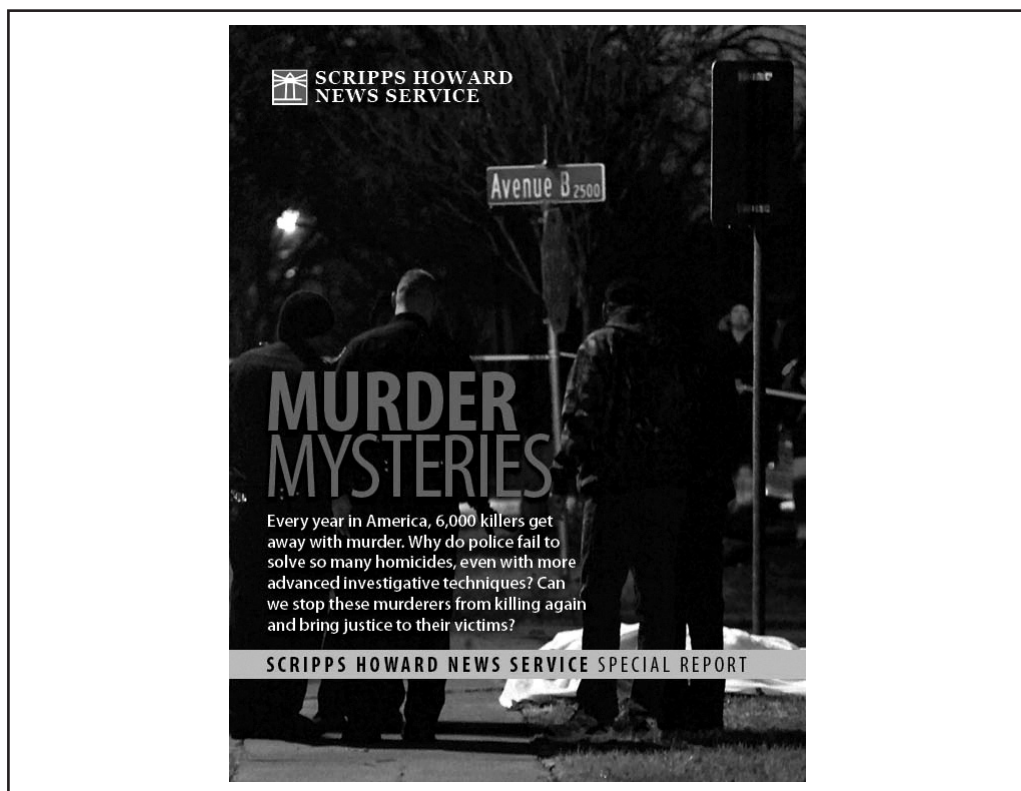


Слика 1-5. Анализайа на военише дневници („Асошиетед прес“)

Нерешени убиства

Еден од моите омилен примери за дата-новинарството е проектот „Нерешени убиства“ на Том Харгроув (Tom Hargrove) од Информативниот сервис „Скрипс Хауард“ (Scripps Howard News Service) (<http://bit.ly/murder-mysteries>). Користејќи ги достапните податоци објавени од властите и барања за пристап до јавни документи, тој изгради детална демографска база на податоци која содржи повеќе од 185.000 нерешени убиства и потоа дизајнираше алгоритам за пребарување за да ги открие можните обрасци и урнеци што би укажувале на можни сериски убијци. Во проектот имате сè: напорна работа, база на податоци што е подобра и од базата што ја поседуваат властите, паметна анализа што ги користи техниките на општествените науки и интерактивна презентација на податоците на интернет така што читателите можат и самите да ги прегледаат и проверуваат.

— Стив Дојџ, Школа за новинарство „Волиер Кронкајш“,
Државен универзитет на Аризона



Слика 1-6. Нерешени убиства (Информативен сервис „Скрипс Хауард“)

Машина за пораки

Ми се допаѓаат приказната „Машина за пораки“ (Message Machine) и „Блог-постот на бубалицата“ (nerd blog post (<http://bit.ly/nerd-blog-post>)) на „ПроПублика“. Целата таа работа почна кога љубопитноста на неколку „џивийерџии“ беше разбудена откако добија различни емаил-пораки од изборниот штаб на (Барак) Обама. Во „ПроПублика“ ги забележаа тие пораки и побараа од нивните посетители да им ги проследат сите емаил-пораки што ги добиле од изборниот штаб. Презентацијата е елегантна, визуелен „diff“ на кој се прикажани неколку различни пораки испратени истата вечер. Она што е сјајно е што тие самите ги собраа податоците (иако се работи за мал примерок, сепак, доволно голем за да може да се раскаже приказната). Она што е уште подобро е што ја раскажуваат приказната за еден феномен во настанување: користењето големи комплети податоци од страна на политичките кампањи за да се испраќаат насочени пораки до точно определени индивидуални лица. Тоа е само навестување за она што доаѓа.

– Брајан Бојер, „Чикаго трибјун“

PRO PUBLICA

Journalism in the Public Interest

Receive our top stories daily

Email address

SUBSCRIBE

Home

Our Investigations

Tools & Data

MuckReads

About Us

Search ProPublica

Message Machine: “You Probably Don’t Know Janet”

by Jeff Larson, Al Shaw and Lois Beckett
ProPublica, March 6, 6:29 p.m.

Tweet 114

Like

E-mail

Print

Campaigns are increasingly tailoring their messages – and their funding requests – using massive databases of personal information about potential voters. Here are six variations of a Thursday night message from the Obama campaign, based on emails submitted by 190 recipients across the country. Click a tab to select an email. Then hover over the other tabs to compare your selection to the others. [Related Story »](#)

EMAIL 1

EMAIL 2

EMAIL 3

EMAIL 4

EMAIL 5

EMAIL 6

155

People in our sample got this email

EMAIL 1 vs. EMAIL 3

+++++ ----

About This Group

Gender

Age

17

74

Previous Donor?

YES

NO

Requested Donation

\$3

\$300

Subject Line

Emails

Janet just found out

151

Janet from Accokeek

3

Meet Janet

1

This email, which touted the chance to win a seat at a dinner with the President, asked some recipients for a donation of \$3 and others for a donation \$25. Many of those who

"You're going to have dinner with the President."

There are only a handful of people who will ever ~~think they'll be picked until they are.~~

Take the chance. ~~Chip~~ hear that sentence. And we're Janet from Accokeek, Maryland. But you Maryland, just ~~trying to show her~~ became one of them.

She learned she's the first guest to be selected for the ~~next dinner with Barack~~ upcoming dinner the President's having ~~dinner with~~ with four supporters.

We're counting down the hours until we ~~draw the next name~~

The ~~second seat could be yours~~ — pick the next winner. Will you be the next person who will hear that sentence?

Pitch in ~~\$XX or more today to be automatically entered~~

\$X or whatever you can ~~today~~.

Janet told us that when she donated ~~\$XX~~ last week she was to help support for the President.

No this campaign today, and you'll be automatically entered ~~when you donate \$XX~~ for the chance to join President ~~Obama together~~ Obama, Janet, and two ~~could be~~ other supporters for dinner sometime soon.

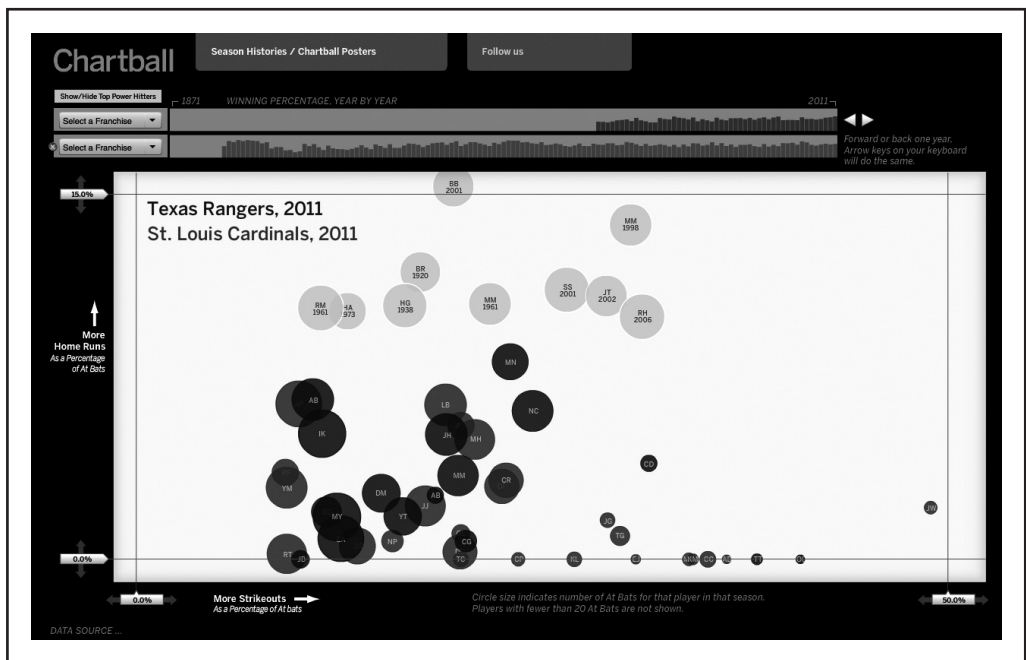
Слика 1-7. Машина за пораки (ПроПублика)

Неколку омилен примери | 37

„Чартбол“ (Chartball)

Еден од моите омилен проект за дата-новинарство е она што Ендрију Гарсија Филипс (Andrew Garcia Phillips) го работи на „Чартбол“ (Chartball, <http://www.chartball.com/>). Ендрију е голем љубител на спортот со незаситен апетит за податоци, извонредно око за дизајн и способност за пишување компјутерски код. Во „Чартбол“ тој не го визуелизира само текот на настаните, туку и детали за успехите и неуспехите на индивидуални играчи и тимови. Тој го дава контекстот, нуди примамливи графички решенија, а работата што ја врши е длабокоумна, забавна и интересна – а јас лично и не сум нешто посебно заинтересирана за спортот!

– Сара Слобин, „Волсґрийй журнал“



Слика 1-8. Победиџе и љоразииџе на џрафикон („Чарџбол“)

Дата-новинарството во перспектива

Во август 2010 година, со неколку колеги од Европскиот центар за новинарство организиравме, во Амстердам, нешто за што сметаме дека е една од првите меѓународни конференции за дата-новинарство. Во тоа време не се разговараше многу за оваа тема, а организациите познати по тоа што работат во оваа област можеа да се избројат на прстите на една рака.

Начинот на кој медиумските организации како што се „Гардијан“ и „Њујорк тајмс“ се справуваа со огромните количества податоци што ги објави „Викиликс“ е важен чекор напред што му донесе повеќе видливост и на терминот „дата-новинарство“. Некаде во тоа време, терминот почна да влегува во многу поширока употреба (заедно со терминот „известување со помош на компјутер“) за да опише на кој начин новинарите ги користат податоците за подобрување на нивното известување и за зајакнување на длабинските истражувања на дадената тема.

Од разговорите што со искусни новинари на полето на дата-новинарство и експерти за новинарство ги водев на „Твитер“ (Twitter) произлегува дека една од најраните формулации за она што денес го препознаваме како дата-новинарство е дадена во 2006 година од страна на Адријан Холовати, основачот на „ЕвриБлок“ (EveryBlock), информативен сервис што им овозможува на корисниците да откријат што се случува во нивното соседство, во нивниот квартал. Во краткиот есеј „Една фундаментална промена низ која мора да поминат инернет-страниците на весниците“ (“A fundamental way newspaper sites need to change”) тој тврди дека новинарите, покрај традиционалните „големи површини текст“, треба да објавуваат структурирани податоци погодни за машинско читање.

Да речеме, на пример, дека еден весник објавил сторија за локален пожар. Можноста таа сторија да ја прочита на мобилниот телефон, сето тоа е фино и убаво. Браво за технологијата! Но, јас всушност сакам да можам да ги истражам суровите факти од таа приказна, еден по еден, со различните слоеви на атрибуција, како и инфраструктура која ќе ми овозможи споредба на деталите за пожарот со детали од претходните пожари: датумот, времето, местото, жртвите, бројот на противпожарната единица, имињата и стажот на пожарникарите што го гаселе пожарот, времето што им било потребно да стигнат на местото на пожарот, како и со детали за наредните пожари кога ќе се случат.

Но, што е разликата помеѓу тоа што е спомнато погоре и другите форми на новинарство што користат бази на податоци или компјутери? Како, и во кој степен, дата-новинарството се разликува од другите форми на новинарство во минатото?

Известување со помош на компјутер и прецизно новинарство

Користењето податоци за подобрување на известувањето и обезбедување структурирани (ако не и погодни за машинско читање) информации за јавноста има долга историја. Веројатно најдиректно релевантно за она што денес го нарекуваме дата-новинарството е *известувањето со помош на компјутер*, или ИПК, (*computer-assisted reporting*, CAR), кое е прв таков организиран, системски приод кон користењето компјутери за собирање и анализирање на податоците со цел да се подобрат вестите.

ИПК прв пат е користено во 1952 година, од страна на „Си-Би-Ес“ (CBS), за предвидување на резултатите на претседателските избори. Од 1960-тите години наваму, новинарите (претежно истражувачки новинари и претежно во САД) се трудат да спроведат независно следење на носителите на моќта преку анализа на базите на податоци на јавните архиви со научни методи. Познато и како „*јавносервисно новинарство*“, застапниците на таквите компјутерски-помогнати техники имаа за цел да ги откријат трендовите, да ги разобличат јавно прифатените знаења и идеи и да ги откријат неправдите што ги вршат јавните власти и приватните корпорации. На пример, Филип Мејер (Philip Meyer) се обиде да ги разобличи добиените извештаи за немирите во Детроит во 1967 година за да покаже дека во нив не учествувале само послабо образованите Јужњаци. Приказните од серијата „Бојата на парите“ (“The Color of Money”) на Бил Дедман (Bill Dedman) од 80-тите години на минатиот век откриваат како политиките за кредитирање на големите финансиски институции се засноваа на системски расни предрасуди. Во својата статија „Што тргна наопаку“ (“What Went Wrong”), Стив Дојг ги анализираше обрасците на уништување предизвикано од ураганот „Ендрју“ во првата половина на 1990-тите години, со цел да разбере каков ефект имаат неуспешните политики и практики на урбано планирање и развој. Известувањето што произлегува од податоци извршува значајна јавна услуга и за тоа ги има добиено највредните и најпознатите новинарски награди.

Во раните 1970-ти години беше скован терминот *прецизно новинарство* за опишување на следниот тип на собирање вести: „примена на истражувачките методи на општествените и бихевиористичките науки во практикувањето на новинарството“ (цитирано од книгата „*The New Precision Journalism*“ на Филип Мејер, достапна на <http://bit.ly/precision-journalism>). Замислата беше дека со прецизно новинарство ќе се занимаваат професионалци во медиумските институции на средната струја со квалификации во областа на новинарството и општествените науки. Тоа се роди како одговор на „новото новинарство“, форма на новинарство во која известувањето применува техники од пишувањето фикција. Мејер сугерира дека научните техники на собирање и анализа на податоците, а не литературните техники, му се потребни на новинарството за да ја исполни својата цел во потрагата по објективност и вистина.

Прецизното новинарство може да се разбере и како реакција на некои од најчесто наведуваните слаби страни на новинарството: зависноста од официјалните соопштенија за печат (подоцна опишано со терминот „churnalism“, пракса на објавување, без проверка и натамошно истражување, официјални соопштенија или агенциски вести за да се одговори на потребата да се исполнат роковите и да се намалат трошоците), склоноста да им се верува на авторитативните официјални извори итн. Според Мејер, тие слабости произлегуваат од недостигот на примена на техниките на информатичките науки и научните методи какви што се анкетите и користењето на јавните архиви. Во 1960-тите години, прецизното новинарство беше користено за претставување на маргинализираните групи и нивните приказни. Според Мејер:

Прецизното новинарство беше начин за ширење на новинарскиот комплет алатки за да се овозможи новинарски увид во темите кои претходно беа недостапни или достапни само во најгруби црти. Тоа особено беше полезно како начин да се слушнат малцинските и дисидентските групи што се бореа за поголема видливост во јавноста.

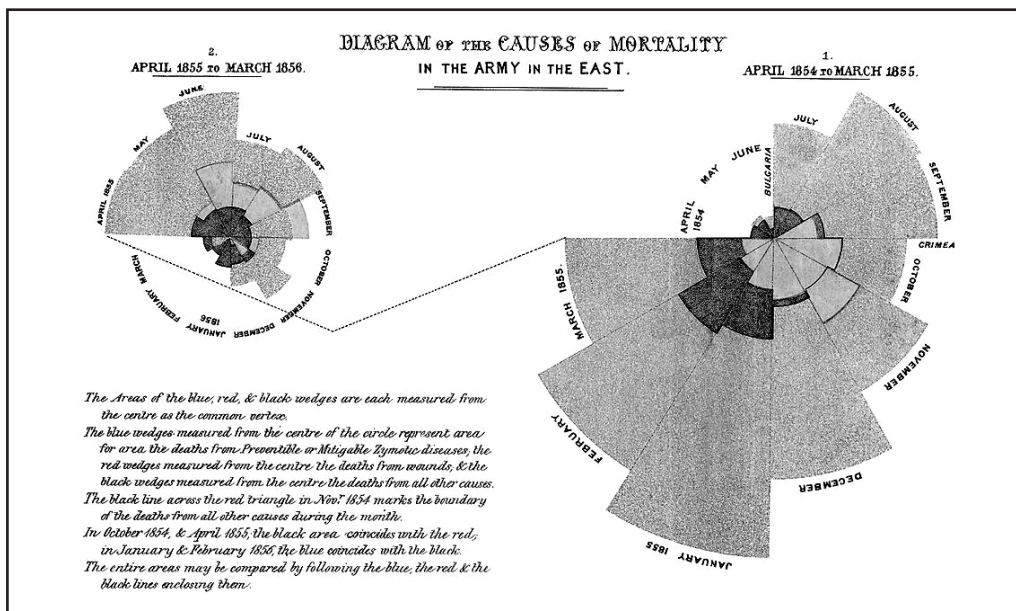
Една влијателна статија за односот помеѓу новинарството и општествениите науки, објавена во 1980-тите години, го одразува денешниот дискурс за новинарството засновано на податоци. Авторите, двајца професори по новинарство од САД, посочуваат дека во 1970-тите и 1980-тите години, сфаќањето на јавноста што се тоа вести се прошири од тесниот концепт за „настани за кои се известува“ на „ситуациско известување“ (или известување за општествените трендови). На пример, преку користење на базите на податоци од пописите и податоците од социолошките истражувања, новинарите можеа да „поминат отаде известувањето за специфични, изолирани настани и да го понудат и контекстот кој им го дава значењето на таквите настани“.

Како што можеше да се очекува, праксата на користење податоци за подобрување на известувањето трае исто толку време колку што постојат податоците. Како што посочува Сајмон Роџерс, првиот пример за дата-новинарство во „Гардијан“ е забележан во 1821 година, кога новинарите објавиле листа на училиштата во Манчестер со бројката на ученици и трошоците на секое училиште поодделно, до која дошле по тајни канали. Според Роџерс, тоа помогнало да се прикаже вистинскиот број на ученици што добиваат бесплатно школување, кој бил многу повисок од официјалните статистики.

Друг ран пример на дата-новинарство во Европа е извонредниот извештај „Стапката на смртност во британската армија“ (“Mortality of the British Army”) на Флоренс Најтингел (Florence Nightingale), објавен во 1858 година. Во нејзиниот извештај до Парламентот, таа користи графикони во застапувањето за подобрувања во санитетот на воените сили на Велика Британија. Најпознат од тие графикони е нејзината „кикиришка“ (“coxcomb”), спирала од секции што ги претставуваат загинатите по месеци, која укажува дека огромно мнозинство од смртните случаи биле резултат на болести што можеле да се спречат, а не на непријателските куршуми.

south end of Edge-hill. In its progress it knocked down several workmen, one of whom was so violently affected in the back of the head, that for relief he had recourse to bleeding; and a girl in a garden had her arm suddenly lifted up by its effect, and felt it bounding for some time after. The coachman of Mr. Duff was struck on the arm whilst on the box, but was merely stunned. A lady near Lichfield, who was sewing at the time, felt the firemen in contact with the needle a sensation resembling that of a slight electric spark. The electric fluid entered the house of Mr. Clare, in Edge Vale, where its progress was not less alarming; to the inmates, than destructive to the premises; and we have never heard of a more surprising escape, than that of the several individuals dispersed in a house, of which almost every room bears testimony of the ravages of the uncontrollable element. It appears probable, from an examination of the aperture which the fluid has made, and the direction in which the bricks, timber, &c. have been forced, that, attracted by the iron railing in front of the house, it entered the wall on one side the door, where it has shattered the bricks, torn to pieces the wood and brick-work between the door-jambs and the arch-way of the door; lifted the boards on the top; shook the fan-light to pieces, burning part of the frame, and leaving a black mark on the paint-work; thence it passed up through the arch door, splitting the bricks and the stone at the bottom of the middle window, the glass of which was shattered to pieces, and the whole frame dislodged and forced into the house. Over the window it forced, in its way to the roof, a large hole, above which the soot appears exactly as if flames had issued from it. Its course appears next to have been towards the chimney: the eaves were shattered to pieces; the ridge stones displaced; many bricks and much cement torn from the wall; and the lead in many places forced. It probably reached the rooms below by the chimneys. In the lower room the stoves, plaster and paper are in several places broken, and the floor, as if searching its way out, scorched the gliding of the chimney glass, and perked the top ornaments, but did not disturb the polished fire-iron, just below. Six squares of glass were driven out in this room. In the room above, in one corner, stood a bundle of rails, to which it made its way, perhaps from the chimney, between the lathing and the wall, as it forced off the plaster, and shot a quantity of it against a chest of drawers, screws or twelve feet distant, evidently with some force.	DAY SCHOOLS.—Establishments				Remarks.	11, June 9, at 10, at Guildhall, London. A Mr. Toms, Tinsell and Bessell, Thurgate street.
	Boys	Girls	Total	Exp.		
Grammar School	155	155	1800	1800	Taught, clothed and boarded.	LEE Jonathan, of Sunderland, in the county Durham, grocer, d. c.; May 15, 19, June 9, 12, at Guildhall, London. Also, Messrs. G. Husban and Ginty, Thurgate-street.
Blue Coat ditto	80	80	2000	2000	Taught and clothed.	FAYN Thomas, and John Daniel Pym, of Canton-street, in the city of London, warehousemen d. c.; May 5, 12, June 9, at 1, at Guildhall, London. Also, Messrs. G. Husban and Ginty, Thurgate-street.
Green Coat ditto	50	50	200	200	And offertory money: do. do.	SMITH John, now of Patrington, in the demesne, in the county of York, grocer, he d. c.; May 11, 12, June 9, at 11, at Dog and Duck Tavern, Kingsdown-on-Hall. Mr. Wainwright, Hull.
Collegiate Church ditto	10	10	100	100	(Suppense)—Taught and clothed.	TATE John, of Liverpool, in the county of Lancashire, provision-merchant, d. c.; May 17, June 9, at 1, at the George Inn, Liverpool. Mr. Denison, Liverpool.
Strageways ditto	10	10	100	100	(Suppense)—Expenses raised by voluntary Subscription.	WARD Joseph, late of Banbury, in the county of Oxford, (but now a prisoner in the King's Bench prison), brewer, d. c.; May 5, 15, June 9, 12, at Guildhall, London. Also, Messrs. Fil and Munday, Holborn.
St. Mary's ditto	12	12	21	40	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
St. John's ditto	9	9	9	40	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
St. Paul's ditto	30	30	20	40	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
Ladies' Jubilee	30	30	20	200	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
Back King-street	21	21	21	40	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
NATIONAL SCHOOLS, Granby-row	194	119	313	600	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
Bolton-street, Salford	200	170	470	600	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
Disasters.	851	381	1232	£5410	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
LANCASTERSHIP SCHOOLS, Marshall-st.	692	225	917	400	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
Widow's, Mole-street	33	32	50	104	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
CATHOLIC SCHOOLS.	198	121	319	104	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
SUNDAY SCHOOLS.	800	581	1271	£554	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
Collegiate Church, Salford	201	205	406	406	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
St. Ann's, Back King-street	50	56	106	106	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
St. Mary's, Back South Parade	130	110	240	240	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
St. Paul's, Green-street	170	163	333	333	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
Turner-street	68	71	139	139	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
Jersey-street	314	281	595	595	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
St. George's, St. John's-street	111	119	230	230	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
St. John's, St. John's-street	118	165	283	283	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
St. James's, St. James's-street	162	198	360	360	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
St. Michael's, Wilkes-street	224	352	576	576	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
St. Peter's, Jackson's-row	120	120	240	240	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
Alport Town	90	90	180	180	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
St. Clement's and St. Luke's	535	1071	1606	1606	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
Demot-street	181	297	478	478	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
St. Stephen's, Bloom-street	120	204	324	324	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
Trinity, King's Head Yard	201	300	501	501	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
Rulme, Duke-street	182	180	362	362	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
All Saints, Oxford road	195	191	386	386	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.
Arldwick	60	110	170	170	(Suppense)—Taught and clothed.	WHARTON Robert and Henry Wharton, of L. Crosby, in the county of Lancashire, joint house-keepers, late coopers in trade, 1, 19, June 9, at 11, at the George Inn, Liverpool. Also, Mr. Hodgson, Liverpool.
	3434	4213	7647	£1079	(Suppense)—Taught and clothed.	WILKINSON Daniel, of Princes-street, Botolph, in the county of Surrey, master-mar merchant, d. c.; May 5, at 10, at Guildhall, London. Messrs. Paterson and Paine, Old Broad-street.

Слика 1-9. Даша-новинарство во „Гардијан“ од 1821 година („Гардијан“)



Слика 1-10. Графиконот на Флоренс Најтинџел за стајката на смртноста во британската армија, (илустрацијата е позајмена од „Википедија“)

Дата-новинарството и известувањето со помош на компјутер

Во моментот се одвива дебата за „континуитетот и промените“ во врска со етикетата „дата-новинарство“ и неговиот однос со претходните новинарски практики за примена на сметачките техники во анализата на збиравите од податоци.

Некои луѓе тврдат дека постои разлика помеѓу ИПК и новинарството засновано на податоци. Тие велат дека ИПК е техника на собирање и анализирање податоци како начин за засилување на известувањето (вообичаено се мисли на истражувачкото известување), додека дата-новинарството е фокусирано на позицијата на податоците во целокупната новинарска работа. Во таа смисла, новинарството засновано на податоци обрнува поеднакво – а понекогаш и поголемо – внимание на самите податоци, а не само на користењето на податоците како средство за пронаоѓање или подобрување на приказните. Од тие причини „Датаблогот на Гардијан“ (Guardian Datablog) или „Тексас трибјун“ ги објавуваат комплетите податоци заедно со приказните – а понекогаш и комплетите податоци самостојно – за да можат луѓето и сами да ги анализираат и истражуваат.

Друга разлика е во тоа што во минатото, истражувачките новинари страдаа од сиромаштвото на информации поврзани со некое прашање на кое се обидуваа да одговорат или со проблем што се обидуваа да го решат. Тоа, се разбира, сè уште е така, но сега постои и неизмерно изобилие од информации, при што новинарите не мора нужно да знаат што да прават со нив.

Тие не знаат како да го извлечат она што е вредно од податоците. Еден таков неодамнешен пример е „Комбинираниот онлајн информатички систем“ (Combined Online Information System - COINS), најголемата база на податоци за јавните трошоци во Велика Британија. Заговарачите на транспарентноста долго бараа пристап до базата на податоци која, по објавувањето, ги збуни повеќето новинари. Како што неодамна ми пиша Филип Мејер: „Порано, кога информациите беа скудни, најголемиот дел од нашите напори беа посветени на „лов и собирање“. Денес, кога имаме изобилство од информации, нивната обработка е од многу поголемо значење“.

Од друга страна, има такви што тврдат дека не постои никаква значајна разлика помеѓу дата-новинарството и известувањето со помош на компјутер. Денес се смета за општо прифатено дека дури и најновите медиумски практики, иако носат нешто ново, си имаат своја предисторија. Наместо да расправаме дали дата-новинарството е новитет или не, далеку поплодно би било да го разгледаме како дел од една подолготрајна традиција, но во однос со новонастанатите околности и услови. Дури и ако навистина нема разлика во нивните цели и во техниките со кои се служат, појавата на етикетата „дата-новинарство“ во почетокот на ова столетие посочува на појавата на една нова фаза во која само обемот на информации што слободно се достапни на интернет – комбинирани со софистицирани алатки фокусирани на корисниците, автопубликацијата и алатките за „краудсорсинг“ („crowdsourcing“, пракса на собирање информации на едно место, при што задачата се делегира на поголема група луѓе) – овозможува повеќе луѓе да работат со повеќе податоци на полесен начин од кога било претходно.

Дата-новинарството е поврзано со масовната информатичка писменост

Дигиталните технологии и глобалната мрежа од корен го менуваат начинот на објавување на информациите. Дата-новинарството е дел од екосистемот на алатки и практики што се појавија со појавата на веб-сајтовите и услугите што работат со податоци. Цитирањето и размената на изворни материјали е дел од природата на хиперлинкуваната структура на Мрежата и е начин на кој денес сме навикнати да се снаоѓаме низ информациите. Ако се вратиме наназад, принципот во основата на хиперлинкуваната структура на Мрежата е пандан на принципот на цитирање што го користат академските трудови. Цитирањето и размената на изворните материјали и податоците на кои се заснова приказната е еден од основните начини на кои новинарството засновано на податоци може да го подобри новинарството воопшто, она што основачот на „Викиликс“ Џулијан Асанж го нарекува „научно новинарство“.

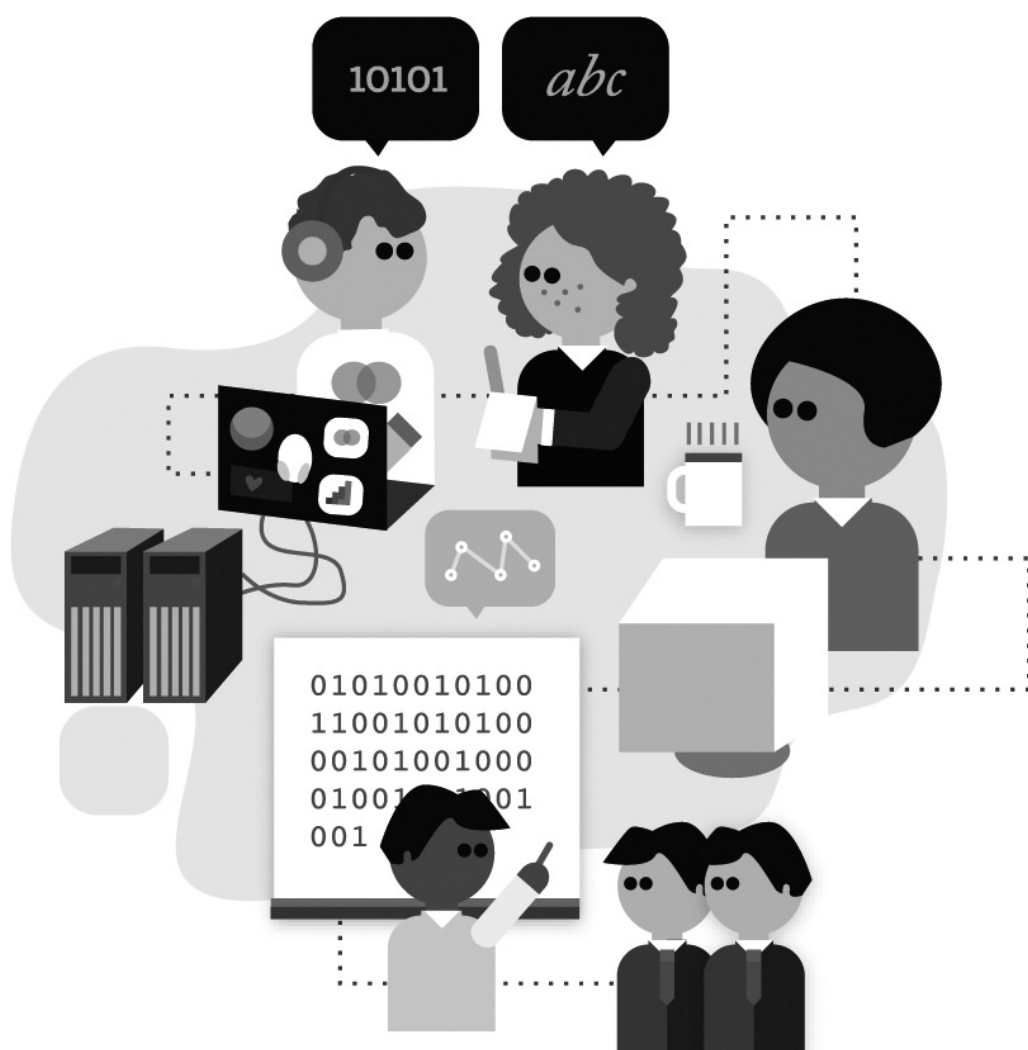
Со тоа што им овозможува на сите да ријат по изворите на податоци и да ги пронајдат информациите што се значајни за нив, како и да ги потврдат или оспорат општоприфатените претпоставки, дата-новинарството во суштина претставува масовна демократизација на ресурсите, алатките, техниките и методологиите што претходно беа користени само од специјалисти во својата област, без оглед дали се работи за истражувачки новинари, научници од областите на општествените науки, статистичари, аналитичари или други стручни лица. Иако во моментот цитирањето и линкувањето до изворите на податоци е особина на дата-новинарството, светот се движи во насока во која податоците се целосно интегрирани во текстурата на медиумите. Новинарите што се занимаваат со дата-новинарство играат значајна улога во обидот да се елиминираат бариерите за разбирањето и работењето со податоците и масовно ја зголемуваат писменоста во користењето на податоците на нивните читатели.

Во овој момент постои една нова заедница од луѓе што се нарекуваат „дата-новинари“ и во голема мера се разликуваат од постарата и позрела заедница на луѓе што се занимаваат со известување со помош на компјутер. Надежта е дека во иднина ќе видиме поцврсти врски помеѓу овие две заедници, на помалку или повеќе ист начин на кој денес ја гледаме соработката на невладините и граѓанските медиумски организации какви што се „ПроПублика“ или „Бирото за истражувачко новинарство“ (Bureau of Investigative Journalism) со традиционалните информативни медиуми во спроведувањето истражувачки проекти.

Иако заедницата на практичари на полето на дата-новинарството поседува поиновативни начини за претставување на податоците и пренесување на приказните, дата-новинарството сигурно може да научи многу нешта од длабоко аналитичниот критички пристап на заедницата што се занимава со ИПК.

— Лилијана Бунеџру, Европски центар за новинарство

Во редакцијата



Каква е положбата на дата-новинарството внатре во редакциите во светот? На кој начин водечките дата-новинари ги убедуваат своите колеги дека објавувањето зборови податоци или лансирањето информативни апликации засновани на податоци е добра идеја? Треба ли новинарите да учат да програмираат или е подобро да работат во тандем со талентирани девелопери на софтверски решенија? Во ова поглавје ќе ја разгледаме улогата што податоците и дата-новинарството ја имаат во Радиодифузната корпорација на Австралија – Еј-Би-Си (Australian Broadcasting Corporation, ABC), Би-Би-Си (BBC), „Чикаго трибјун“ (Chicago Tribune), „Гардијан“ (the Guardian), „Тексас трибјун“ (the Texas Tribune) и во онлајн-изданието на „Цајт“ (Zeit Online). Ќе научиме како да откриеме и да ангажираме квалитетни девелопери, како да ги натераме луѓето да се ангажираат на некоја тема преку организирање „хакатони“ (hackathon - маратон во програмирање) и други настани, како да соработуваме преку државните граници, како и за развивањето нови бизнис-модел за поддршка на дата-новинарството.

Активностите на Еј-Би-Си на полето на дата-новинарството

Австралиската радиодифузна корпорација е национален јавен радиодифузер во Австралија. Со годишен буџет од околу една милијарда австралиски долари, Еј-Би-Си одржува седум радиомрежи, 60 локални радиостаници, три дигитални телевизиски сервиси, еден нов меѓународен телевизиски сервис и онлајн-платформа на која оди се побогатата понуда на дигитални содржини и содржини произведени од корисниците. Според последните достапни податоци, корпорацијата има повеќе од 4.500 вработени со полно работно време, од кои околу 70 проценти работат на производство на содржини.

Ние сме национален радиодифузер и горди сме на нашата независност; иако сме финансирани од владата, законски сме одвоени од неа на пристојна дистанца. Ја следиме традицијата на независното јавно-сервисно новинарство. Еј-Би-Си се смета за информативна организација со најголема доверба на публиката во Австралија.

Живееме во возбудливи времиња. Под водството на директорот Марк Скот (Mark Scott), кој има претходно искуство во раководењето со печатени медиуми, производителите на содржини во Еј-Би-Си беа охрабрани да бидат „агилни“, еден од омилените зборови на корпоративската мантра.

Се разбира, тоа е лесно да се каже, но не и толку лесно да се оствари.

Сепак, неодамна имавме една иницијатива за која требаше да го охрабри таквиот став – натпреварувачки процес во кој од вработените се бараше да ги „продадат“ своите идеи за мултиплатформски проекти за да добијат средства за нивен развој.

Така се зачна и првиот проект за дата-новинарство во Еј-Би-Си.

Некаде во почетокот на 2010 година, и самиот влегов во една таква сесија на предлагање идеи за да го претставам мојот предлог пред тројца виши службеници „задолжени за нови идеи“.

Ја прецвакував идејата подолго време, алчно голтајќи го дата-новинарството што го нудеше денес легендарниот „Датаблог на Гардијан“ (а тоа да ми ти било само предавање, како што сфатив подоцна).

Настапив со тврдење дека воопшто не се сомневам дека, во наредните 5 години, Еј-Би-Си ќе има свој сопствен оддел за дата-новинарство. Сметав дека тоа е неизбежно. Прашањето беше како ќе стигнеме таму и кој ќе го направи првиот чекор.

За оние читатели што не знаат ништо за Еј-Би-Си, замислете си една огромна бирократија градена повеќе од 70 години. Радиото и телевизијата отсекогаш биле нејзината основна понуда. По воспоставувањето на веб-страницата, во последната деценија понудата на содржини се прошири на текст, фотографија и степен на интерактивност што дотогаш не можевме ни да го замислиме. Интернет-просторот ја принуди Еј-Би-Си повторно да размисли како го дели колачот (парите) и каков колач (содржина) сака да испече.

Се разбира, тоа е процес кој тече.

Со дата-новинарството ситуацијата беше малку поинаква. Програмата Влада 2.0 (Government 2.0) (откривме дека во Австралија таа почесто се крши отколку што се почитува) понуди нови начини за пренесување приказни што дотогаш беа затрупани во купишта единици и нули.

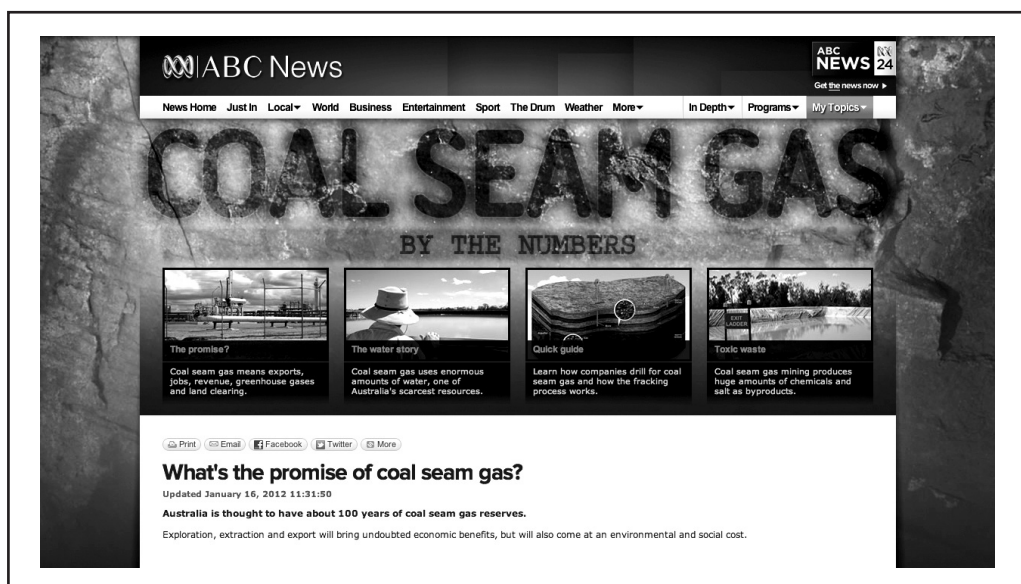
Сето тоа им го кажав на луѓето што присуствуваа на претставувањето на мојот предлог. Спомнав и дека ќе треба да ги идентификуваме новите вештини што ќе ни требаат и да ги обучиме новинарите за работа со новите алатки. Ни требаше само некаков проект за да го притиснеме копчето „Плеј“.

Ми ги дадоа парите што ги барав.

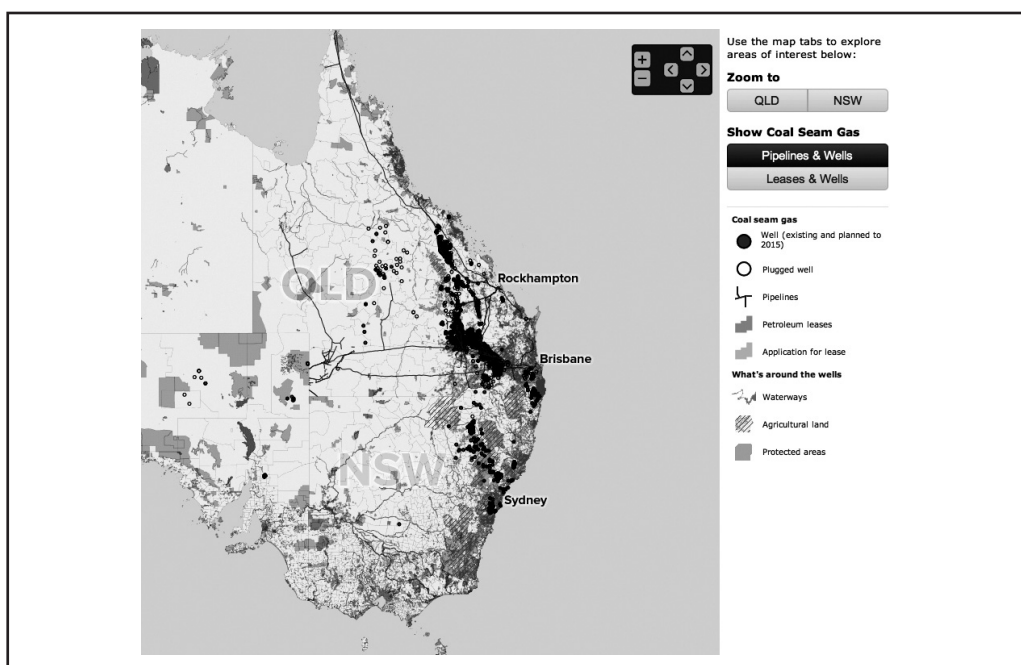
Мултиплатформскиот проект на Еј-Би-Си и „Еј-Би-Си њуз онлајн“ (ABC News Online) официјално стартуваше на 24 ноември 2011 година, со сторијата „Фактите за резервите на гас во наоѓалиштата на јаглен“ (Слика 2-1) (“Coal Seam Gas by the Numbers”).

Сторијата се состоеше од пет страни со текст, интерактивни мапи и визуелизации на податоци.

Не се работеше исклучиво за дата-новинарство, туку за хибридно новинарство што произлезе од миксот на луѓе во тимот и од самата приказна, која и понатаму останува едно од најгорливите прашања во Австралија.



Слика 2-1. „Фактиите за резервиите на гас во наоѓалиштите на јаглен“ (Еј-Би-Си њуз онлајн)



Слика 2-2. Интерактивна мапа на гасните полиња и дозволиите за експлоатација во Австралија (Еј-Би-Си њуз онлајн)

Врв на сè беше интерактивната мапа (Слика 2-2) на која беа прикажани резервите на гас во рудниците со јаглен и издадените дозволи за експлоатација во Австралија. Корисниците можеа да пребаруваат по локација и да преминуваат од приказот на резервите на приказот на издадените дозволи. Корисниците можат да ја зумираат мапата за да видат кој го вршел истражувањето, каков е статусот на лежиштето и датумот на кој почнало дупчењето. На друга мапа беа прикажани локациите на кои се врши истражување на гасните резерви заедно со локациите на подземните води во Австралија.

Имавме и посебни визуелизации за прашањето на солта која се јавува како нус-производ, како и за можностите за производство на вода, во зависност од тоа кое сценарио ќе биде избрано.

Друг дел од проектот го истражуваше прашањето на испуштање хемикалии во сливот на една локална река.

Нашиот тим

- Веб-девелопер и веб-дизајнер
- Новинар што раководеше со тимот
- Истражувач, ангажиран на определено време, со експертиза на полето на извлекување податоци, работа со „Ексел“ (Excel) и прочистување на податоците
- Помлад новинар ангажиран како хонорарец
- Извршен продуцент што го користевме како консултант
- Еден академски консултант со експертиза на полето на извлекување податоци, графичка визуелизација и напредни истражувачки вештини
- Ги користевме услугите на еден проектен менаџер, а во администрацијата ни помогна мултиплатформскиот оддел на Еј-Би-Си
- Значајно е да се каже дека имавме на располагање и една референтна група новинари и други лица што ги консултиравме по потреба

Од каде ги добивме податоците?

Податоците за интерактивните мапи беа „стругани“ (“scraped”) од датотеките со геодетски податоци (“shapefiles”, со наставка “.shp”) спуштени од владините интернет-страници.

Податоците за солта и водата беа извадени од неколку различни извештаи.

Податоците за испуштање хемикалии беа извадени од еколошките дозволи издадени од властите.

Што научивме?

„Фактите за резервите на гас во наоѓалиштата на јаглен“ беше амбициозен проект, како по својата содржина, така и по својот обем. Според мене, најважно беше да видиме што сме научиле од проектот и што би можеле да направиме на поинаков начин во следниот проект?

Проектот за дата-новинарство собра во една канцеларија многу луѓе што во нормални услови не би се сретнале во Еј-Би-Си: лаички кажано, „пискарала“ (hack, жаргонски назив за новинар, заб.прев.) и хакери. Многумина од нас практично не говореа ист јазик и дури воопшто не го ценеа она што го работи другата група. Статистичкото новинарство ги разорува старите навики!

Следат неколку практични нешта што ги научивме:

- Собирањето на тимот на едно место е од клучно значење. Нашиот девелопер и нашиот дизајнер работеа на друга локација и доаѓаа кај нас само на состаноци. Тоа дефинитивно не е оптималното решение! Сместете ги сите во иста просторија со новинарите.
- Извршниот продуцент што ја имаше улогата на консултант се наоѓаше на сосема друг кат од зградата. Требаше да бидеме сместени многу поблиску, ако ништо друго, поради можноста понекогаш да наминеме едни кај други.
- Изберете приказна којашто е водена единствено од податоците.

Големата слика: Неколку идеи

Големите медиумски организации треба да преземат активности за градење на капацитетите потребни за справување со предизвиците што со себе ги носи дата-новинарството. Јас мислам дека во техничките одделенија на медиумите работат многу „бубалици“ и хакери што едвај чекаат некој да им чукне на врата. Значи, ни требаат работилници на кои ќе ги споиме „пискаралата и хакерите“, место каде што прикриените „бубалици“, помладите новинари, веб-девелоперите и дизајнерите ќе си играат, а поискусните новинари ќе ги споделуваат своите вештини и ќе ја вршат улогата на ментори. Задачата е: сними ги податоците и факај се за работа!

Дата-новинарството по природа е интердисциплинарно. Тимовите што се занимаваат со дата-новинарство се составени од луѓе што претходно не би работеле заедно. Дигиталниот простор, од своја страна, ги брише границите.

Живееме во едно фрагментирано политичко тело кое не верува никому. Бизнис-моделот што порано обезбедуваше професионално и независно новинарство – со сите негови несовершености – се наоѓа на работ на колапсот. Би требало да се прашаме, како што многумина веќе се прашуваат, како би

изгледал светот без одржлива четврта сила? Американскиот новинар и интелектуалец Волтер Липман (Walter Lippman) забележал во 1920-тите години дека е „општо прифатено мислењето дека здраво јавно мнение не е можно без пристап до вестите“. Неговата изјава е исто толку вистинита и денес. Во 21-иот век, сите „висат“ во блогосферата. Тешко е да се разликуваат професионалните новинари од спинерите, лажливците, симулантите и интересните групи. Безмалку сите веб-страници или извори можат да се направат да изгледаат веродостојни, измазнети и чесни во намерите. Импресумите што ни влеваа толку многу доверба ги снемуга, оставени да умрат во некој ендек. А во овој нов простор на „џанк-новинарство“, на брзо новинарство, хиперлинковите можат до бескрај да го водат читателот кон други, уште побескорисни извори што изгледаат одлично и го заработуваат хиперлинкувањето во дигиталната соба со сидови од огледала. Техничкиот термин за тоа е: глупостите заглупуваат (изворно на англиски „bullshit baffles brains“, заб.прев.).

Во дигиталниот простор сите се раскажувачи, така? Е, не е така. За да преживее професионалното новинарство – под што подразбирам раскажување што ги прифаќа принципите на етичка, урамнотежена и храбра потрага по вистината – тогаш занаетот мора да се наметне и во дигиталниот простор. Дата-новинарството е само уште една од алатките што ќе ги користиме за снаоѓање во дигиталниот простор. Таму ќе правиме мапи, ќе превртуваме, одвојуваме, филтрираме, вади-ме, за да ја пронајдеме приказната среде сите тие нули и единици. Во иднина ќе работиме рака под рака со хакерите, девелоперите, дизајнерите и програмерите на компјутерски код. Таквата транзиција бара сериозен пристап кон градењето на капацитетите. Ни требаат директори на вести што разбираат како се поврзани дигиталното и новинарството за да започнат да инвестираат во изградбата.

– Венди Карлајл, Австралиска радиодифузна корпорација (Wendy Carlisle, Australian Broadcasting Corporation)

Дата-новинарството во Би-Би-Си

Поимот „дата-новинарство“ може да покрива цел дијапазон дисциплини и се користи на различни начини во новинските организации, така што можеби ќе помогне ако дефинираме што подразбираме ние, во Би-Би-Си, под „дата-новинарство“. Општо земено, поимот се однесува на проектите што користат податоци за една од следните цели:

- овозможување на читателите да ги дознаат информациите што се релевантни лично за нив;
- откривање приказна што е значајна, а претходно не била позната;
- помагање на читателите подобро да разберат некое комплексно прашање.

Овие категории можат да се преклопуваат, а во онлајн-опкружувањето често имаат корист од определено ниво на визуелизација.

Дадете му лична нота

Интернет-страницата на „Би-Би-Си њуз“ (BBC News, <http://www.bbc.com/news/>), веќе подолго од една деценија користи податоци за да им понуди услуги и алатки на корисниците.

Најпостојан пример, за прв пат објавен во 1999 година, е нашата „табела на пласманот на лигата на училишта“ (school league tables), која ги користи податоците што властите ги објавуваат на годишно ниво. Читателите можат да ги пронајдат локалните училишта преку внесување на поштенскиот број на некоја географска област и да направат споредба меѓу нив врз основа на повеќе показатели. Новинарите што го следат образованието, пред објавувањето на табелата, исто така соработуваат со тимот од девелопери во барањето приказни во податоците.

Кога почнувавме со таа работа, на интернет немаше официјална страница што на јавноста ѝ даваше можност да ги погледне тие податоци. Денес, кога и Министерството за образование нуди сопствена слична услуга, ние го префрливме фокусот на нашата понуда на приказните што произлегуваат од податоците.

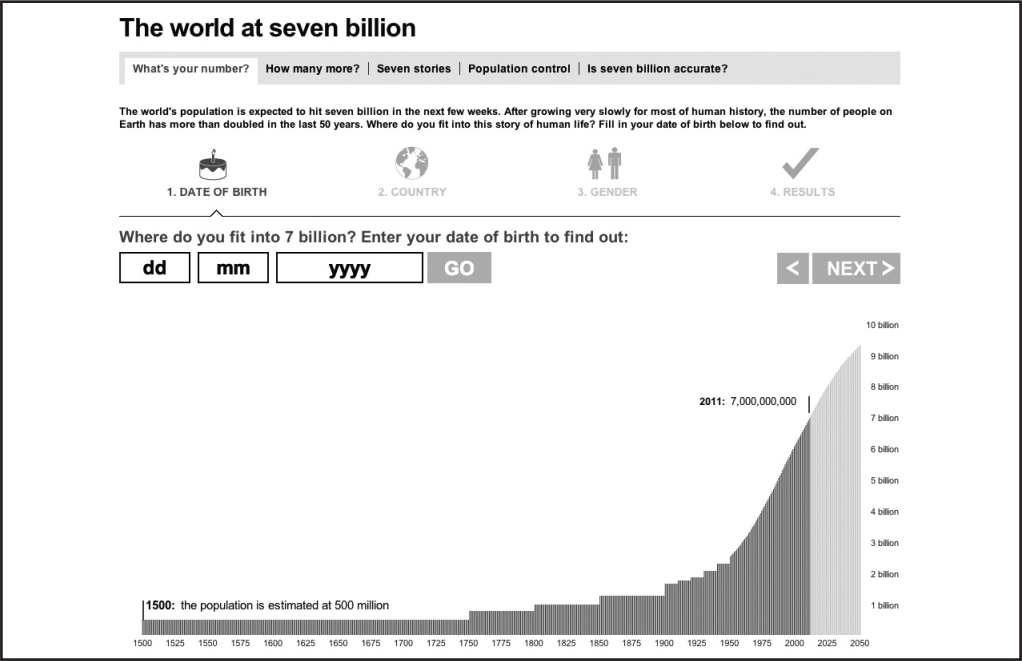
Обезбедувањето податоци за кои постои очигледен интерес во јавноста мора да биде главната задача на тоа поле. Специјалниот извештај „Сите смртни случаи на сите патишта“ (“Every death on every road”) е еден скорешен пример на проект преку кој објавивме една огромна група на податоци што во нормални услови не се достапни за пошироката јавност. Обезбедивме можност за пребарување по поштенски број, така што корисниците можат да ги пронајдат локациите на сите смртни случаи што се случиле на патиштата во Велика Британија во последната деценија.

Визуелизиравме некои од главните факти и бројки што ги извадивме од податоците што ни ги даде полицијата, а за да му дадеме на проектот повеќе динамика и човечки лик, се здруживме со „Лондонската асоцијација на служби за итна помош“ (London Ambulance Association) и со радиото и телевизијата „Би-Би-Си Лондон“ (BBC London) за да ги следиме сообраќајните несреќи во главниот град како што се случуваат. Сето тоа беше објавувано на интернет, во живо, како и преку „Твитер“ (Twitter) со „хаштаг“ (hashtag, тараба, заб.прев.) „#crash24“, а судирите беа внесувани на географска мапа веднаш штом некој известеше за нив.

Едноставни алатки

Освен што обезбедуваме истражување на големи групи на податоци, имавме успех и на полето на создавање едноставни алатки за корисниците за обезбедување информации што се релевантни за нив. Таквите алатки се наменети за оние што немаат време да се внесат во пообемни анализи. Способноста за лесно споделување некој личен факт е нешто што го вградивме во нашите стандарди.

Еден „порелаксиран“ пример за таквиот приод е сторијата „Светот одбројува 7 милијарди: Кој е вашиот број?“ (“The world at 7 billion: What’s your number?”), чие објавување се поклопи со денот на кој светот и официјално ја надмина бројката од 7 милијарди жители. Со внесување на денот на раѓање, корисниците можеа да откријат кој бил нивниот „реден број“ во глобалната популација во моментот на раѓањето и да го споделат тој број на „Твитер“ или на „Фејсбук“ (Facebook). Апликацијата користеше податоци што ги добивме од Фондот за развој на населението на ОН. Сторијата беше многу популарна и стана најмногу споделуваниот линк на „Фејсбук“ во Велика Британија во 2011 година.



Слика 2-3. Светот одбројува седум милијарди (Би-Би-Си)

Друг скорешен пример е буџетскиот калкулатор на Би-Би-Си (BBC Budget calculator), кој им овозможува на корисниците да откријат колку подобра или полоша ќе биде нивната ситуација по стапувањето на сила на буџетот на Канцеларот за трезор (официјалното име на министерот за финансии на Велика Британија) – и потоа да ја споделат таа бројка. За таа цел соработувавме со сметководствената фирма „КПМГ ЛЛП“ (KPMG Limited Liability Partnership), која ни ги даде пресметките засновани на годишниот буџет, а ние потоа напорно работевме за да создадеме атрактивен интерфејс што ќе ги охрабри корисниците да ја извршат задачата.

„Копање“ по податоците

Каде е новинарството во целата таа приказна? Една потрадиционална дефиниција вели дека дата-новинарството претставува пронаоѓање приказни во чистите податоци. Дали во базата на податоци пред вас има некаква ексклузивна приказна? Дали бројките се точни? Дали потврдуваат или оспоруваат некаков проблем? Секој новинар што се занимава со дата-новинарство или со известување со помош на компјутер мора да си ги постави тие прашања. Сепак, може да се потроши многу време во просејување на податоците од некоја масивна група во надеж дека ќе откриеме нешто значајно.

Заклучивме дека најпродуктивен приод е влегувањето во партнерство со истражувачки тимови или програми што ја поседуваат потребната експертиза и имаат време за истражување некоја приказна. „Панорама“ (Panorama), актуелно-политичката емисија на Би-Би-Си, работеше неколку месеци, заедно со „Центарот за истражувачко новинарство“, на собирање податоци за платите во јавниот сектор. Резултат од тој потфат се еден ТВ-документарец и еден специјален извештај на интернет, „Платите во јавниот сектор во броеви“ (“Public Sector pay: The numbers”), каде сите податоци беа објавени, визуелизирани и анализирани за секој сектор засебно.

Како и кај партнерството со истражувачки новинари, од клучно значење е да имате на располагање повеќе новинари што имаат специјализирани познавања. Кога еден член на тимот – колегата инаку покрива економија – ги анализираше податоците за кретењето на потрошувачката објавени од Владата, заклучи дека Владата ги преувеличува тие податоци. Како резултат ја имавме ексклузивната приказна „Како да разбереме што велат податоците“ (“Making sense of the data”), проследена со јасна визуелизација, што ја доби наградата на Кралското статистичко друштво (Royal Statistical Society).

Разбирање на проблемот

Кај дата-новинарството не мора да се работи само за ексклузивни приказни што никој друг не ги забележал. Задачата на тимот за визуелизација на податоците е да го искомбинира добриот дизајн со јасна новинарска приказна за да добие корисникот искуство на кое нема да може да му одолее. Интересните визуелизации на вистинските податоци можат да се искористат за да се обезбеди полесно разбирање на некое прашање или приказна, па ние во Би-Би-Си често го користиме тој приод во пренесувањето на приказните. Една техника што ја користевме во нашиот бројач на побарувања во Велика Британија е да ги претставиме на разнобојна мапа (слична на метеоролошките мапи) на која јасно се гледаат промените што настанале со текот на времето (<http://bbc.in/KF7IKU>).

Фичерот „Мрежа на задолженост во Еврозоната“ (“Eurozone debt web”) ја истражува замрсената мрежа на кредитирање внатре во поединечните др-

жави-членки. Секогаш е добро да се понуди визуелно објаснување на некое комплицирано прашање, со користење бои и пропорционални стрелки, во комбинација со јасен и некомплицирани текст. Значајно е да се има предвид потребата да се охрабри корисникот да го разгледа „фичерот“ или да ја следи приказната, без да се почувствува надвлдаеан од бројките.

Преглед на тимот

Тимот што се занимава со дата-новинарство за потребите на интернет-страницата „Би-Би-Си нюз“ (BBC News) има 20 членови – новинари, дизајнери и девелопери.

Освен проекти што работат со податоци и визуелизации, тимот ги произведува и сите инфографики и интерактивни мултимедијски фичери на информативниот веб-сајт. Заедно, тие формираат збирка од раскажувачки техники што ги нарекуваме визуелно новинарство (visual journalism). Кај нас нема луѓе чие работно место гласи „дата-новинар“, туку од сите новинари и уредници што се дел од тимот се бара да ги поседуваат основните вештини за работа со апликации за сметководствени табели какви што се „Ексел“ (Excel) и „Гугл док“ (Google Docs) во анализата на податоците.

Централно прашање за сите проекти што работат со податоци се техниките вештини и советодавната улога на нашите девелопери, како и вештините за визуелизација на нашите дизајнери. Иако секој од нас на прво место е новинар, дизајнер или девелопер, ние продолжуваме напорно да работиме за да ги подобриме нашето разбирање и вештините на полето на експертиза на другите членови на тимот.

Основни алатки за обработка на податоците се „Ексел“, „Гугл док“ и „Фјуџн тејблс“ (“Fusion Tables”). Тимот, иако во помала мера, ги користеше и базите на податоци „МајЕсКјуЕл“ (MySQL) и „Аксес“ (Access), а и „Солр“ (Solr) за истражување поголеми групи податоци, како и „RDF“ и „SPARQL“ како основа во барањето начини за моделирање на настаните со користење “Linked Data”-технологии. Девелоперите, исто така, се служат со програмскиот јазик што го претпочитаат, сеедно дали се работи за „Екшнскрипт“ (ActionScript), „Питон“ (Python) или „Перл“ (Perl), за споредување, анализа или разложување на групата на податоци со која работиме. „Perl“ делумно го користиме и во дел од изданијата.

За истражување и визуелизација на географски податоци ги користиме мапите на „Google“, „Bing Maps“ и „Google Earth“, заедно со „ArcMAP“ на „Esri“.

За графичките решенија го користиме пакетот апликации „Adobe Suite“, вклучително и „After Effects“, „Illustrator“, „Photoshop“ и „Flash“, иако денес сè поретко објавуваме прикази во „Flash“ на интернет-страницата, затоа што „JavaScript“ – а особено „jQuery“ и други „JavaScript“-библиотеки, какви што се „Highcharts“, „Raphael“ и „D3“ – подобро ги исполнуваат нашите потреби за визуелизација.

– Бела Харел и Ендрју Лајмдорфер, Би-Би-Си (Bella Hurrell and Andrew Leimdorfer, BBC)

Како функционира тимот за новински апликации на „Чикаго трибјун“

Тимот за новински апликации на „Чикаго трибјун“ е една весела банда хакери што работи во составот на редакцијата. Соработуваме со уредниците и новинарите за да помогнеме во: 1) истражувањето и пренесувањето на сториите, 2) илустрирањето на сториите на интернет и 3) градењето на секогаш свежи и нови онлајн-ресурси за добрите луѓе од „Чикаголенд“.

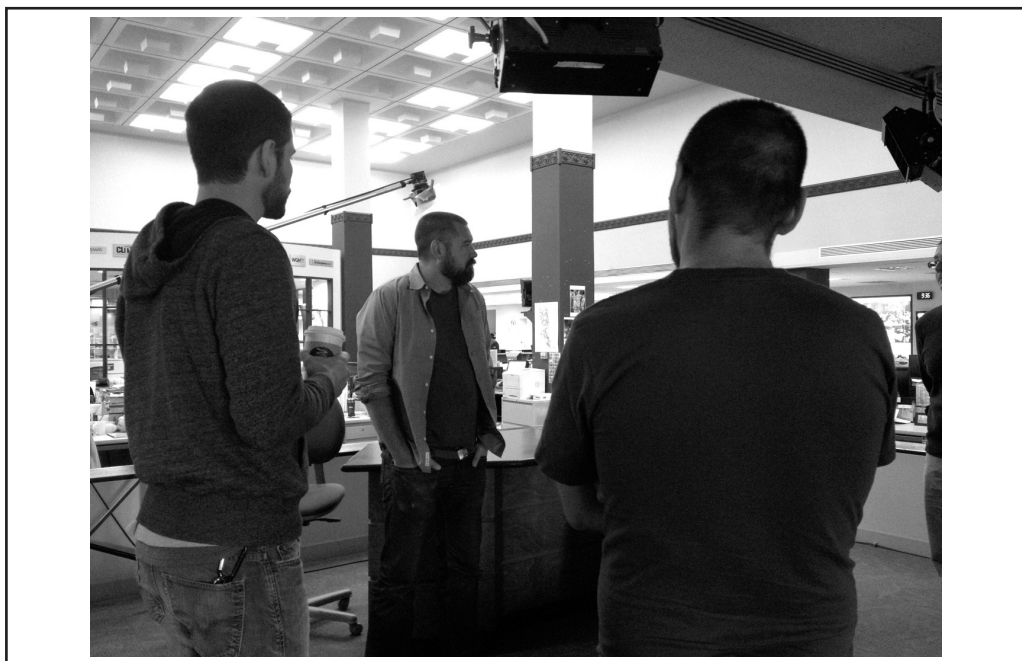
За нас е важно што сме дел од редакцијата. Обично си наоѓаме работа во директен разговор со новинарите. Тие знаат дека за нас претставува задоволство кога можеме да им помогнеме во пишувањето на „стругач на екрани“ (“screen scraper”) за некој „склепан“ владин веб-сајт, да „исцераме“ цел куп „PDF“ документи или на друг начин да ги трансформираме податоците во нешто што потоа може да се анализира. На некој начин, се работи за „промотивен попуст“ за нашиот тим; на тој начин дознаваме за потенцијалните проекти за работа со податоци на самиот почеток.

За разлика од повеќето тимови што се занимаваат со оваа работа, нашиот беше основан од технолози за кои влегувањето во новинарството значеше промена на кариерата. Некои од нас стекнаа магистратура по новинарство по неколку години поминати во пишување код за комерцијални потреби, а други дојдоа од заедницата што се залага за отворено владеење.

Многу сме агилни. За да се осигураме дека секогаш сме на иста линија, секое утро почнуваме со 5-минутен состанок „од нога“. Често програмираме во парови; двајца девелопери на една тастатура често се попродуктивни од двајца девелопери што работат на две тастатури. За повеќето проекти, доволна е една седмица да произведеме резултати, но кај подолготрајните проекти работиме во итерации од по една седмица и на секои седум дена им го покажуваме дотогаш сработеното на заинтересираните страни (обично се работи за новинари и уредници). Нашата мантра е: „Ако нешто работиш погрешно, откриј го тоа што порано“. Ако работиш погрешно, треба тоа да го знаеш што е можно порано, особено ако треба да го завршиш кодот до определен рок!

Повторливото хакирање, со рокови, има една огромна предност: постојано ги унапредуваме и подобруваме нашите алатки. Секоја недела ќе направиме една или две нови апликации и тогаш, за разлика од нормалните софтверски работилници, можеме да ги оставиме настрана и да преминеме на следниот проект. Супер е што разменуваме информации со новинарите и секоја недела учиме нешто ново.

Сите идеи за нови апликации потекнуваат од новинарите и уредниците во редакцијата. Тоа, мислам, нè разликува од другите тимови за апликации во другите редакции кои често излегуваат со свои сопствени идеи. Изградивме цврсти лични и професионални односи во редакцијата и луѓето таму знаат дека кога имаат податоци, можат да ни се обратат.



Слика 2-4. Тимот за новински апликации на „Чикаго трибјун“ (фотографија: Хедер Билингс)

Поголемиот дел од нашата работа во редакцијата е да им даваме поддршка на новинарите. Им помагаме да копаат низ податоците, ги претвораме „PDF“-датотеките во сметководствени табели, „стругаме“ интернет-страници итн. Тоа е услуга што радо ја даваме затоа што ни овозможува мошне рано да добиеме увид во работата со податоци што се врши во редакцијата. Дел од таа работа резултира во некаква новинска апликација: мапа, табела, а понекогаш дури и цела веб-страница.

Порано, апликацијата ја врзувавме за пишаната сторија, но тоа не ни носеше многу сообраќај. Денес апликациите се поставени на врвот на нашата веб-страница, или блиску до врвот, а апликацијата води до сторијата, што е добро решение и за апликацијата и за сторијата. Имаме и посебна секција на веб-страната за тоа што го работиме, но таа нема многу посетители. Не сме изненадени што е така. Колку често слушате некој да каже: „Е, денес баш би сакал да разгледам некои податоци!“

Сакаме кога ги посетуваат нашите страници. Секако, ни се допаѓаат и профалбите од нашите колеги, но од тоа не се живее. Мотивацијата секогаш треба да ја барате во влијанието што го имате; врз животите на луѓето, врз законите, во повикувањето на политичарите на одговорност итн. Пишаната статија ќе се

осврне на трендот и ќе му даде човечки лик со неколку анегдоти. Но, што да прави читателот откако ќе ја прочита приказната? Дали неговото семејство е безбедно? Дали неговите деца добиваат соодветно образование? Нашата работа е песна ако им помага на читателите да ја откријат сопствената приказна во податоците. Примери за персонализирано дело што имало влијание се и нашите апликации „Извештаи за безбедноста во домовите за целодневна грижа“ (Nursing Home Safety Reports) и „Училишни свидетелства“ (School Report Card).

– Брајан Бојер, Чикаго Трибјун

Зад сцената во „Датаблогот на Гардијан“

Кога го лансиравме „Датаблогот“ (Datablog), немавме поим кој би можел да се интересира за суровите податоци, статистики и визуелизации. Како што рече еден човек од мојата канцеларија, на прилично висока позиција: „Што ќе му е тоа на кого било?“

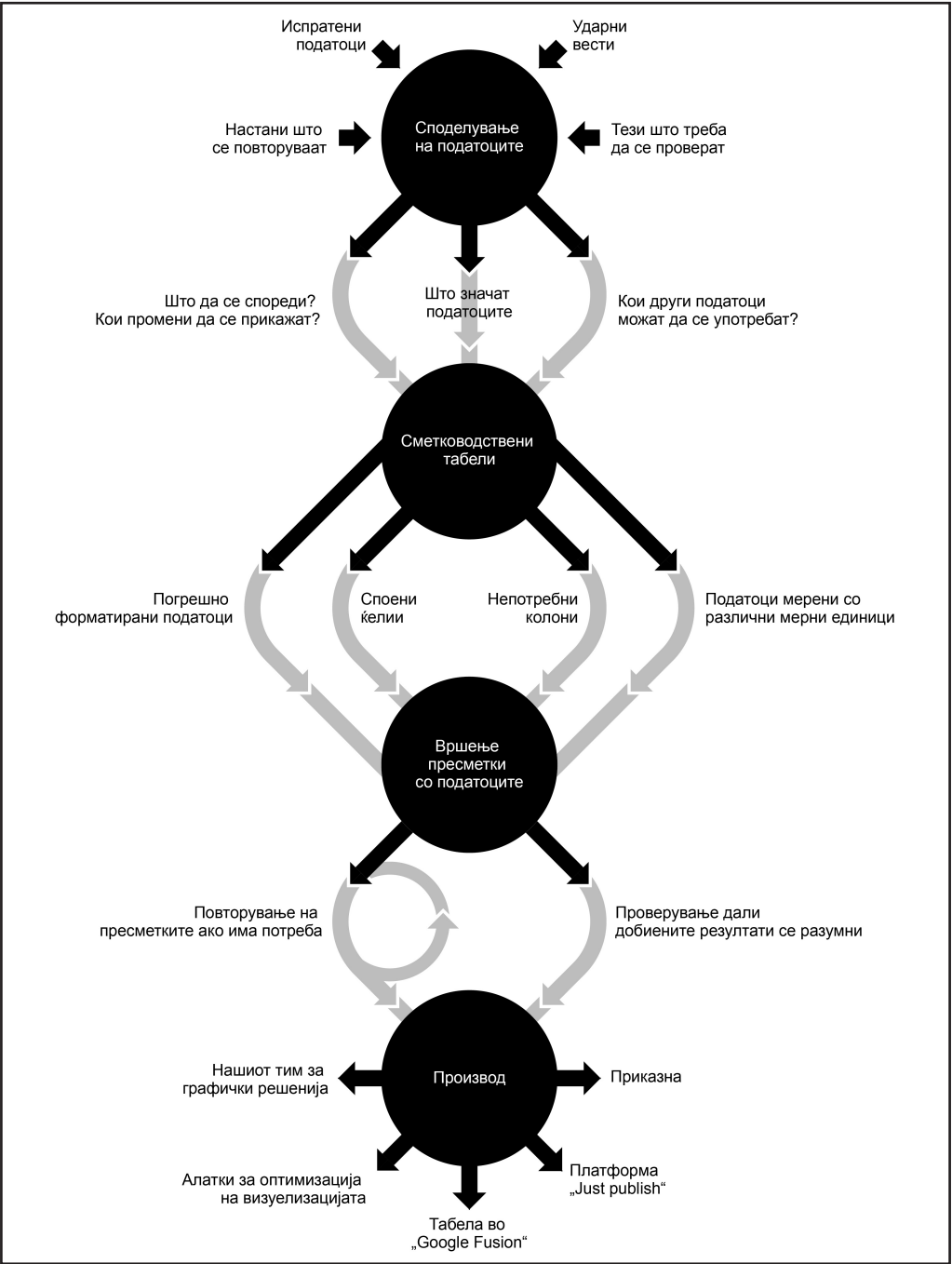
„Датаблогот на Гардијан“, чиј уредник сум, беше замислен како мал блог на кој ќе ги нудиме податоците што служат како основа за нашите приказни. Тој денес се состои од насловна страница (<http://guardian.co.uk/data>); содржи резултати од пребарувања на податоците што ги чуваат владите во целиот свет и глобални развојни податоци; визуелизации подготвени од графичките уметници на „Гардијан“ и собрани од интернет; и алатки за истражување на податоците за јавната потрошувачка. Секој ден ги користиме сметководствените табели на „Google“ за споделување на целосните податоци на кои се заснова нашата работа; ги визуелизираме и анализираме тие податоци и потоа ги користиме како извор на стории за весникот и за веб-сајтот.

Како уредник на вести и новинар што работи со графички прикази, собирањето нови групи на податоци и борбата со нив за да сфатам кои се вестите за тој ден, за мене беше логичен продолжеток на она што веќе го работев.

Прашањето што ми беше поставено веќе беше одговорено. Претходните неколку години беа неверојатно добри за јавните податоци. Обама ги отвори трезорите со податоци на владата на САД со првиот закон што го предложи кога стана претседател, а неговиот пример наскоро го следеа и други влади со нивните веб-страници за чување податоци: Австралија, Нов Зеланд и веб-страната Data.gov.uk на британската влада.

Го имавме скандалот со личните трошоци на членовите на Парламентот, најнеочекуваниот пример за дата-новинарство во Велика Британија – и како последица на тоа, денес имаме ситуација во која Вестминстер се обврза да објавува огромни количества податоци секоја година.

Имавме општи избори на кои сите главни политички партии се обврзаа на транспарентност на податоците и ги отворивме нашите трезори за чување по-



Слика 2-5. Визуелизација на производниот процес во „Датаблогот на Гардијан“ („Гардијан“)

датоци за целиот свет. Неколку весници одделија дел од скапоцениот простор на нивните страници за објавување на базата на податоци „COINS“ на Министерството за финансии.

Во исто време, како што Мрежата исфрла на површината сè повеќе и повеќе податоци, читателите од целиот свет стануваат сè позаинтересирани за суровите/голите факти што стојат зад вестите. Кога го лансиравме „Датаблогот“, мислеам дека главна публика ќе ни бидат девелоперите што развиваат апликации. Всушност, се работи за луѓе што сакаат да знаат повеќе за емисијата на јаглерод диоксид, имиграцијата од Источна Европа, загинатите во Авганистан, дури и колку пати „Битлси“ го искористиле зборот „љубов“ во своите песни (за оние што тоа ги интересира, точната бројка е 613 пати).

Постепено, работата на „Датаблогот“ влијаеше на приказните со кои се соочувавме и ги дополнувавме. Понудивме 458.000 документи за трошоците на членовите на Парламентот на „краудсорсинг“ и ги анализиравме деталните податоци за тоа кој член на Парламентот какви сметки поднесол за надомест. Им помогнавме на нашите корисници да ги истражат деталните бази на податоци за трошењата на Министерството за финансии и ги објавивме податоците зад вестите.

Но вистинскиот пробој за дата-новинарството почна, во пролетта 2010 година, со една сметководствена табела: 92.201 редови податоци, од кои секој ред содржеше детален преглед на една воена акција во Авганистан. Тоа беа воените дневници на „Викиликс“ (WikiLeaks). То ест, тоа беше само првиот дел. Следеа уште две епизоди: Ирак и дипломатските телеграми. Официјалното име за првите два дела беше „SIGACTS: the US military Significant Actions Database“ (Базата на податоци за значајните акции на воените сили на САД).

За новинските организации сè се сведува на географијата – и на близината на редакцијата. Ако сте блиску до редакцијата, лесно можете да предлагате приказни и да бидете дел од процесот; од друга страна, важи правилото „далеку од очите, далеку од срцето“. Пред „Викиликс“, ние бевме сместени на друг кат, заедно со графичкото одделение. По „Викиликс“, нè сместија на истиот кат, веднаш до редакцијата. Тоа значи дека сега полесно можеме да ѝ предлагаме идеи на редакцијата, а на новинарите им е полесно да се сетат на нас кога им треба помош во подготовката на приказните.

Не толку одамна, новинарите контролираа што од официјалните податоци ќе биде прикажано. Ќе напишевме стории за броевите и ќе ги објавевме за благодарната јавност која не беше заинтересирана за суровите статистики. Идејата да дозволиме присуство на сурови информации во весниците се сметаше за анатема.

Таа динамика денес е променета до непрепознатливост. Нашата улога е да се претвориме во толкувачи; да им помогнеме на луѓето да ги разберат податоците, а понекогаш и само да ги објавиме затоа што се интересни сами по себе.

Но, броевите без анализа се само броеви, и тука настапуваме ние. Кога британскиот премиер тврдеше дека немиритите од август 2011 година не беа мотиви-

вирани од сиромаштијата, ние ги објавивме адресите на оние што учествуваа во немирите со индикатори за сиромаштијата за да ја покажеме вистината зад неговите тврдења.

Зад сите стории што се плод на дата-новинарски пристап стои цел еден процес. Тој процес се менува цело време затоа што постојано користиме нови алатки и техники. Некои луѓе велат дека решението е човек да стане некој вид суперхакер, да пишува код и да се внесе целиот во програмскиот јазик „SQL“. Можеби некој навистина ќе реши да го примени тој приод. Сепак, голем дел од работата ја вршиме само во „Ексел“.

Прво ги лоцираме податоците или ги добиваме од различни извори, од најновите вести, од владините податоци, преку новинарски истражувања итн. Потоа размислуваме што би можеле да сториме со тие податоци; дали можеби треба да ги споиме со некоја друга група на податоци? Како да ги прикажеме промените што се случуваат со текот на времето? На таквите сметководствени табели често им е потребно сериозно подредување – сите тие изместени колони и ќелии споени на чуден начин не ни помагаат многу. Се разбира, под претпоставка дека не се работи за „PDF“-датотека, најлошиот формат за прикажување податоци во историјата на човештвото.

Често официјалните податоци се проследени со дополнителни официјални шифри; секое училиште, болница, изборен округ или локален орган на власта си има своја шифра за идентификација.

И државите имаат такви идентификациски шифри (за Обединетото Кралство, на пример, тој знак е „ГБ“). Шифрите се корисни затоа што можеби сакате да споите некои групи податоци, а вчудовидувачки е колку различните начини на пишување и редоследот на зборовите пречат во таа намера. Ги имате Бурма и Мјанмар, на пример, или Округот Фајет (Fayette County) во САД (има 11 окрузи со тоа име во повеќе сојузни држави, од Џорџија до Западна Вирџинија). Шифрите ни овозможуваат да го споредиме сличното со слично.

На крајот од процесот доаѓа резултатот: дали тоа ќе биде приказна, или графикон, или визуелизација? Кои алатки ќе ги користиме? Најдобрите алатки што ги имаме се бесплатните алатки што ни овозможуваат набрзина да произведеме нешто со нив. Пософистицираните графички решенија ги произведува нашиот развоен тим.

Тоа значи дека често ги користиме графиконите на „Гугл“ (Google charts) за помали графикони и статистички пити, или табелите на „Гугл фјужн“ (Google Fusion Tables) за брзо и лесно креирање мапи.

Тоа можеби изгледа како новина, ама навистина не е така.

Во првото издание на „Манчестерски Гардијан“ (The Manchester Guardian) од сабота, 5 мај 1821 година, вестите се објавени на последната страница, како и кај другите весници од тоа време. Првата статија на насловната страница беше оглас за изгубен пес од видот лабрадор.

Меѓу расказите и извадоците поезија, одвај една третина од последната страница е посветена на фактите. Сеопфатна табела со трошоците на училиштата во регионот што никогаш претходно „не биле изложени пред очите на јавноста“, пишува авторот потпишан со иницијалите „Н. Х.“.

Н. Х. сакал тие податоци да бидат објавени, зашто инаку фактите би биле оставени за нив да известуваат несоодветно образовани свештеници. Неговата мотивација ја следела логиката дека „таквите информации содржани овде се вредни; затоа што, без да го знаеме степенот до кој образованието... превладува, дури и најдобрите ставови за состојбата во општеството и неговиот иден напредок нужно мора да се погрешни“. Со други зборови, како може едно општество да се подобри ако луѓето не знаат што се случува околу нив?

Дури и денес не можам да замислам подобро објаснување за тоа што го работиме. Денес она што некогаш беше сторија за последната страница може да стане вест за насловната страница.

– *Џајмон Роџерс, Гардијан (Simon Rogers, The Guardian)*

Дата-новинарството во онлајн-изданието на „Џајт“ (Zeit Online)

Проектот за споредба на богатството врз основа на резултатите на тестот ПИСА (PISA based Wealth Comparison; Програмата за меѓународно оценување на учениците ПИСА е програма на ОБСЕ за оценување на образовните системи во целиот свет преку тестирање на учениците од средните училишта што се спроведува на секои три години, заб. прев.) е интерактивна визуелизација што овозможува споредба на стандардите на живот во различни држави. Проектот користи податоци од сеопфатниот извештај на ОБСЕ за рангирањето на образованието во светот, ПИСА 2009 (PISA 2009), објавен во декември 2010 година. Извештајот се заснова на прашалникот во кој 15-годишни ученици одговараат на прашања во какви услови живеат нивните семејства.

Идејата беше да се анализираат и визуелизираат податоците за да се добие единствен начин на споредба на стандардите на живот во различни држави.

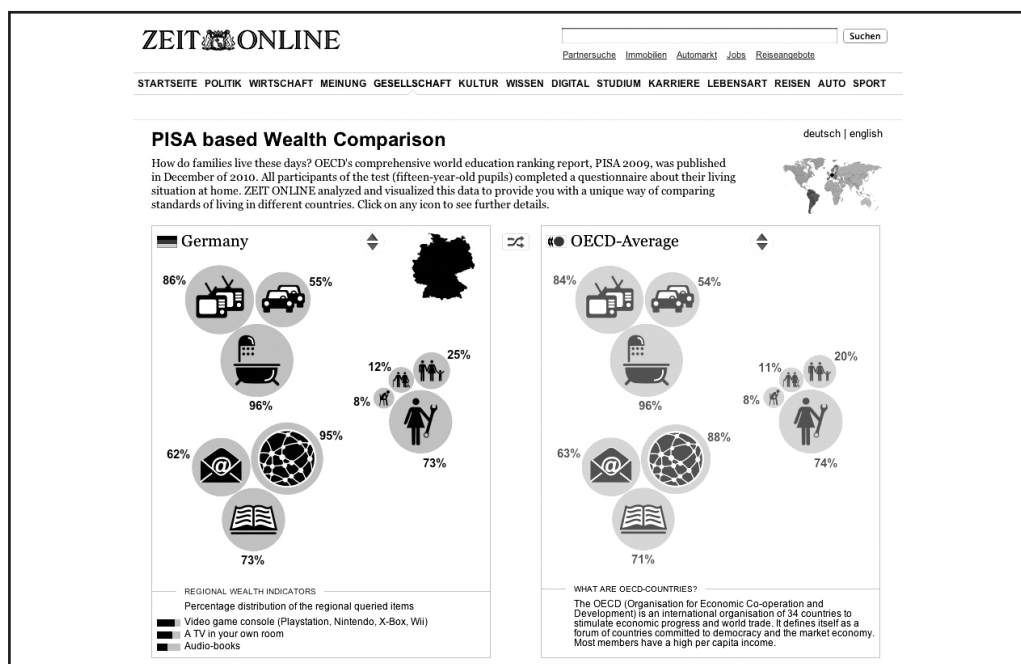
Прво, нашиот внатрешен уредувачки тим одлучи кои факти изгледаат како корисни за споредба на животниот стандард и би требало да се визуелизираат, вклучувајќи ги и следните:

- Материјалното богатство (бројот на ТВ уреди, автомобили и бањи и тоалети во домот)
- Семејната состојба (дали со семејството живеат и бабата и дедото, процентот на семејства со едно дете, невработеноста на родителите, дали мајката е вработена или не)

- Пристап до извори на знаење (интернет-конекција во домот, фреквенцијата на користење електронска пошта и количеството книги што ги поседува семејството)
- Три дополнителни показатели за степенот на развој за секоја од опфатените земји

Со помош на тим од дизајнери што работи внатре во редакцијата, тие факти беа преведени во лесно разбирливи илустрации. Направивме насловна страница со дизајн кој овозможуваше споредба помеѓу различните земји меѓу кои можеше да се бира, кои беа претставени како да играат карти.

Потоа ги контактиравме луѓето од германската „Мрежа за отворени податоци“ (German Open Data Network) за да побараме девелопери што можат да ни помогнат во реализацијата на проектот. Таа заедница на високо мотивирани луѓе ни ја посочи Грегор Аиш (Gregor Aisch), еден мошне талентиран дизајнер на информации, за да го напишат компјутерскиот код за апликацијата што ќе значеше остварување на нашите соништа (единствениот услов што го поставивме беше да не користат „Flash“!). Грегор создаде мошне квалитетна интерактивна визуелизација стилизирана како убави меурчиња, по урнекот на библиотеката „Рафаел џаваскрипт“ (Raphaël-Javascript Library).



Слика 2-6. Споредба на богатството врз основа на резултатите на ранжирањето ПИСА („Џајт онлајн“)

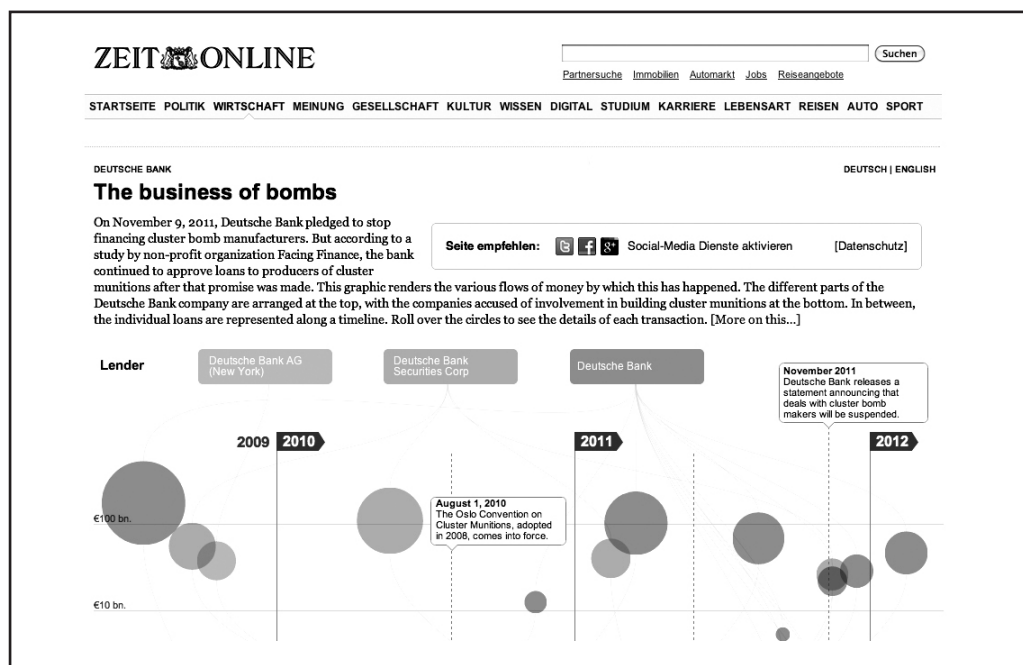
Како резултат од соработката добивме еден мошне успешен интерактивен приказ што привлече многу сообраќај. Лесно е да се направи споредба помеѓу кои било две држави, што значи дека приказот е корисен како референтна алатка. Тоа значи дека можеме да го користиме во нашата секојдневна новинарска работа. На пример, ако известуваме за нешто што е поврзано со животот во Индонезија, можеме лесно и брзо да вградиме графичко решение со споредба на животните услови во Индонезија и Германија. Преносот на знаење на тимот внатре, во нашата организација, претставуваше извонредна инвестиција за идните проекти.

Ние што работиме во “Zeit Online” откривме дека проектите во областа на дата-новинарството ни носат многу сообраќај и ни помагаат да ги ангажираме читателите и публиката на нови начини. На пример, состојбата во атомската централа во Фукушима, по цунамито што ја погоди Јапонија, предизвика голем интерес и за неа многу се известуваше. По истекувањето на радиоактивниот материјал, беше евакуиран целиот простор во радиус од 30 километри околу електраната. Луѓето можеа да прочитаат и да погледнат многу информации за евакуацијата. “Zeit Online” пронајде еден иновативен начин да ѝ го објасни влијанието на тие настани на својата публика во Германија. Прашавме колку луѓе живеат во близина на атомска електрана во Германија? Колку луѓе живеат во радиус од 30 километри од најблиската нуклеарна централа? Една карта покажува колку луѓе би требало да бидат евакуирани ако слична ситуација би настанала во Германија. Како резултат добивме реки сообраќај; всушност, целиот проект стана вирален во сферата на социјалните медиуми. Проектите во областа на дата-новинарството релативно лесно можат да се приспособат за објавување на други јазици. Ние направивме верзија на англиски јазик за близината на населените места до нуклеарните електрани во САД, што беше извонреден магнет за сообраќај кон нашата интернет-страница. Новинските организации сакаат да бидат препознаени од читателите како авторитативни извори на кои може да им се верува. Откривме дека проектите на полето на дата-новинарството, во комбинација со можноста за нашите читатели да ги прегледаат и повторно да ги искористат суровите податоци, ни носат високо ниво на доверба.

Веќе две години, Одделението за истражување и развој и главниот уредник на „Цајт онлајн“ Волфганг Блау (Wolfgang Blau) го застапуваат ставот дека дата-новинарството е значаен начин на раскажување стории. Транспарентноста, кредибилитетот и инволвираноста на корисниците се значајни делови на нашата филозофија. Токму поради тоа, дата-новинарството е природен дел од нашата сегашна и идна работа. Визуелизациите на податоци можат да му донесат нова вредност на начинот на восприемање на една приказна и претставуваат атрактивен начин за уредувачкиот тим да ги претстави своите содржини.

На пример, на 9 ноември 2011 година, „Дојче банк“ (Deutsche Bank) јавно презеде обврска да престане со финансирањето на производителите на касетни бомби. Сепак, според една студија на непрофитната организација „Фејсинг фајненсе“

(Facing Finance), банката продолжи да им одобрува заеми на производителите на касетна муниција и по даденото ветување. Визуелизацијата што ја подготвливме врз основа на тие податоци им ги прикажува на читателите различните текови на парите. Уделите на компанијата „Дојче банк“ се наведени во горниот дел на визуелизацијата, а компаниите обвинети за вмешаност во производство на касетна муниција на дното. Во средината на визуелизацијата се претставени индивидуалните заеми во вид на хронолошки приказ. Кога курсорот на глумчето поминува преку круговите се прикажуваат деталите за секоја поединечна трансакција. Сепак, визуелизацијата им дозволува на читателите да ги разберат и истражат финансиските врски и меѓузависности на поинтуитивен начин.



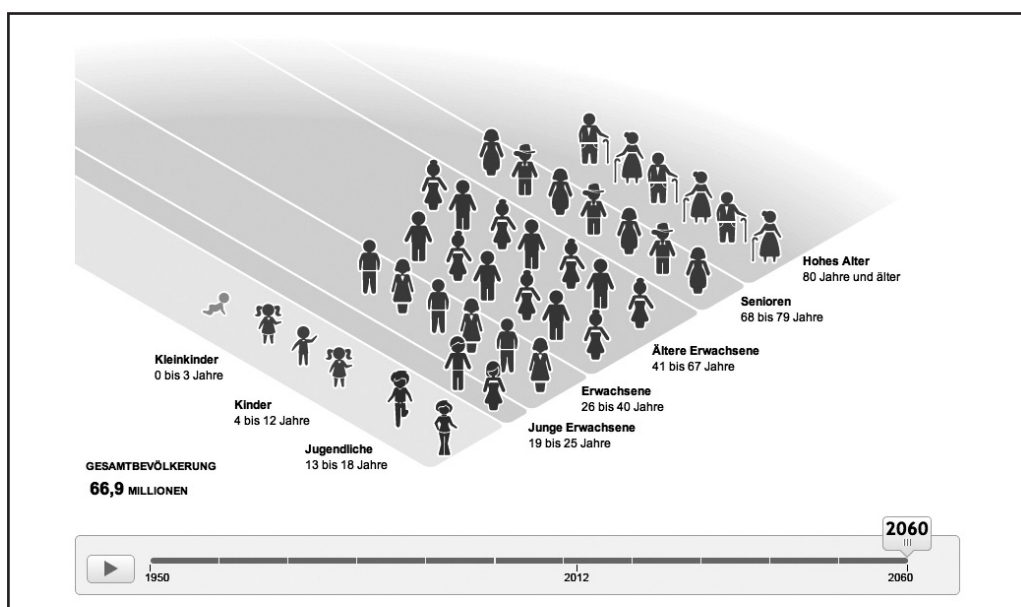
Слика 2-7. Бизнисот на производство на бомби („Цајт онлајн“)

Да разгледаме еден друг пример: Федералниот завод за статистика на Германија објави една извонредна група податоци за животниот век во Германија, вклучувајќи и модели на различни демографски сценарија до 2060 година. Типично, таквите податоци се претставуваат во форма на популациска пирамида, како што е пирамидата на Федералниот завод за статистика.

Во соработка со колегите од научната рубрика, се обидовме на нашите читатели да им обезбедиме подобар начин за разгледување на проектираните

демографски податоци за нашето општество во иднината. Во нашата визуелизација претставуваме статистички репрезентативна група од 40 лица на различна возраст, што живеат во различни периоди од 1950 до 2060 година. Ги организиравме во осум различни групи. Изгледа како групна фотографија на германското општество во различни точки во времето. Истите податоци, визуелизирани со традиционална популациска пирамида, ни даваат само апстрактен осет за ситуацијата, додека групата во која има деца, млади, средовечни и стари лица значи дека читателите полесно можат да се идентификуваат со податоците. Треба само да го притиснете копчето „плеј“ за да тргнете на пат низ времето долг 11 децении. Исто така, можете да ја внесете вашата година на раѓање и полот и да станете дел од групната фотографија: така ќе го видите вашето демографско патување низ децениите и очекуваниот животен век.

– Саша Фенор, „Цајт онлајн“ (Sascha Venohr, Zeit Online)



Слика 2-8. Визуелизација на демографски податоци („Цајт онлајн“)

Како се најмува хакер

Едно прашање што новинарите редовно ми го поставуваат е: „Како да најдам програмер што ќе ми помогне во работата на мојот проект?“ Немојте да се залажувате дека тоа е еднонасочен процес; граѓански ориентираните хакери и лицата што работат со податоци често се исто толку орни да стапат во контакт со новинари.

Помеѓу корисниците на алатките и услугите за работа со податоци, новинарите се оние што ја имаат најголемата моќ. Од гледна точка на девелоперите, новинарите размислуваат латерално за можните начини на користење на алатките засновани на податоци во контексти за кои девелоперите не размислувале претходно (повратните информации во таков случај се бесценети!). Тие, исто така, помагаат во воспоставувањето на контекстот и во промоцијата на проектите и помагаат тие да добијат релевантност. Се работи за симбиотски однос.

За среќа, тоа значи дека без оглед на тоа дали сакате да најмате хакер или барате можни соработници, а имате ограничен буџет, многу веројатно е дека веќе постои некој што е заинтересиран да ви помогне.

Прашањето е, како да го пронајдете? Арон Пилхофер (Aron Pilhofer) од „Њујорк тајмс“ вели:

Можеби ќе откриете дека во вашата организација веќе работат луѓе што ги поседуваат сите вештини кои ви се потребни, само не работат во вашата редакција. Прошетајте наоколу, посетете ги одделенијата за технологија и информатички технологии и веројатно ќе го најдете она што го барате. Од еднаква важност е правилно да ја цените културата на програмерите: пријдете некому што има компјутер кој наликува на компјутерот на Слика 2-9... веројатно токму тоа е лицето што го барате.



Слика 2-9. Хакериите лесно се познаваат по налепниците што ѓордо ги красат нивните компјутери (фотографија: Луси Чејмберс)

Следат уште неколку идеи:

Посйавейше оглас на веб-страница за слободни работни места

Идентификувајте ги веб-страниците што како целна група ги имаат девелоперите кои работат со различни програмски јазици. На пример, поставете оглас на страницата „Огласник за работни места на Питон“ (Python Job Board, <http://www.python.org/community/jobs/>).

Контакттирајте ги релевантниите „мејлинг-листии“

Такви листи, на пример, се NICAR-L и Data Driven Journalism.

Контакттирајте ги релевантниите организации

На пример, ако сакате да прочистите или да „истругате“ некои податоци од интернет, можете да контактирате некоја организација каква што е „Скрејпервики“ (Scraperwiki), која има обемен адресар на проверени програмери што се подготвени да помогнат.

Зачленете се во релевантни групи/мрежи

Побарајте ги иницијативите каква што е „Хекс/хакерс“ (Hacks/Hackers), група што на едно место ги собира новинарите и техничарите. Локални организации на „Хекс/хакерс“ сега се појавуваат насекаде во светот. Исто така, можете да поставите оглас или повик на нивниот билтен со понуди за работа.

Заедниците што се интересираат за локални прашања

Пробајте да направите брзо пребарување за определена област на стручност во географската област во која живеете и работите (на пример, “javascript” + “london”). Интернет-страниците каква што е Meetup.com, исто така може да послужат како одлична појдовна точка.

„Хакаџони“ и натпреварувања

Без оглед дали има парична награда или не, натпреварувањата во изработка на апликации и визуелизации и развојните денови често се плодна почва за воспоставување нови врски и соработка.

Прашајте „џик“!

„Гиковите“ (geek, „гик“ е жаргонски израз за лице со ексцентрично однесување што пред сè е заинтересирано за прашања поврзани со новите технологии, заб.прев.) се дружат со други гикови. Личните препораки се добар начин за пронаоѓање соработници.

— Луси Чејмберс, Фондација „Отворено знаење“ (Lucy Chambers, Open Knowledge Foundation)

Хакерски вештини

Откако ќе пронајдете хакер, како ќе знаете дали го бидува за тоа што го работи? Го прашавме Алистер Дент (Alastair Dant) од „Гардијан“ да ни каже како да забележиме кој е добар хакер:

Добриите хакери го кодираат целиот пакет

Кога си имате работа со цврсти рокови, подобро е да знаете од сè по малку отколку да бидете вистински мајстор во само една работа. Новинските апликации бараат „кротење“ на податоците, динамички графички решенија и многу храброст.

Добриите хакери ја гледаат пошироката слика

Холистичкиот начин на размислување ја претпочита вредноста на наративот пред техничките детали. Лично, по прво би слушнал една единствена нота отсвирена со чувство отколку долга, технички совршено отсвирена соло-делница. Откријте кој сака да работи заедно со дизајнер.

Раскажуваат добра приказна

Наративната презентација бара на нештата да им го најдеме местото во просторот и времето. Откријте на кој свој проект се најгорди и побарајте да ве спроведат низ процесот на изработката; така ќе можете да ја проверите и нивната способност за комуникација, а не само нивото на техничко знаење што го поседуваат.

Хакериите разговараат за проблемите

Секоја брза изработка бара ангажман на мешани тимови што работат на една заедничка цел. Секој од учесниците треба да ги почитува своите колеги и да биде подготвен да преговара со нив. Непредвидените проблеми често бараат брза промена на плановите и колективен компромис.

Хакериите ги усовршуваат своите знаења

Технологијата се развива со голема брзина. Тешко е да се држи чекор со тој развој. Бидејќи познавам добри девелопери што потекнуваат од различни места и ситуации, најчестата заедничка особина им е подготвеноста да научат нови работи ако за тоа постои потреба.

— Луси Чејмберс, Фондација “Отворено знаење”, извадено од интервју со Алистер Дент, шеф на одделението за интерактивни технологии во „Гардијан“

Како да го пронајдете девелоперот од вашите соништа

Разликата во продуктивноста помеѓу добриот и извонредниот девелопер не е линерна – таа е експоненцијална. Успешниот избор на соработници е од огромно значење. За жал, таквиот избор не е секогаш лесен. Доволно е што е тешко да се избере меѓу кандидатите ако и самиот не сте искусен технички менаџер. Ако го додадеме и проблемот со висината на платите во новинските организации, предизвикот станува навистина голем.

Ние, во „Трибјун“, користиме два приоди во регрутирањето нов кадар: емоционалната исполнетост и привлечноста на техничката страна на работата. Емоционалниот приод се сведува на следното: новинарството е од суштинско значење за функционирањето на демократијата. Ако работите кај нас, можете да го промените светот. Од техничката страна, ја промовираме можноста за учење нови работи. Нашите проекти се мали, брзи и повторливи. Секој проект си носи свој комплет алатки, свој јазик, своја тема (безбедноста од пожари, пензискиот систем) што треба да се совладаат. Редакцијата е еден голем лабораториски сад. Никогаш порано не сум раководел со тим што научил толку многу, толку брзо, како овој тим.

Што се однесува до прашањето каде да барате, ние ја имавме среќата да пронајдеме извонредни хакери во заедницата за отворена влада. Мејлинг-листата „Sunlight Labs“ е место на кое добронамерните „бубалици“ што работат на „посрани“ работни места ги минуваат бессоните ноќи. Друг потенцијален извор е „Код за Америка“ (Code for America). Секоја година, по неколку момци и девојки излегуваат од „Код за Америка“ во потрага по следниот голем проект. Како бонус, „Код за Америка“ користи мошне ригорозен приемен процес; тие веќе ги провериле кандидатите за вас. Денес, новинари заинтересирани за програмирање, исто така, излегуваат и од новинарските школи. Можеби се неискусни, но имаат потенцијал за продавање.

Најпосле, не е доволно само да се вработат девелопери. Она што ви треба е технички менаџмент. Девелоперот што работи сам (особено ако е тучушто излезен од новинарска школа, без искуство внатре во индустријата) неизбежно ќе донесе и некои лоши одлуки. Дури и најдобрите програмери, оставени сами на себе, ќе изберат да работат нешто што е интересно од техничка страна, наместо она што е значајно за вашата публика.

Тоа работно место наречете го „уредник за новински апликации“, „проектен менаџер“, наречете го како сакате. Исто како и на оние што пишуваат, и на програмерите им требаат уредници, ментори, некој што ќе ги притиска да ги испочитуваат роковите во производството на софтвер.

– Брајан Бојер, „Чикаго Трибјун“

Користење надворешна експертиза преку „хакатони“

Во март 2010 година, организацијата за дигитална култура „СЕТУП“ (SETUP) од Утрехт организираше настан под име „Хакирање на новинарството“ (Hacking Journalism). Целта на настанот беше да ја охрабри соработката помеѓу девелоперите и новинарите.

„Хакатоните ги организираме за да направиме „кул“ апликации, но ние не знаеме како да ги препознаеме интересните приказни што се кријат во податоците. Тоа што го работиме има мало значење за општеството“, рекоа програмерите. „Го препознаваме значењето на дата-новинарството, но ги немаме потребните технички знаења да ги изработиме работите што ни требаат“, рекоа новинарите.



Слика 2-10. Новинариите и девелопериите на Региохак (Фотографија: Хајнце Хавинџа (Heinze Havinga))

Работата во регионален весник ме научи дека таму нема ниту пари ниту волја да се вработи програмер во редакцијата. Во тоа време, дата-новинарство беше непозната работа за холандските весници.

Моделот на организирање „хакатони“ совршено одговараше на задачата; релаксирана атмосфера за заедничка работа, со многу пица и енергетски пијалаци. „РегиоХак“ (RegioHack) е хакатон организиран од мојот работодавач, регионалниот весник „Де Стентор“ (De Stentor), неговото сестринско издание „ТЦ Тубантија“ (TC Tubantia) и „Саксион хогеншолен еншеде“ (Saxion Hogenschoolen Enschede). Настанот, всушност, се одржа во просториите на „Саксион“.

Правилата беа дека секој може да се пријави за учество во 30-часовен „хакатон“. Ние ги обезбедивме храната и пијалациите за учесниците. Целта беше да имаме 30 учесници, поделени во шест групи. Тие групи требаше да се фокусираат на различни теми – криминал, здравство, транспорт, безбедност, стареење на населението и носителите на моќ. Трите главни цели на настанот за нас беа:

Да ѝ пронајдеме ѝ приказни

За нас, дата-новинарство е нова и непозната работа. Ползата од него можеме да ја покажеме само преку добро подготвени стории. Планиравме да произведеме барем три стории засновани на начелата на дата-новинарството.

Да ги ѝоврземе луѓето едни со други

Ние новинарите не знаеме како се извршува дата-новинарството, ама барем не ни глумиме дека знаеме. Ставајќи ги новинарите, студентите и програмерите да поминат 30 часа во иста просторија, сакавме тие меѓу себе да ги разменуваат своите знаења и ставови.

Да организираме дружење

Весниците не организираат многу општествени настани, а „хакатони“ уште помалку. Сакавме да провериме дали еден таков настан ќе даде резултати. Целиот настан можеше, всушност, да излезе прилично напнат: 30 часа поминати со странци, сите говорат на својот жаргон, си ја чукате главата за основни поими и прашања и работите надвор од зоната во која се чувствувате најудобно. Со тоа што решивме тоа да биде дружење (веќе ги спомнав пицата и енергетските пијалаци), сакавме да создадеме атмосфера во која новинарите и програмерите ќе се чувствуваат комфортно и ќе соработуваат ефективно.

Пред настанот, „ТЦ Тубантија“ објави интервју со вдовицата на еден полицаец која беше напишала книга за годините што нејзиниот починат сопруг ги поминал работејќи како полицаец. Таа имаше и документ на кој беа наведени сите регистрирани убиства во источните делови на Холандија, список што нејзиниот сопруг почнал да го води во 1945 година и го дополнувал редовно. Во нормални услови, тој документ би го објавиле на нашата интернет-страница. Овој пат, изготвивме една „командна табла“ (dashboard) со употреба на софверската апликација „Табло“ (Tableau). Исто така, блогиравме за настанувањето на таа активност на интернет-страницата на нашиот „РегиоХак“ (RegioHack).

За време на „хакатонот“, една проектна група се зафати со темата на развојот на училиштата и стареењето на населението во нашиот географски ре-

гион. Преку нивната визуелизација на идните предвидувања и проекции, можеме да видиме кои градови ќе се најдат во неволја по неколкугодишен пад на бројот на новозапишани деца во училиштата. Тие сознанија ги искористиме да подготвиме статија за тоа како тој процес ќе влијае врз училиштата во нашиот регион.

Започнавме уште еден мошне амбициозен проект по име „Двестетемина од Твенте“ („De Tweehonderd van Twente“) за да видиме кој ја држи моќта во нашиот регион и да направиме база на податоци на највлијателните луѓе во регионот. Списокот на влијателни луѓе ќе го направиме преку пресметка, најверојатно во „Гугл“ или слична апликација, за тоа кој поседува најголем број врски со моќни организации. Од списокот можеби ќе произлезе серија статии, но и сам за себе ќе биде моќна алатка за новинарите. Кој е поврзан со кого? На базата можете да ѝ поставувате прашања и да ја користите во секојдневната работа. Базата, исто така, ќе има и културна вредност. Неколку уметници веќе ни пристапија со барање да ја искористат базата на податоци, кога ќе ја завршиме, во изработката на интерактивни уметнички инсталации.

По „РегиоХакот“, забележавме дека новинарите почнаа да гледаат на датановинарството како на одржлив додаток на традиционалното новинарство. Моите колеги продолжија да ги користат и надградуваат техниките што ги на-



Слика 2-11. Околу дајта-новинарството се раѓаат нови заедници (Фотографија: Хајнце Хавинџа)

учија тој ден за да подготват поамбициозни и технички подотерани проекти, како што е една база на податоци која ги содржи административните трошоци на домувањето. Тие податоци ги искористив да подготвам интерактивна мапа изработена со апликацијата „Fusion Tables“. На пример, побаравме од читателите да си играат со податоците и ги „краудсорсиравме“ резултатите преку <http://bit.ly/scratchbook-crowdsourcing>. Откако добивме многу прашања за начинот на изработка на мапа во „Fusion Tables“, решив да снимам и видео упатство.

Што научивме? Научивме многу, но, исто така, се сретнавме и со бројни пречки, од кои ви ги претставуваме следните четири:

Дали да се почне со прашањата или со податоците?

Безмалку сите проекти заглавија во фазата на барање информации. Во најголем број случаи, проектите почнаа со некое новинарско прашање. Но што потоа? Кои податоци се достапни? Каде да ги најдете? Кога ќе ги пронајдете податоците, дали тие даваат одговор на вашето прашање? Новинарите, кога истражуваат за потребите на некоја статија, обично знаат каде можат да ги пронајдат информациите што им требаат. Кај дата-новинарството, повеќето новинари воопшто не знаат какви информации им стојат на располагање.

Недоволно техничко знаење

Дата-новинарството е, пред сè, техничка дисциплина. Понекогаш мора да „стругате“, понекогаш мора да програмирате за да ги визуелизирате вашите резултати. За одлично дата-новинарство, потребно е да покриете два аспекти: новинарски увид од страна на искусен новинар и технички знаења што ги поседува сестран познавач на дигиталниот домен. Тие две работи не беа честа појава на „РегиоХакот“.

Дали е весел?

Учесниците најчесто користеа една група податоци за барање вести, наместо да ги пребаруваат врските што постојат помеѓу различни извори. Причината лежи во тоа што верификацијата на вестите произведени преку дата-новинарство бараат определени познавања од областа на статистиката.

Постои ли вообичаена, руиниска постојайка?

Сето погоре кажано се сведува на заклучокот дека не постои рутинска постапка. Учесниците поседуваат некакви вештини, но не знаат како и кога да ги употребат. Еден новинар го спореди сето тоа со печење колач. „Ги имаме сите состојки: брашно, јајца, млеко итн. Е тогаш ги ставаме сите во кеса, ја протресуваме убаво и се надеваме дека од неа ќе излезе готов колач“. Навистина, ги имаме сите состојки, но не го знаеме рецептот.

Што понатаму? Нашите први искуства со дата-новинарството можат да им помогнат на други новинари или програмери што сакаат да влезат во истото поле на работа, а ние работиме на еден извештај во таа смисла.

Исто така размислуваме како да продолжиме со „РегиоХак“ во форма на „хакатон“. Откривме дека се работи за забавен, поучен и продуктивен пристап што претставува одличен вовед во дата-новинарството.

Но, за успешно дата-новинарство, мора тој пристап да го интегрираме во работата во редакцијата. Новинарите мора да размислуваат не само за цитати, соопштенија, состаноци на општинските совети, туку и за податоците. Со организацијата на „РадиоХакот“, им докажавме на нашите читатели дека дата-новинарството не е само некаква последна мода. Сега пишуваме подобри статии засновани на подобри информации, а во исто време им нудиме на читателите разновидни статии и во печатената верзија и на интернет.

– Џери Ферманен (Jerry Vermanen), NU.nl

Следење на парите: Дата-новинарството и прекуграничната соработка

Истражувачките новинари и граѓаните што се заинтересирани за расветлување на организираниот криминал и корупцијата што влијае врз животите на милијарди луѓе во целиот свет, со секој нов ден добиваат претходно невиден пристап до информации. Огромни количества податоци се објавени на интернет од страна на владите и други организации и изгледа дека сите можат да дојдат до сè повеќе информации. Од друга страна, корумпираните владини функционери и организираниот криминал истовремено прават сè што можат да ги скријат информациите, а со тоа и своите недела. Тие се обидуваат да ги држат луѓето во незнаење додека си ги договараат матните зделки што ги разоруваат општествата на сите нивоа и предизвикуваат конфликти, глад или други кризи.

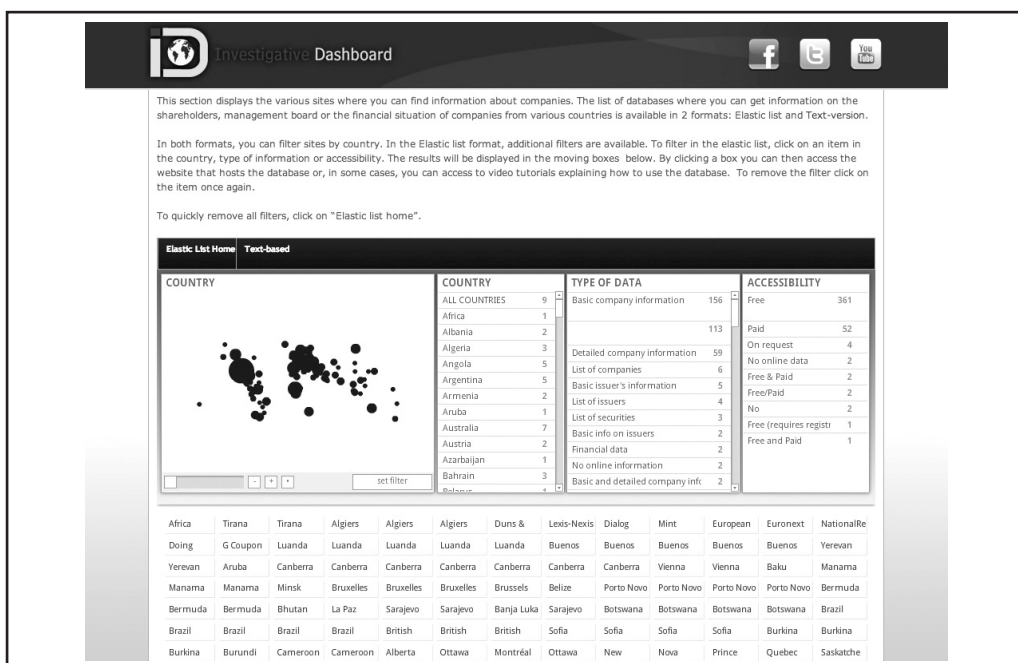
Истражувачките новинари имаат должност да ги откријат таквите недела и со тоа да ги оневозможат коруптивните и криминалните механизми.

Постојат три основни упатства што, ако се следат доследно, водат до добро и темелно новинарство при истражувањето на големите случаи на корупција и криминал, дури и во најнеповолни околности:

Размислувајте вон граници на вашата држава

Во многу случаи, полесно е да се дојде до информации од странство отколку во државата во која работи истражувачкиот новинар. Можеби ќе се покаже дека информациите собрани од странство, преку странски бази на информации или со користење на законите за пристап до информации на други земји, биле токму она што ви треба за да ја составите сложувалката на вашето истражување. Најпосле, криминалците и корумпираните функцио-

нери ретко ги чуваат парите во земјата од каде што ги украде. Поверојатно е дека ќе ги чуваат во странски банки или ќе ги инвестираат во странство. Криминалот е глобализиран. Бази на податоци што им помагаат на истражувачките новинари во следењето на движењето на парите во светот можат да се пронајдат на многу места на интернет. На пример, „Истражувачката контролна табла“ (Investigative Dashboard, <http://investigativedashboard.org/>) им овозможува да го следат движењето на парите преку границите.



Слика 2-12. Истражувачката контролна табла (Проект за известување за организирано криминал и корупција (OCCRP))

Искористување на меѓагоспојетските мрежи за истражувачко новинарство

Истражувачките новинари во светот се собираат во организации какви што се „Проектот за известување за организирано криминал и корупцијата“ (Organized Crime and Corruption Reporting Project), „Африканскиот форум за истражувачко новинарство“ (African Forum for Investigative Reporting), „Арапските репортери за истражувачко новинарство“ (Arab Reporters for Investigative Journalism) и „Глобалната мрежа за истражувачко новинарство“ (Global Investigative Journalism Network). Новинарите, исто така, можат да ги искористат и професионалните новинарски платформи каква што е „Ај-Деј Нет“ (IJNet), каде секојдневно се разменуваат глобални информации поврзани со новинарската работа. Многу од новинарите со-

брани во мрежи работат на слични теми и се соочуваат со слични ситуации, така што размената на информации и методи е логична работа. Таквите мрежи се поврзани со мејлинг-листи или преку групи на социјалните мрежи, така што е лесно да се стапи во контакт со колеги новинари и да им се побара информација или корисен совет. На таквите форуми и мејлинг-листи можат да се најдат и идеи за нови истражувања.

Искористете ја технологијата и соработувајте со хакери

Постојат софтверски решенија што им помагаат на истражувачките новинари во пристапот до информациите и нивната обработка. Различни видови софтвер помагаат во бришењето на „шумот“, во копањето и разбирањето големи количества податоци и во пронаоѓањето на документите потребни да се објави и поткрепи приказната. Постојат повеќе готови софтверски програми што можат да се искористат за анализа, собирање и толкување на информациите и, што е поважно, истражувачките новинари треба да се свесни дека постојат многу програмери подготвени да помогнат, само треба да бидат прашани. Таквите програмери или хакери знаат како да дојдат до информациите и како да ракуваат со нив, а можат да бидат од голема помош и во истражувачката работа. Таквите програмери, од кои многумина се членови на глобалните движења за отвореност на податоците, можат да бидат сојузници од непроценлива вредност во борбата против криминалот и корупцијата, способни да им помогнат на новинарите во собирањето и анализирањето на информациите.

Добар пример за таква соработка помеѓу програмерите и граѓаните е „Скрејпервики“ (ScraperWiki), интернет-страница на која новинарите можат да ги прашаат програмерите за помош во извлекувањето податоци од интернет-страници. „Истражувачката контролна табла“ нуди цел список готови алатки што можат да им помогнат на новинарите во собирањето, обликувањето и анализирањето на податоците.

Ползата од гореспомнатите упатства е видлива во многу примери.

Еден добар пример за тоа е работата на Кадија Исмаилова (Khadija Ismaylova), искусна истражувачка новинарка од Азербејџан која работи во услови на оскуден и ограничен пристап до информациите. Исмаилова мора секојдневно да се расправа со секакви пречки за да ѝ понуди на јавноста во Азербејџан добри и веродостојни информации. Во јуни 2011 година, Исмаилова, која работи за канцеларијата на Радио Слободна Европа во Баку, објави дека ќерките на претседателот на Азербејџан, Илхам Алиев (Ilham Aliyev), се тајни сопственички на брзорастечката телекомуникациска компанија „Азерфон“, преку удел во оф-шор компанијата со седиште во Панама. Компанијата тогаш имаше 1,7 милиони претплатници, покриваше 80 одсто од територијата на државата и беше (во тој момент) единствен понудувач на „3Г“ (3G) услуги во Азербејџан. Исмаилова потроши три години на обидите да открие кои се сопствениците на телекомуникациската компанија, но владата одби да ги от-

крие информациите за акционерите и повеќе пати отворено лажеше за сопственоста на компанијата. Дури тврдеа дека компанијата е во сопственост на германската компанија „Сименс АГ“ (Siemens AG), тврдење што германската корпорација веднаш го негираше. Азербејџанската новинарка успеа да открие дека сопственици на „Азерфон“ се неколку приватни компании регистрирани во Панама. Нејзе ѝ изгледаше како да нема каде понатаму, сè додека не доби помош од странство. Во почетокот на 2011 година, Исмаилова разбра преку „Истражувачката контролна табла“ (Investigative Dashboard) дека компаниите со седиште во Панама можат да се следат со помош на апликација развиена од програмерот и активист Ден О’Хајгин (Dan O’Huiginn). Со помош на таа алатка, таа најпосле откри дека двете ќерки на претседателот се инволвирани во телекомуникациската компанија преку бизниси регистрирани во Панама.

О’Хајгин всушност создаде алатка која им помага на новинарите од целиот свет да ја истражуваат корупцијата – Панама, познато оф-шор прибежиште, наголемо ја користат корумпирани функционери кои таму ги засолнуваат украдените пари [од пријателите на поранешниот египетски претседател Хосни Мубарак (Hosni Mubarak) до „валканите“ државни функционери од Балканот или Латинска Америка]. Она што програмерот-активист го создаде се нарекува „веб-стругач“, метода што дозволува извлекување и преобликување на информациите на начин што овозможува нивно користење од страна на истражувачите. О’Хајгин го „иструга“ регистарот на компании на Панама затоа што тој регистар, иако отворен, дозволува пребарување само ако истражувачкиот новинар го знае точното име на бараното трговско друштво. Тоа ги ограничува можностите за истражување, знаејќи дека новинарите обично пребаруваат по име на лицето чија имотната состојба ја проверуваат. Тој ги извлече податоците и создаде нова веб-страница на која можеше да се пребарува и по име на лице. Новата веб-страница им овозможи на истражувачките новинари од многу земји да бараат информации, да ги проверуваат имињата на владините функционери и парламентарците и да проверат дали се тајни сопственици на фирми во Панама (како што беше случај со роднините на претседателот на Азербејџан).

Следењето на советите наведени погоре носи и други предности, освен по-добриот пристап до информации. Една од тие предности е поврзана со намалувањето на штетата и обезбедувањето подобра заштита за истражувачките новинари што работат во непријателско опкружување. Тоа се должи на фактот што кога работи како дел од мрежа, новинарот не е сам; истражувачкиот новинар соработува со колеги од други земји, па на криминалците им е потешко точно да утврдат кој е одговорен за откривањето на нивните недела. Како резултат, владите и корумпираните потешко можат да преземат одмазднички дејства.

Треба да се има предвид и дека информациите што не изгледаат многу вредни во една географска област можат да бидат од клучно значење во друга. Размената на информации што се одвива преку истражувачките мрежи може да доведе до објавување многу значајни стории. На пример, информацијата

дека романски државјанин е уапсен во Колумбија со еден килограм кокаин најверојатно не е вест што ќе се најде на насловните страници на весниците во Богота, но може да се покаже како извонредно важна за романската јавност ако локален новинар успее да открие дека лицето фатено со дрогата работи во Владата во Букурешт.

Ефикасното истражувачко новинарство е резултат на соработка меѓу истражувачки новинари, програмери и сите други што сакаат да ги користат податоците за создавање почисто, поправично и поправедно глобално општество.

– *Паул Раду, Проект за извештување за организираниот криминал и корупцијата (Paul Radu, Organized Crime and Corruption Reporting Project)*

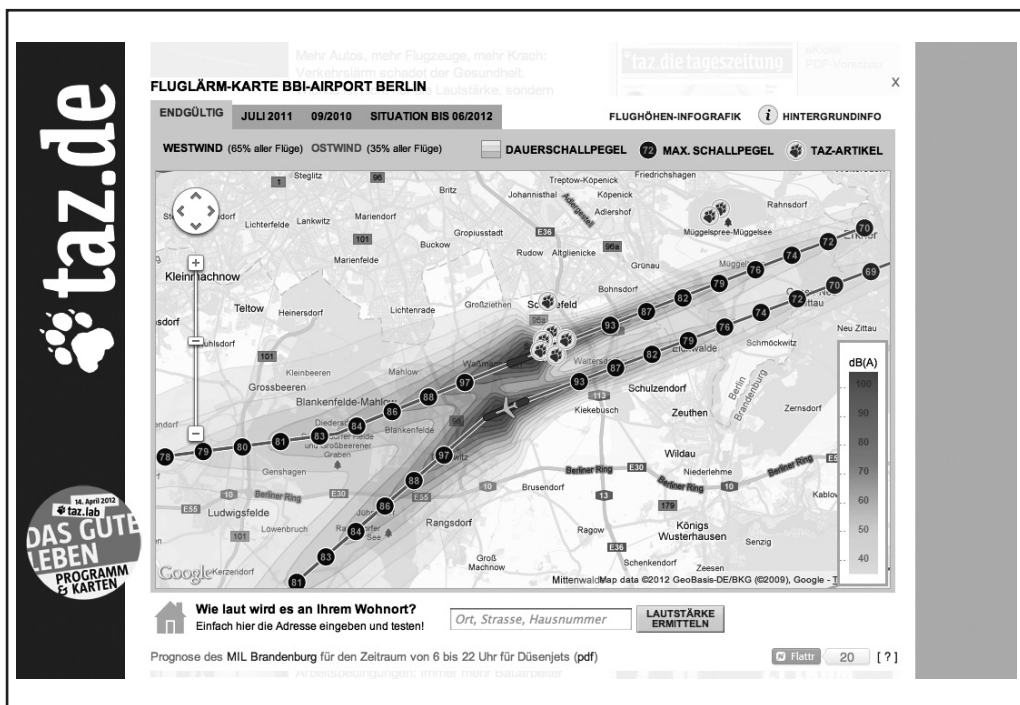
Нашите стории се во форма на компјутерски код

Организацијата „Опен дата сити“ (OpenDataCity) ја основавме кон крајот на 2010 година. Во тоа време, во Германија не се случуваше ништо што би наликувало на дата-новинарство.

Зошто го работиме ова? Многу пати слушаме како луѓето што работат во весниците и радиодифузните медиуми велат: „Не, не сме подготвени да основаме единица посветена на дата-новинарството во нашата редакција. Сепак, радо би прифатиле понуда некој друг да ја работи таа работа за нас“.

Колку што сме запознаени, ние сме единствената специјализирана компанија што се занимава исклучиво со дата-новинарство во Германија. Во моментот сме тројца: двајца од нас имаат искуство во новинарството, а третиот партнер има одлични познавања во областа на пишувањето компјутерски код и визуелизацијата. Соработуваме со неколку слободни хакери, дизајнери и новинари.

Во последниве 12 месеци спроведовме проекти од областа на дата-новинарството со четири весници и им понудивме обука и консултантски услуги на повеќе медиумски работници, лица што научно ги проучуваат медиумите, и новинарски школи. Првата апликација што ја подготвивме беше „ТАЗ“ (TAZ), интерактивна алатка за бучавата што се слуша околу новиот аеродром во Берлин. Следниот проект што вреди да се спомне беше апликацијата за задржување на податоците за користењето на мобилниот телефон од страна на еден германски политичар, што го работевме заедно со онлајн-изданието на „Цајт“ (ZEIT Online). За тој проект, во Германија ја добивме наградата „Гриме онлајн“ (Grimme Online Award) и наградата „Лид“ (Lead Award), како и наградата за онлајн-новинарство на „Асоцијацијата за онлајн-новинарство“ (Online Journalism Association) од САД. Во времето на пишувањето на тој проект, веќе нè чекаше работа на неколку други проекти, од едноставни интерактивни инфографики до дизајнирање и развивање еден вид „посреднички софтвер“ (middleware) за дата-новинарство.



Слика 2-13. Мапа на букавајта на Аеродромот во Берлин (Taz.de)

Се разбира, наградите помагаат да изградите добра репутација. Сепак, кога зборуваме со издавачите кои треба да ги одобрат проектите, нашата аргументација не се фокусира на тоа дека треба да инвестираат во дата-новинарството затоа што тоа ќе им донесе награди. Таквите инвестиции, пред сè, се однесуваат на привлекувањето на вниманието на публиката во подолг временски период, на одржлив начин. Т.е., да градите нешто поради неговото долгорочно влијание, не поради вест која ќе биде забораена по само неколку дена.

Следат три аргументи што ги имаме користено да ги убедиме издавачите дека треба да се зафатат со подолгорочни проекти:

Проектирањето во областа на дата-новинарството немаат рок на траење

Во зависност од нивниот дизајн, на апликациите за дата-новинарство може да им се додава нов материјал. Тие не се наменети само за корисниците, туку можат да се користат и во интерното известување и во анализа на функционирањето. Ако сте загрижени поради можноста и конкуренцијата да има корист од вашата инвестиција, можете да задржите некои функции или податоци само за интерна употреба.

Можеће да го надградувате минатиот тирод

Кога спроведувате проект по принципите на дата-новинарството, често креирате делови од компјутерски код што можат да се надградат и повторно да се употребат. Тоа може да го преполови времето што е потребно за следниот проект, затоа што подобро ќе знаете што треба (и што не треба) да правите, а и ќе ги имате делчињата што ќе ви послужат како основа.

Инвестициите во дата-новинарството се исплатливи

Проектите водени од податоци се поевтини од традиционалните маркетиншки кампањи. Онлајн-медиумите често инвестираат во проекти какви што се „Оптимизација за пребарувачите на интернет“ [Search Engine Optimisation (SEO)] или „Маркетинг на пребарувачите на интернет“ [Search Engine Marketing (SEM)]. Објавениот проект заснован на податоци, во нормални услови, носи многу кликови и интерес, а може да стане и вирален. Издавачите, по правило, вложуваат помалку во таквите активности отколку во обидите да се постигне истото внимание со кликови и линкови преку SEM.

Она што го работиме не се разликува многу од работата на другите медиумски агенции: обезбедуваме апликации или услуги за информативните медиуми. Разликата можеби е во тоа што ние себеси се гледаме, пред сè и над сè, како новинари. Ако нè прашате нас, ние произведуваме статии или стории, иако не во форма на зборови и слики, аудио- или видеоснимки, туку во форма на компјутерски код. Кога говориме за дата-новинарството, ние нужно говориме за технологија, софтвер, уреди и како да ги искористиме во пренесувањето на приказните.

На пример, тукушто завршивме една апликација која, во реално време, извлекува податоци од веб-страницата на германската железница со помош на „стругач“, овозможувајќи ни го развојот на интерактивен „Монитор на железничкиот сообраќај“ за „Зидојче цајтунг“ (Train Monitor for Süddeutsche Zeitung), што ги покажува доцнењата на возовите на долгите релации во реално време. Податоците во апликацијата се ажурираат секоја минута, а обезбедуваме и програмски интерфејс на апликацијата (API). Тој проект го почнавме пред неколку месеци и веќе собравме огромна група податоци што расте од час во час. Сега веќе се работи за стотици илјади редови податоци. Проектот им овозможува на корисниците да ги прегледаат податоците во реално време и да ги истражуваат архивите за претходните месеци. Приказната што ја пренесуваме на крај ќе биде дефинирана во голема мерка од индивидуалните акции на корисниците.

Во традиционалното новинарство, поради линеарниот карактер на пишаните или радиодифузните медиуми, мораме да мислиме на почетокот и крајот, на текот на приказната, на должината и аголот на нашата статија. Кај дата-новинарството, сето тоа изгледа малку поинаку. Да, и тука има некаков почеток. Луѓето ја посетуваат веб-страницата и добиваат прв впечаток за интерфејсот. Но, од тој момент натаму се оставени сами на себе. Можеби ќе останат една минута, можеби половина час.

Наша задача, како новинари што се занимаваат со податоци, е да ја обезбедиме потребната рамка или опкружување. Освен програмирањето и сегментите на управување со податоците, мораме да смислиме како најдобро да ги дизајнираме искуствата. Корисничкото искуство (User Experience UX) главно произлегува од (графичкиот) кориснички интерфејс (Graphical User Interface (GUI)). Во крајна линија, тоа е точката на која се гледа успехот или пропаста на еден проект. Можете да го имате најдобриот компјутерски код како основа за возбудлива група на податоци. Но ако не чини насловната страница, никому нема да му е грижа за кодот.

Има уште многу за учење и експериментирање. За среќа, тука е индустријата на видеоигри која во последниве неколку децении вовела многу иновации поврзани со дигиталните наративи, екосистеми и прикази. Значи, кога развиваме апликации за дата-новинарство, треба да гледаме како функционира дизајнот на компјутерските игри и како тие ги раскажуваат приказните. Зошто незадолжителните игри како „Тетрис“ (Tetris) се толку забавни? И зошто отворените светови во игрите во кои играчот може да ја менува и да влијае на околината, како што се „Гренд тефт ауто“ (Grand Theft Auto) или „Скајрим“ (Skyrim), се толку возбудливи?

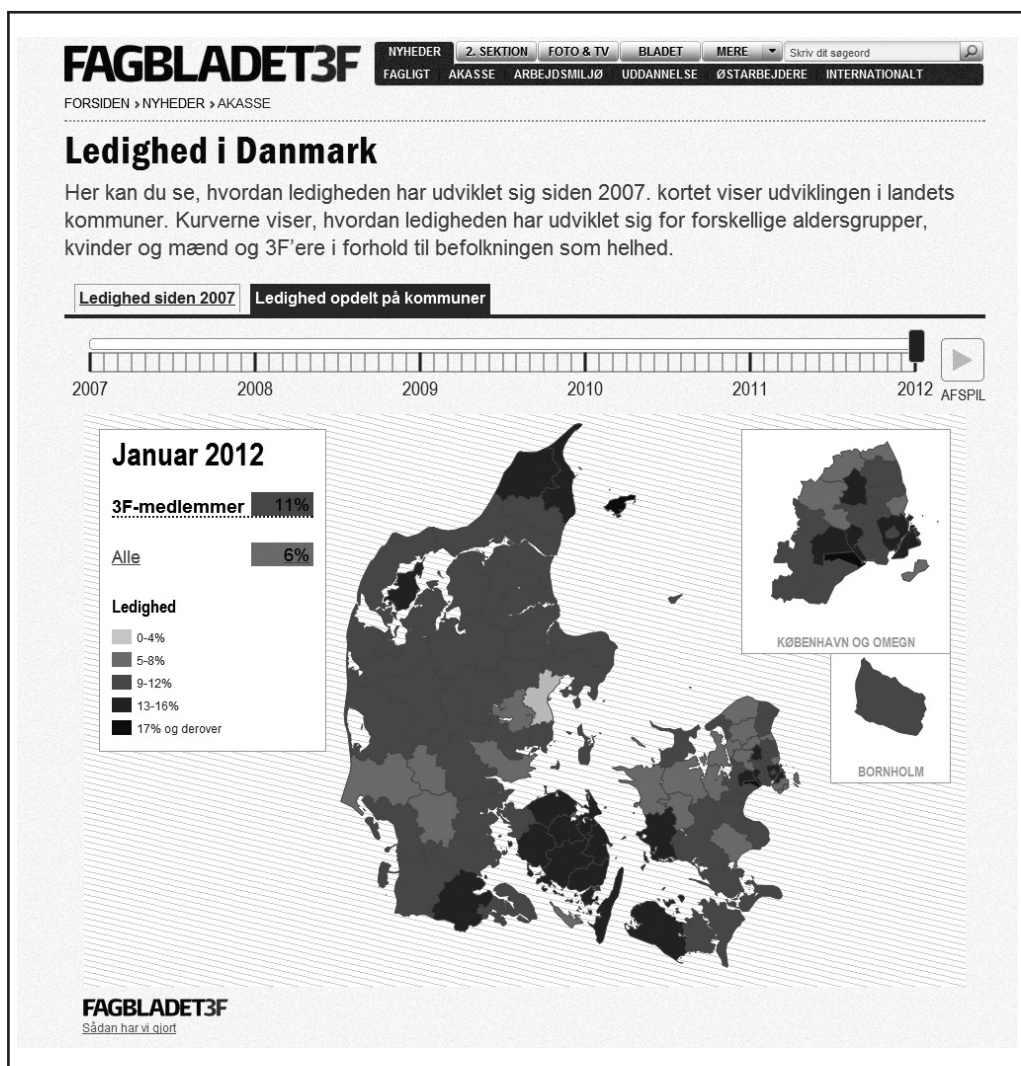
Ние сметаме дека дата-новинарство ќе остане со нас долго време. За неколку години, дата-новинарството ќе биде природен дел од редакциите, од проста причина што новинските веб-сајтови ќе мора да се променат. Количеството податоци достапни за јавноста ќе продолжи да расте. За среќа, новите технологии и натаму ќе ни овозможуваат нови и нови начини на раскажување приказни. Некои од тие приказни ќе бидат засновани на податоци, а многу апликации и сервиси ќе бидат новинарски по својата природа. Она што е интересно е какви стратегии ќе развијат редакциите за да го помогнат тој процес. Дали ќе решат да создадат тимови од новинари кои работат со податоци што ќе бидат интегрален дел од редакцијата? Дали ќе имаат одделенија за истражување и развој (R&D), нешто како „старт-ап“-оддели внатре во компаниите? Или делови од работата ќе бидат доделени на специјализирани компании како надворешни добавувачи? Сè уште сме на самиот почеток и само времето ќе каже што и како ќе биде.

— *Лоренц Мацаџи (Lorenz Matzat), OpenDataCity*

Кас и Мулвад: Полуготови содржини за интересните групи

Медиумите основани од лица што се заинтересирани за некоја тема е сектор во настанување, главно запоставен од теоретичарите на медиумите, а можат да имаат огромно влијание како преку мрежите на интернет така и со обезбедување содржини за информативните медиуми. Можеме да ги дефинираме како (вообичаено онлајн) медиуми контролирани од заинтересирани лица, организации или институции, а се користат за промовирање определени интереси или заедници. Невладините организации, на пример, вообичаено основаат такви медиуми; но, такви медиуми основаат и организирани групи потрошувачи, професионални здруженија, синдикати итн. Клучното ограничување на нивното влијание врз јавното мислење или другите заинтересирани страни најчесто е недостигот на капацитет за откривање значајни информации, недостиг кој е поголем дури и од сличниот проблем што го имаат информативните медиуми кои се соочуваат со намалени редакции. Кас и Мулвад (Kaas & Mulvad), данска профитна корпорација, е едно од првите истражувачки медиумски претпријатија што обезбедува стручни капацитети наменети, пред сè, за таквите медиуми на интересни групи.

Фирмата е основана во 2007 година со издвојување од „Данскиот институт за известување со помош на компјутер“ (Danish Institute for Computer-Assisted Reporting (Dicar)), непрофитна организација што им продаваше истражувачки стории на медиумите и обучуваше новинари на полето на анализа на податоците. Основачите, Томи Кас (Tommy Kaas) и Нилс Мулвад (Nils Mulvad), претходно работеа како новинари во информативната индустрија. Нивната нова фирма нуди нешто што тие го нарекуваат „податоци плус новинарски сознанија“ (содржина што останува полуобработена и на која ѝ треба доработка или дополнително уредување), пред сè, на медиумите на интересните групи, кои ја финализираат содржината како нови изданија или стории што ги дистрибуираат преку информативните медиуми или преку сопствените медиумски активности (како што се интернет-страниците). На списокот на директни клиенти имаат владини институции, фирми за односи со јавноста, синдикати и невладини организации каква што е „ЕУ Транспаренси“ (EU Transparency) или „Светскиот фонд за природа“ (World Wildlife Fund). Работата на невладините организации вклучува следење на субвенциите за земјоделците и одгледувачите на риба, редовни извештаи за активностите на лобистите во ЕУ подготвени со „стругање“ на соодветните интернет-страници. Индиректните клиенти ги вклучуваат фондациите што финансираат проекти на невладините организации. Фирмата соработува и со индустријата за производство на вести; еден таблоиден весник, на пример, ги купува нивните сервиси за следење на славните личности.



Некои од нивните проекти во областа на дата-новинарството се:

Мапа на невработеноста за 3Ф (3F)

Визуелизација на податоци со клучните статистики за невработеноста во Данска, подготвена за данскиот Синдикат на неквалификувани работници „3Ф“.

Животни услови за 3Ф

Уште еден проект за синдикатот „3Ф“ кој ги покажува разликите во животните услови во различни делови на Данска. Мапата користи 24 различни показатели.

Задолженоста за „Угебревет А4“ (Ugebrevet A4)

Проект за пресметување на „индексот на задолженост“ и визуелизација на разликите во приватниот сектор.

Опасни објекти во Данска (Dangerous Facilities in Denmark)

Проект за мапирање и анализа на опасните построенија во близина на детските градинки и други институции за целодневна грижа, спроведено за „Бјорн и Унге“ (“Børn&Unge”), магазин што го издава Данскиот сојуз на едукатори за деца и млади (оригиналната кратенка е „BUPL“).

Податоци за корпоративната одговорност за „Веспас“

Визуелизација на податоците за корпоративната одговорност во пет области, подготвена за данскиот производител на ветреници „Вестас“ (Vestas), со текст што се генерира автоматски. Се ажурира автоматски, секое тримесечие, со податоци извадени од 400 веб-страници од целиот свет сведени на една единствена производна единица.

Мапа на презимињата за „Експеријан“ (Experian)

Внесете го вашето презиме и проверете како е дистрибуирано во различни географски региони во Данска.

Мапа на „смешковци“ (Smiley Map) за „Екстра бладеј“

Секој ден, „Кас и Мулвад“ ги извлекуваат податоците за сите инспекции што откриле неправилности во производството или продажбата на храна и подготвуваат мапа за потребите на данскиот таблоид „Екстра бладеј“ (Ekstra Bladet) (мапата можете да ја видите некаде на средината на интернет-страницата на весникот).

„Кас и Мулвад“ не се првите новинари што соработуваат со медиуми на интересни групи. Еколошката организација „Гринпис“ (Greenpeace), на пример, рутински ангажира новинари како соработници во подготовката на извешта-

ите. Сепак, не знаеме за друга фирма што нуди услуги засновани на податоци за медиумите на интересните групи; многу почесто се работи за индивидуални новинари што соработуваат со индивидуални невладини организации како известувачи, уредници или пишувачи. Во моментот, медиумите што известуваат со помош на компјутери се фокусирани на пребарување и објавување (сетете се на „Викиликс“); „Кас и Мулвад“ повторно се јавуваат како иноватори поради нивниот фокус на анализата на податоците. Нивниот пристап бара, освен програмерски вештини, и разбирање кој вид информации има најголеми шанси да произведе влијателна сторија. Она што сигурно можеме да го кажеме е дека сите оние што сакаат да ги пресликаат нивните услуги веројатно ќе мора да ги стекнат тие две вештини преку некој вид партнерство, затоа што ретки се лицата што ги поседуваат и двете вештини.

Процеси: Иновативни информатички технологии плус анализа

Фирмата спроведува околу 100 проекти годишно, во траење од неколку часа до неколку месеци. Таа постојано инвестира во проекти што ги прошируваат нејзиниот капацитет и нејзината понуда. Услугата на следење на славните е еден од проектите што почна како експеримент. Друг експеримент вклучуваше „стругање“ на интернет во потрага по вести за заплени домови поради неплатени рати од кредити и подготовка на мапа на која се бележеа таквите случаи. Партнерите велат дека прв критериум во изборот на проектите е дали тие уживаат во работата и дали ќе научат нешто од проектот; потрагата по потенцијален пазар почнува дури по дефинирањето на новата услуга. Тие велат, без размислување, дека им било тешко да развиваат нови методи и нови бизниси додека работеле во индустријата на вести.

Мулвад коментира:

Немаме уредници или шефови што решаваат кои проекти можеме да ги работиме, кој софтвер или хардвер можеме да го купиме. Алатите ги набавуваме според потребите на проектот, на пример, најдобрите решенија за „стругање“ на текстови или „рударење“ по податоци. Сакаме да станеме предводници на полето на нови решенија во тие области. Се обидуваме да пронајдеме клиенти што се подготвени да ги платат нашите услуги, но, ако проектот ни се види забавен, ќе го спроведеме и за пониска цена.

Создадена вредност: Лични и компаниски брендови и приходи

Вкупниот обрт во 2009 година изнесуваше приближно 2,5 милиони круни, или околу 336.000 евра. Фирмата, исто така, ја одржува репутацијата на основачите како водечки новинари, а како резултат на таа репутација се одржува постојаната побарувачка за нивните услуги како предавачи и говорници. За возврат, нивните јавни настапи го поддржуваат брендот на фирмата.

Клучни сознанија од овој пример

- Кризата на намален капацитет на индустријата на вести истовремено е и криза на недоволната искористеност на капацитетите. Кас и Мулвад мораа да ја напуштат индустријата на вести за да го работат она што сметаат дека вреди да се работи, и тоа им се исплатува. Ништо не ги спречува новинските организации да ја усвојат таа вредност во својата работа.
- На некои пазари сигурно постои исплатлив пазар за „полуготови“ содржини што им служат на интересите на интересните групи.
- Сепак, таквата можност го раѓа прашањето колкава контрола новинарите можат да задржат врз презентацијата и користењето на нивниот труд од страна на трети лица. Да потсетиме дека тоа прашање веќе постои внатре во индустријата на вести (таму уредниците можат да наметнат промени на новинарскиот производ), а тоа прашање постоеше и во другите медиумски гранки (во филмската индустрија, на пример, каде кавгите помеѓу режисерите и студијата околу „финалната монтажа“ се далеку од ретка појава). Тоа не значи дека медиумите на интересните групи се коцкаат со ризиците, но таа дилема нема ни да исчезне туку така. Ќе мора да се обрне повеќе внимание на етичките аспекти на таа растечка реалност и на растечкиот пазар.
- Од гледна точка на приходите, не е доволно да се има само еден производ или услуга. Успешните претпријатија што се занимаваат со контрола на општествените процеси би требало да го прифатат пристапот на одржување портфолио, во кој консултантските услуги, обуката, говорите и другите услуги ќе носат дополнителен приход како поддршка за главниот бренд.

— Приреген исечок од книжата *“Disruptive News Technologies: Stakeholder Media and the Future of Watchdog Journalism Business Models”* на Марк Ли Хантнер и Лук Н. Ван Вассенхофе (Mark Lee Hunter and Luk N. Van Wassenhove), *INSEAD Working Paper*, 2010.

Бизнис модели на дата-новинарството

Среде сиот интерес и сета надеж што се поврзуваат со новинарството засновано на податоци, се јавува едно прашање кое постојано ги интересира редакциите: кои се можните бизнис-модели?

Иако треба да сме претпазливи кога даваме предвидувања, навраќањето на неодамнешната историја и моменталната состојба во која се наоѓа медиумската индустрија ни дава извесен увид. Денес имаме бројни новински организации што имаат корист од усвојувањето нови приоди.

Поимите како што се „дата-новинарство“ и популарниот поим „наука за податоците“ можеби звучат како да се однесуваат на нешто сосема ново, но не мора да е така. Тие нови имиња се само начин да се означи поместувањето што добива на интензитет во последните неколку децении.

Многу новинари навидум се несвесни за големината на приходите што веќе се остваруваат преку собирање, анализа и визуелизација на податоците. Се работи за бизнис на рафинирање на информациите. Алатките и технологиите за обработка на податоци овозможуваат расветлување на мошне сложени прашања како што се меѓународните финансии, јавниот долг, демографските трендови, образованието итн. Поимот „деловно разузнавање“ опишува цел еден спектар од концепти на информатичките технологии што имаат цел да обезбедат јасен увид на она што се случува во трговските друштва и корпорации. Заработувачката на големите, профитабилни компании на нашава доба, како што се „Мекдоналдс“ (McDonalds), „Зара“ (Zara) и „Ејч енд ем“ (H&M), зависи од постојаното следење на податоците. Сега засега добро им оди.

Она што во моментот се менува е фактот дека алатките што се развиваат за користење во тој простор стануваат достапни и за другите домени, вклучувајќи ги и медиумите. И, секако, постојат новинари што тоа го разбираат. Да го разгледаме примерот на „Табло“ (Tableau), компанија што произведува пакет на софтверски алатки за визуелизација. Или движењето „Биг дата“ (Big Data movement), каде технолошките компании користат софтверски пакети (често со отворен изворен код) за копање по купишта податоци од кои, во милисекунда, извлекуваат различни сознанија.

Тие технологии сега можат да се применат и во новинарството. Тимовите на „Гардијан“ и „Њујорк тајмс“ постојано ги поместуваат границите на ова поле во повој. Она што сега го гледаме е само врв на ледениот брег.

Прашањето е, како новинарството од тоа да направи пари? Големiot глобален пазар кој сега се отвора, сè уште зависи од трансформацијата на јавно достапните податоци во нешто што можеме да го обработиме: давање видливост и човечки лик на податоците. Сакаме да можеме да воспоставиме однос со големите броеви што секојдневно ги слушаме во вестите – што му значат на секој од нас сите тие милиони и милијарди.

Постојат неколку мошне профитабилни медиумски компании кои се водат од податоците, и тие само го применија тој принцип порано од сите останати. Тие уживаат во здрави стапки на раст и понекогаш навистина импресивни профити. Еден пример за таква компанија е „Блумберг“ (Bloomberg). Компанијата одржува околу 300.000 терминали и обезбедува финансиски податоци за своите корисници. Ако работите во бизнисот со пари, тоа е вистинската алатка за вас. Секој терминал доаѓа со тастатура со копчиња означени со различни бои и до 30.000 опции за преглед, споредба, анализа и помош во носењето на одлуката што понатаму. Таа основна дејност генерира заработувачка од

6,3 милијарди американски долари годишно – според проценката изнесена во една статија на „Њујорк тајмс“ од 2008 година (<http://www.nyti.ms/IQcRgY>). Како резултат, „Блумберг“ вработува новинари лево и десно, го купи загубарот со славно име „Бизнис вик“ (Business Week) итн.

Друг пример е канадскиот медиумски конгломерат што денес го знаеме под името „Томсон Ројтерс“ (Thomson Reuters). Почнаа со еден дневен весник, потоа купија неколку познати изданија во Велика Британија, а пред две децении решија целосно да излезат од бизнисот со весници. Наместо тоа, нивниот раст се заснова на информативните услуги што на клиентите од неколку индустриски гранки им нудат продлабочена перспектива. Ако не сте сигурни како да заработите пари од специјализираните информации, ви советувам само да го прочитате делот за историјата во одредницата за компанијата на интернет-енциклопедијата „Википедија“ (Wikipedia).

Или да го земеме примерот на „Економист“ (Economist). Магазинот изгради одличен, влијателен бренд од своите медиумски сегменти. Истовремено, „Економист интелиџенс јунит“ (Economist Intelligence Unit) денес работи повеќе како консултантска фирма, известува за релевантните трендови и дава економски прогнози за безмалку сите држави во светот. Вработува стотици новинари и тврди дека опслужува околу 1,5 милиони клиенти во целиот свет.

Постојат и бројни специјализирани сервиси водени од податоци што си имаат своја „ниша“ и можат да послужат како инспирација: „еМаркетер“ (eMarketer) во САД нуди споредби, графикони и совети за сите што се интересираат за интернет-маркетинг; „Штифтунг варентест“ (Stiftung Warentest) во Германија, институција што се занимава со проверка на квалитетот на производите и услугите; во Германија ја имаме и „Статиста“ (Statista), новооснована компанија што помага во визуелизацијата на јавно достапните податоци.

Во целиот свет во моментот расте бран на нови организации во овој сектор што, нормално, покриваат широк дијапазон области; на пример „Тајметрик“ (Timetric, <https://timetric.com/>), компанија што има за цел „повторно да го размисли деловното истражување“, „Опен корпорейтс“ (OpenCorporates, <https://opencorporates.com/>), „Касаби“ (Kasabi, <http://blog.kasabi.com/>), „Инфочимпс“ (Infochimps, <http://www.infochimps.com/>) и „Дата маркет“ (Data Market, <http://datamarket.com/>). За многу од нив слободно можеме да кажеме дека во основа се работи за експерименти, но земени заедно можат да се сметаат како знак на промените што доаѓаат.

Потоа ги имаме јавните медиуми, заспаниот цин во областа на новинарството засновано на податоци. Во Германија, годишно во овој сектор влегуваат 7,2 милијарди евра. Новинарството е посебен производ: ако е добро сработено, тогаш не станува збор само за заработување пари, туку и за исполнување значајна општествена улога. Кога еднаш ќе стане јасно дека дата-новинарството обезбедува подобри, поверодостојни сознанија до кои се доаѓа на полесен на-

чин, дел од тие средства може ќе биде искористен за отворање нови работни места во редакциите.

Кај дата-новинарство не се работи само за тоа да се биде прв, туку и да се биде извор на информации на кој му се верува. Во повеќеканалниот свет во кој живееме, лесно е да се привлече големо внимание. Довербата, од друга страна, е сè поограничен ресурс. Новинарите што се занимаваат со дата-новинарство можат да помогнат во споредбата, синтезата и презентацијата на разновидни и често тешко разбирливи извори на информации на начин што на читателите им дава вистински увид во комплексните прашања. Наместо да рециклираат соопштенија и да прераскажуваат приказни што ги слушнале од друг, дата-новинарите на читателите можат да им понудат јасен, разбирлив и, по можност, приспособлив поглед на нештата преку интерактивни графики и директен пристап до примарните извори. Тоа не е мала работа и си има своја вредност.

Значи, кој е најдобриот приод за амбициозните новинари кои сакаат да се занимаваат со податоци што ќе им овозможи влез на ова поле и можност да ја убедат управата да поддржи такви иновативни проекти?

Првиот чекор секако би требало да вклучува разгледување на можностите во непосредната околина: би се рекло, јаболката на ниските гранки што се на дофат на раката. На пример, можеби веќе имате на располагање збирки структурирани текстови и податоци што можете да ги искористите. Одличен пример за тоа е „Базата на податоци за убиства“ (Homicide database) на „Лос Анџелес тајмс“ (Los Angeles Times). Таму податоците и визуелизацијата ја претставуваат самата срцевина, не се нешто што искрснало отпосле. Уредниците ги собираат податоците за сите кривични дела што ќе ги пронајдат и пишуваат статии врз основа на нив. Со текот на времето, таквите збирки само растат, се продлабочуваат и стануваат сè повредни.

Можеби нема да успее од прва, но, со време, успехот ќе си дојде сам по себе. Еден показател што влева надеж е дека „Тексас трибјун“ и „ПроПублика“, две компании за кои можеме да тврдиме дека се појавија по ерата на печатените медиуми, известиле дека финансирањето на нивните непрофитни новинарски организации ги достигна проектираните нивоа многу порано од планираното.

Познавањето на сите аспекти на податоците – било во општа смисла или како специјалист фокусиран на еден аспект на синџирот на исхрана во работата со податоци – нуди вредна перспектива за оние што веруваат во новинарството. Токму како што еден познат германски издавач кажа во едно неодамнешно интервју: „Сега постои и таа нова група луѓе што себеси се нарекуваат дата-новинари. Тие веќе не сакаат за нив да останат само ронките“.

– Мирко Лорени, „Дојче веле“

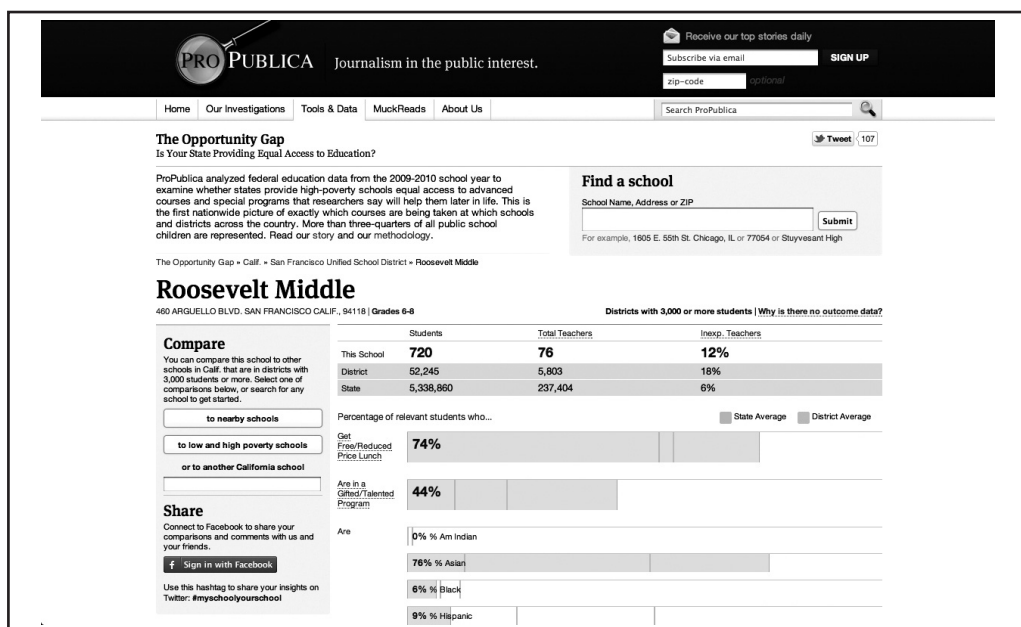
Студии на случаи



Во ова поглавје ќе го насочиме погледот на она што се случуваше зад сцената во неколку проекти од областа на дата-новинарството – од апликации развивани за само еден ден од истражувања што траеле цели девет месеци. Ќе видиме како изворите на податоци биле искористени за подобрување на известувањето за секакви можни теми, од избори до јавната потрошувачка, од улични немири до корупција, од успехот што го постигнале училиштата до цената на водата. Освен од големи медиумски организации какви што се „Би-Би-Си“, „Чикаго трибјун“, „Гардијан“, „Фајненшал тајмс“ (Financial Times), „Хелсингин саномат“ (Helsingin Sanomat), „Ла насион“ (La Nación), „Волстрит журнал“ (Wall Street Journal), и „Цајт онлајн“ (Zeit Online), ќе учиме и од помали иницијативи какви што се „Калифорнија воч“ (California Watch), „Хак/Хакерс Буенос Аирес“ (Hack/HackersBuenos Aires), „ПроПублика“ и една група локални граѓански новинари од Бразил што настапуваат под името „Пријателите на Жануарија“ (Friends of Januária).

Разлики во можностите (The Opportunity Gap)

Организацијата „Опортјунити гап“ (Opportunity Gap - Разлики во можностите) користеше претходно необјавени податоци за граѓанските права на Министерството за образование на САД и преку нив покажа дека некои држави, на пример, Флорида, успеале да обезбедат еднакви услови и за богатите и за сиромашните ученици, нудејќи им приближно еднаков пристап до училишните предмети што се предаваат на напредно ниво. Други држави, на пример, Канзас, Мериленд и Оклахома, нудат помалку можности во околните каде што живеат семејства со пониски просечни приходи.



Слика 3-1. Проектот „Опортјунити гап“ (ПроПублика)

Беза вклучени податоците за сите училишта во околните што имаат 3.000 или повеќе ученици. Во истражувањето беза претставени повеќе од три четвртини од сите деца што посетуваат јавни училишта. Еден новинар од нашата редакција ги собра, а директорот на одделението за известување со помош на компјутер детално ги прочисти податоците.

Проектот траеше околу три месеци. Сè на сè, шест луѓе работеа на приказната и на изработката на новинската апликација: двајца уредници, еден новинар, едно лице задолжено за ИПК и двајца девелопери. За повеќето од нас тоа не беше единствениот проект на кој работевме во тој период.

Проектот навистина бараше комбинација од вештини: добро познавање на областа, разбирање на најдобрите практики во работата со податоци, дизајнирање и пишување компјутерски код итн. Уште позначајно, проектот бараше способност да се пронајде приказна во податоците. Беа потребни и уреднички интервенции, не само во приказната, туку и во новинската апликација.

За чистење и анализа на податоците користевме главно „Ексел“ и скрипти за прочистување, како и софтверската апликација „МС Аксес“ (MS Access). Новинската апликација беше програмирана во софтверската рамка „Руби он реилс“ (Ruby on Rails) со обилно користење „Јаваскрипт“ (JavaScript).

Покрај статијата со основните наоди, известувањето се состоеше и од интерактивна апликација што им овозможи на читателите да ја разберат оваа огромна група податоци собрани на национално ниво и да ги пронајдат примерите во него со кои можат да се идентификуваат. Со нашата апликација, читателите можеа да го пронајдат своето локално училиште – на пример, Централното средно училиште (Central High School) во Њуарк, Њу Џерси – и веднаш да видат како му оди на училиштето во различни области. Потоа можеа да го притиснат копчето „Спореи со училишта од побогати и посиромашни окрузи“ (Compare to High and Low Poverty Schools) и да ги разгледаат податоците за други средни училишта, нивната релативна сиромаштија и во која мера нудат предмети како виша математика, напредна наставна програма и слично. Во нашиот пример, Централното средно училиште се граничи со Средното училиште „Милбурн“. „Опортјунити геп“ покажува дека само еден отсто од учениците во „Милбурн“ добиваат бесплатен или топол оброк по намалена цена, додека 72 отсто посетуваат барем еден предмет од програмата за напредна настава. На другиот крај од спектарот, 85 отсто од учениците во средното училиште „Интернешнал хај“ (International High) добиваат бесплатен оброк или оброк по намалена цена, а само еден процент од нив посетуваат предмети од програмата за напредна настава.

Преку овој пример читателите можат да искористат нешто што добро го познаваат – локалното средно училиште – за да разберат нешто што не го знаат: како е распределен пристапот до образование, но и степенот до кој сиромаштијата е фактор за предвидување на пристапот.

Апликацијата ја интегриравме во „Фејсбук“, така што читателите можеа да се пријават на „Фесјбук“, а нашата апликација автоматски ги известуваше кои училишта би можеле да бидат интересни за нив.

Сите наши новински апликации привлекуваат одличен сообраќај, а особено сме горди на начинот на кој оваа апликација раскажува една комплексна приказна; поточно, им помага на читателите сами да ги раскажат своите приказни.

Како и кај повеќето проекти засновани на податоци што ги чуваат властите, на податоците им е потребно многу прочистување. На пример, иако има само околу 30 можни предмети во програмата за напредна настава, некои училишта пријавија дека имаат стотици такви предмети. Поради тоа беше

потребно рачно да се проверат таквите податоци, како и многу телефонски разговори кога сметавме дека за некои податоци ни треба потврда или треба да бидат коригирани.

Напорно работевме за да бидеме сигурни дека апликацијата ќе ги раскаже и „далечните“ и „блиските“ приказни. Значи, апликацијата требаше на читателите да им обезбеди една широка, апстрактна приказна на национално ниво; можност да го споредат пристапот до образование во поединечните сојузни држави. Сепак, бидејќи таквите апстракции знаат да ги збунат читателите кои не можат да сфатат што значат таквите податоци на нив лично, сакавме читателите да можат да ги пронајдат училиштата што тие ги посетувале и да ги споредат со побогатите и посиромашните училишта во нивниот регион.

Мој совет за младите новинари што сакаат да се занимаваат со дата-новинарство е да ја познаваат материјата на која работат, да бидат љубопитни и да прашуваат! Сите правила што се применуваат во другите видови новинарство важат и овде. Мора правилно да се пренесат фактите, приказната треба да биде добро раскажана и – од клучно значење е вашата новинска апликација да биде усогласена со приказната што ја пишувате. Ако не се усогласени и постои расчекор помеѓу нив, или приказната или апликацијата ќе биде погрешна.

Исто така, ако сакате да научите да пишувате компјутерски код, најважно е да почнете да го работите тоа. Ако повеќе сакате да учите на курсеви, или со помош на книги или видео, тогаш најпрво размислете за некоја добра идеја за проект и за рокот за негово спроведување. Ако во глава имате некоја приказна што може да се објави само како новинска апликација, нема да ве спречи тоа што не знаете како да програмирате!

– Скоџ Клајн, ПроПублика (*Scott Klein, ProPublica*)

Деветмесечна истрага на европските структурни фондови

Во 2010 година, „Фајненшал тајмс“ и „Бирото за истражувачко новинарство“ [Bureau of Investigative Journalism (BIJ)] ги здружија силите во едно истражување на европските структурни фондови. Целта беше да се провери кои се корисниците на структурните фондови и да се види дали парите се користат за вистински цели. Со вкупно 347 милијарди евра доделени во период од седум години, Структурните фондови се втората по големина програма за субвенции во Европската Унија. Програмата постои веќе неколку децении, но, освен неколку широки и општи прегледи, малку се знаеше за тоа кој добил средства од таа програма. Како дел од промените во тековната рунда на респределување на фондовите, властите се обврзани да го објават списокот на корисници, заедно со опис на проектите и вкупната сума доделена од ЕУ и од националните фондови.



Слика 3-2. Истйражување на Структурниите фондови на ЕУ („Фајненшал тајмс“ и „Биро за истйражувачко новинарстйво“)

Проектниот тим работеше на проектот девет месеци и беше составен од најмногу 12 новинари и еден програмер вработен на полно работно време. Само собирањето на податоците траеше неколку месеци.

Од проектот произлегоа серија стории што „Фајненшал тајмс“ и БИН ги објавија во период од пет дена, еден радио документарец произведен од Би-Би-Си и неколку телевизиски документарци.

Пред да се зафатите со толку голем проект, мора да се осигурате дека наодите се оригинални и дека на крај ќе имате неколку добри стории што нема да ги има никој друг.

Процесот беше поделен во неколку засебни чекори.

1. Идентификување кој ги има податоците и како се чуваат

Генералниот директорат за региони на Европската комисија има свој портал кој води до интернет-страниците на регионалните тела што ги објавуваат податоците. Ние сметавме дека Комисијата би требало да одржува една сеопфатна база со проектни податоци до кои би можеле да пристапиме директно или преку барање за слободен пристап до информациите. Таква база на податоци не пос-

тои, барем не таква што нуди ниво на детали какви што ни беа потребни. Откривме дека многу од хиперлинковите што ги нуди Комисијата се неисправни, како и дека повеќето од овластените тела ги објавуваат податоците во „PDF“-формат, наместо во формати што се погодни за анализа, на пример, „CSV“ или „XML“.

Еден тим од 12 лица работеше на идентификација на најновите податоци и внесување на податоците во една унифицирана сметководствена табела. Знаејќи дека полињата за внесување на податоците не се униформни (на пример, насловите на категориите беа наведени на различни јазици, некои групи податоци користеа различни валути, а во некои табели податоците беа поделени на фондови на ЕУ и национални фондови), ни требаше максимална прецизност во пренесувањето и описот на внесените податоци од секоја поединечна група.

2. Снимање и подготовка на податоците

Следниот чекор се состоеше од снимање на сите сметководствени табели, „PDF“-документи и, во некои случаи, „стругање“ на оригиналните податоци од интернет.

Потоа секоја група податоци требаше да се стандардизира. За нас беше најважно да ги извлечеме податоците од „PDF“-датотеките, што понекогаш содржеа и по неколку стотини страници. Поголемиот дел од таа задача ја извршивме со користење на апликациите „UnPDF“ и „ABBY FineReader“ што овозможуваат извлекување на податоците во формати какви што се „CSV“ или „Excel“.

Исто така, требаше да се провери, и по неколку пати ако треба, дали алатките за извлекување од „PDF“-документи правилно ги фатиле податоците. Проверката ја направивме со филтрирање, сортирање и сумирање на вкупните износи (за да се осигураме дека соодветствуваа со податоците наведени во „PDF“-документите).

3. Креирање база на податоци

Програмерот подготви една база на податоци што работеше во системот „SQL“. Секој од подготвените документи беше искористен како елемент вграден во генералната „SQL“-база на податоци. Еднаш дневно, сите индивидуални датотеки беа внесувани во една голема „SQL“-база на податоци, која можеше да се пребарува на брзина, со клучни зборови, преку кориснички интерфејс.

4. Проверка и анализа

Тимот ги анализираше податоците на два основни начини:

Преку насловната страница на базата на податоци

Тоа се состоеше од пребарување со внесување клучни зборови (на пример, „тутун“, „хотел“, „компанија А“) во пребарувачот. Со помош на сервисот за преведување на „Гугл“ (Google Translate), вграден во функцијата за пре-

барување на нашата база на податоци, клучните зборови беа преведувани на 21 јазик и се добиваа соодветни резултати. Резултатите од пребарувањето можеа да се снимаат за потоа новинарите да продолжат понатаму да истражуваат за потребите на нивните проекти.

Со макроанализа на целата база на податоци

Повремено ќе снимевме цела група податоци која потоа можеше да се анализира (на пример, преку користење клучни зборови, со собирање на податоците по држава, регион, тип на трошок, број на проекти по корисник итн.) Информациите за нашите приказни ги добивавме со тие две методи, но и со теренско и „деск“-истражување.

Проверката на интегритетот на податоците (со собирање на податоците за доделените средства и споредба со она што овластените тела велеа дека било доделено) ни одземаше многу време. Еден од главните проблеми беше што овластените тела, во повеќето случаи, ги даваа само износите од „фондовите на ЕУ и националните фондови“. Според правилата на ЕУ, секоја програма може, со европски средства, да финансира само определен процент од вкупните трошоци. Нивото на учество на фондовите на ЕУ се определува, на ниво на програма, преку таканаречената стапка на кофинансирање. Секоја програма (на пример, програмата за регионална конкурентност) се состои од многубројни проекти. На ниво на проект, технички може да се случи некој проект да биде целосно финансиран, а друг да не добие ништо од европските фондови, сè додека, ставени заедно, европското финансирање на ниво на програма не ја надминува одобрената стапка на кофинансирање.

Тоа значеше дека треба да ги провериме сите поединечни износи доделени од ЕУ што ги наведовме сториите за компаниите што користат европски средства.

– Синџија О’Мурку, „Фајнениш џајмс“ (Cynthia O’Murchu, Financial Times)

Кризата на „Еврозоната“

Ние, значи, ја следиме кризата на „Еврозоната“. Секој дел од неа. Драмата што се одвива кога владите се расправаат едни со други, а животните заштеди одат во неповрат; реакциите на светските лидери, мерките на штедење и протестите против тие мерки. „Волстрит журнал“ секој ден објавува графикони што ги прикажуваат изгубените работни места, падот на БДП, падот на светските пазари. Бројките постојано растат до ниво кога станувате рамнодушни.

Уредниците на насловната страница („Page One“) свикаа состанок на кој расправавме за начинот на кој ќе известуваме за крајот на годината и, додека ја напуштавме просторијата по состанокот, се фатив сам себе како се прашу-

вам: како ли им било на оние што поминале низ сево ова? Дали сега е исто како во 2008, кога бев отпуштен од работа, а „црните“ вести не престануваа? Дома зборувавме за работните места, за работата и за парите секоја вечер, знаевме да заборавиме дека тоа би можело да ја вознемири ќерка ми. Најлошо беше за викенд. Се обидував да го одречам стравот што цврсто ме држеше за гуша и анксиозноста што ме стегаше во градите.

Дали истото сега им се случува на семејствата во Грција? Во Шпанија?

Се свртев и влегов во канцеларијата на Мајк Ален (Mike Allen), уредникот на насловната страница, и му ја понудив идејата да ја раскажеме приказната за кризата преку искуствата на семејствата во Еврозоната, така што најпрво ќе ги провериме статистичките податоци, ќе ги лоцираме демографските профили кои ќе помогнат да разбереме што, всушност, се подразбира како семејство и сето тоа ќе го прикажеме заедно со фотографии, илустрации и интервјуа, аудиоснимки на различните генерации. Ќе користиме убави портретни фотографии, снимени изјави – и податоци.

Кога се вратив на мојата работна маса, напишав кратко резиме и нацртав едно лого.



Слика 3-3. Резимето на „Кризата на „Еврозоната““ („Волстријџ журнал“)

Следните три недели трчав по бројки: статистики за браковите, стапката на смртност, големината на семејствата и здравствените трошоци. Читав за

заедничкиот живот и стапките на развод, прегледувајќи истражувања за благостојбата и стапките на штедење. Ги прегледувајќи податоците на националните заводи за статистика, разговарав со Одделот за население на ОН, ММФ, Евростат и ОБСЕ сè додека не пронајдов економист што целата кариера ја посветил на проучување на семејствата. Тој ми посочи една истражувачка што се занимава со составот на семејствата, а таа, од своја страна, ми укажа на повеќе „бели книги“ за темата што ме интересира.

Со помош на уредникот Сем Енрикез (Sam Enriquez) го стеснивме бројот на земји што ќе ги опфатиме. Собравме тим за разгледување на можните визуелни приоди, како и да размислиме кои новинари би можеле да ги произведат потребниот текст, звучните записи и сторијата. Мет Крег (Matt Craig), уредник на фотографија на насловната страница, се зафати со задачата да определи кои фотографии ќе работат на проектот. Мет Мареј (Matt Murray), заменикот на одговорниот уредник задолжен за вестите од светот, испрати допис до шефовите на дописништвата со барање нивните новинари да помогнат во реализацијата на проектот. (Таквата поддршка од управата беше од клучно значење).

Но, пред сè, податоците. Наутро ќе ги експортирав податоците во сметководствени табели и ќе направев графикони за да ги идентификувам трендовите: намалување на заштедите, исчезнување на пензиите, мајки што повторно бараат работа, здравствените трошоци, заедно со јавниот долг и невработеноста. Секое попладне ќе ги разгледувајќи тие податоци во гроздови, споредувајќи ги државите меѓу себе за да пронајдам некаква приказна.

ITALY_onlychildren.xlsx				
Distribution of households by number of children, 2007 (2)				
	1 child	2	3+	
Sweden	43.3	40.6	12.8	3.1
Finland	42.7	39.2	13.5	4.1
Denmark	41.3	43.4	12.5	2.1
Netherlands	38.8	42.7	14.1	4.1
France	45.3	39.9	11.7	3.1
Germany	48.6	39.5	9	1
Austria	50.1	37.2	10.2	2.1
Belgium	44.5	36.8	13.7	1
Luxembourg	44.8	46	8.1	1.1
Ireland	43.8	35.2	16	1
Italy	55.2	37.9	6.1	0.1
Spain	55.2	39.9	3.9	0.1
Portugal	61.4	33.7	4	1
Greece	46.4	47.9	4.3	1.1
Cyprus	42.5	46.8	8.5	2.1
Hungary	49.5	36.9	10.5	3.1
Estonia	58	32.9	7.5	1.1
Latvia	62.8	29.5	5.8	1.1
Lithuania	59.7	31.4	6.8	2.1
Slovenia	49.7	41.5	7.2	1.1
Slovakia	53.7	36	8.3	1
Poland	53.5	35.2	8.6	2.1
EU-25	49.5	38.9	9	2.1
EU-15	48.7	39.5	9.2	2.1
AME	52.5	36	9.2	2.1

Слика 3-4. Оценувањето на корисноста на некоја згрупа на податоци може да одземе многу време (Сара Слобин)

Тоа го правев цела недела пред целосно да се изгубам и да почнам да се сомневам во себе. Можеби грешев во пристапот. Можеби тоа не беше прашање за државите, туку се работеше за татковци и мајки, за деца и нивните баби и дедовци. Количеството податоци растеше. И се намалуваше, исто така. Понекогаш со часови собирав информации за тие на крајот да не ми кажат ништо. Да заклучам дека сум „риел“ низ напдно погрешена група на бројки. Едноставно, понекогаш се работеше за застарени податоци.

А количеството информации повторно се зголемуваше, кога сфатив дека сè уште имам неодговорени прашања и дека сè уште не ги разбирам семејствата.

Морав да разберам, да му дадам некаква форма. Затоа, набрзина направив серија графикони со апликацијата „Illustrator“ и почнав да ги преуредувам и аранжирам податоците.

Како што настануваа графиконите, пред очи ми излегуваше јасна и смислена слика на семејствата.

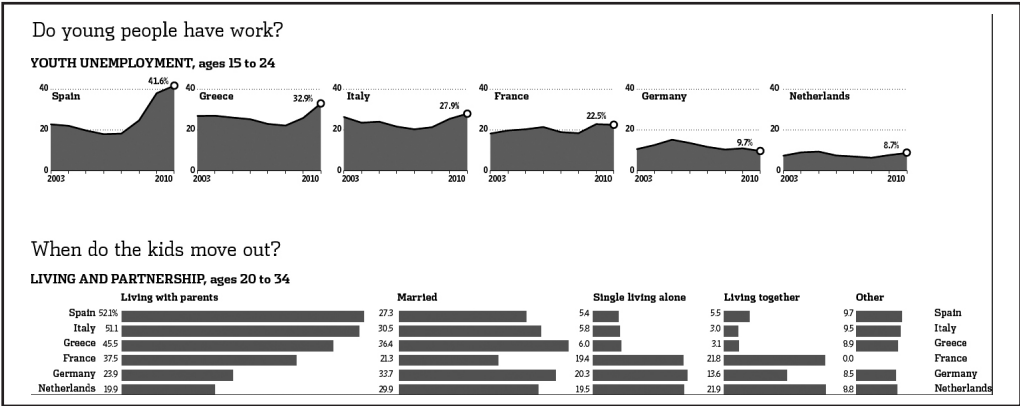
Почнавме со работа на проектот. Зборував со сите новинари поединечно. Им ги испратив графиконите, идејата во општи црти, со отворена покана да пронајдат приказни за кои тие мислат дека имаат значење, дека ќе им ја приближат кризата на нашите читатели. Ни требаше едно мало семејство во Амстердам, и поголеми семејства во Шпанија и Италија. Сакавме да слушнеме повеќе генерации за да видиме како личните искуства ги обликуваат одговорите.

Од тој момент натаму, станував рано за да ја проверам електронската пошта, поради временската разлика. Новинарите испраќаа прекрасни теми, резимеа и за мене неочекувани изненадувања.

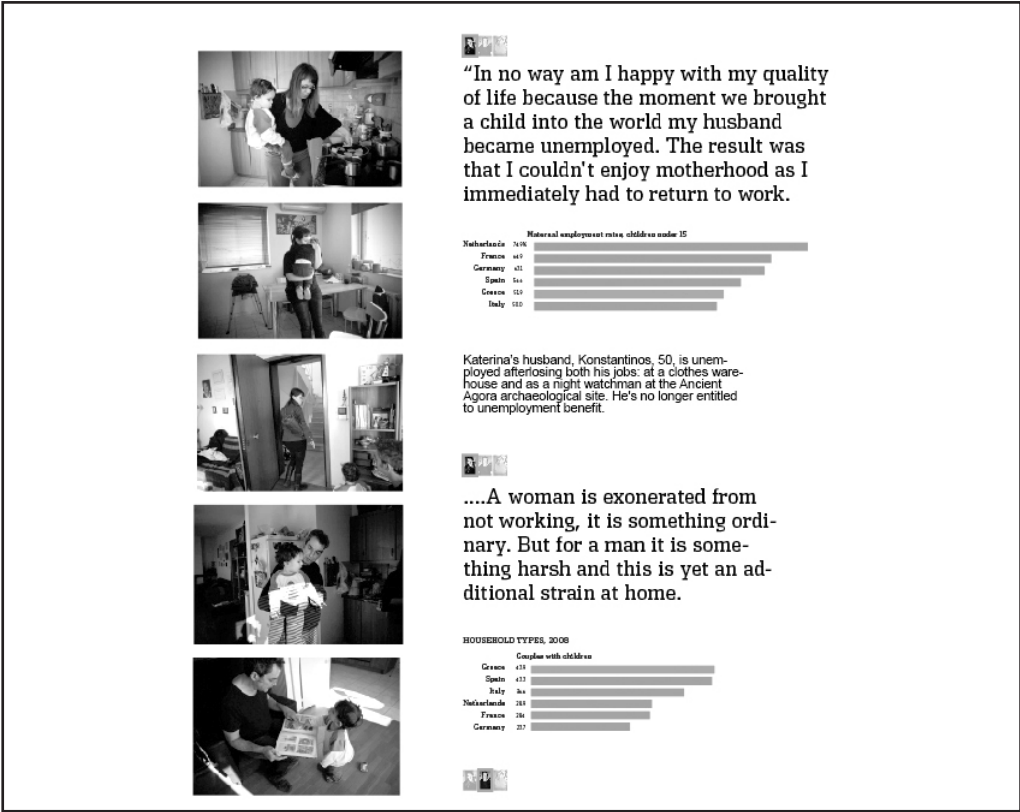
Што се однесува до фотографијата, знаевме дека сакаме портрети на сите генерации. Идејата на Мет беше неговите фоторепортери да го следат секој член на семејството во текот на еден ден од животот. Тој избра фото-репортери што известувале од целиот свет, па дури и од војни. Мет сакаше секоја од индивидуалните серии фотографии да заврши на масата за вечера. Сем сугерираше да ги прикажеме и менијата.

Потоа само требаше да чекаме и да видиме каква приказна ќе ни раскажат фотографиите. Да видиме што ќе ни кажат семејствата. Го дизајниравме интерактивниот приказ. Палетата на бои ја позајмив од еден стрип со Тинтин и работевме на интерактивните аспекти. И кога сè беше склопено и ги имавме нашите приказни, повторно додадовме неколку (но не многу) од оригиналните графикони.

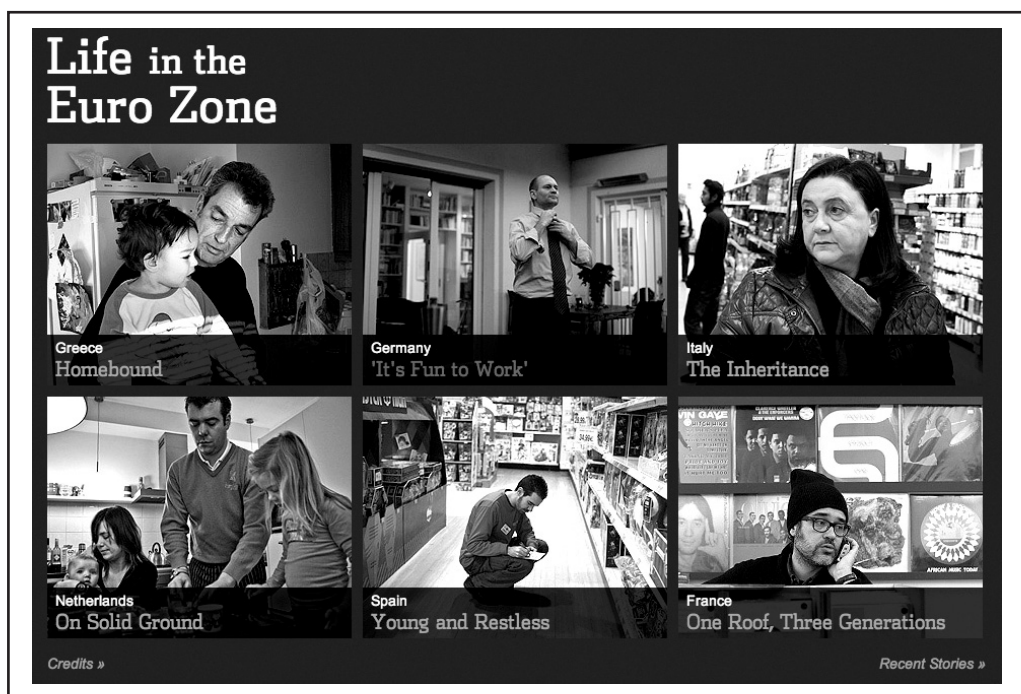
Сосема доволно да ја акцентираме секоја индивидуална приказна, да им дадеме острина на темите. Податоците служеа како пауза во приказната што ја кажувавме, можност за промена на брзината.



Слика 3-5. Графичка визуелизација: разбирање на трендовите и обрасците скриени во податоците (Сара Слобин)



Слика 3-6. Бројките претставуваат живи луѓе; вредноста на податоците лежи во индивидуалните приказни што ги раскажуваат („Волсирџиј журнал“)



Слика 3-7. Живоїтої во „Еврозоната“ („Волсїрийї журнал“)

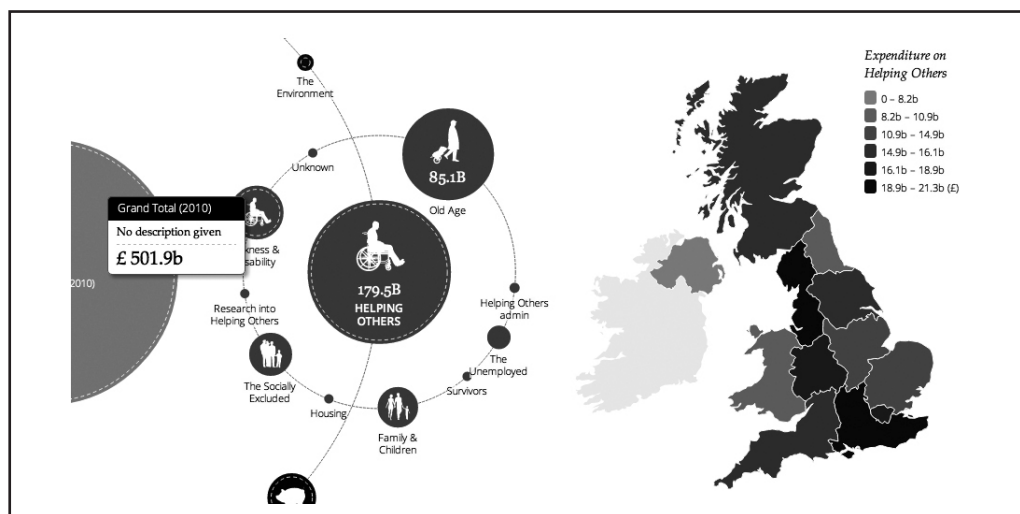
Најпосле, луѓето беа нашите податоци – нивните фотографии и приказни. Податоците ја вградуваа секоја приказна и ја одржуваа тензијата помеѓу државите.

До моментот на објавувањето, токму пред Нова година, додека размислувавме што нè чека зад хоризонтот, веќе ги знаев имињата на сите членови на сите семејства за кои известувавме. И ден-денес се прашувам што прават сега. И не ми е гајле ако вам не ви изгледа како проект за обработка на податоци. Затоа што успеавме, моментите документирани во „Животот во Еврозоната“, приказните на луѓето што се собираат на семејна вечера и зборуваат за работата и животот, да ги споделиме со нашите читатели. Тоа беше можно затоа што ги разбравме податоците.

– Сара Слобин, „Волсїрийї журнал“

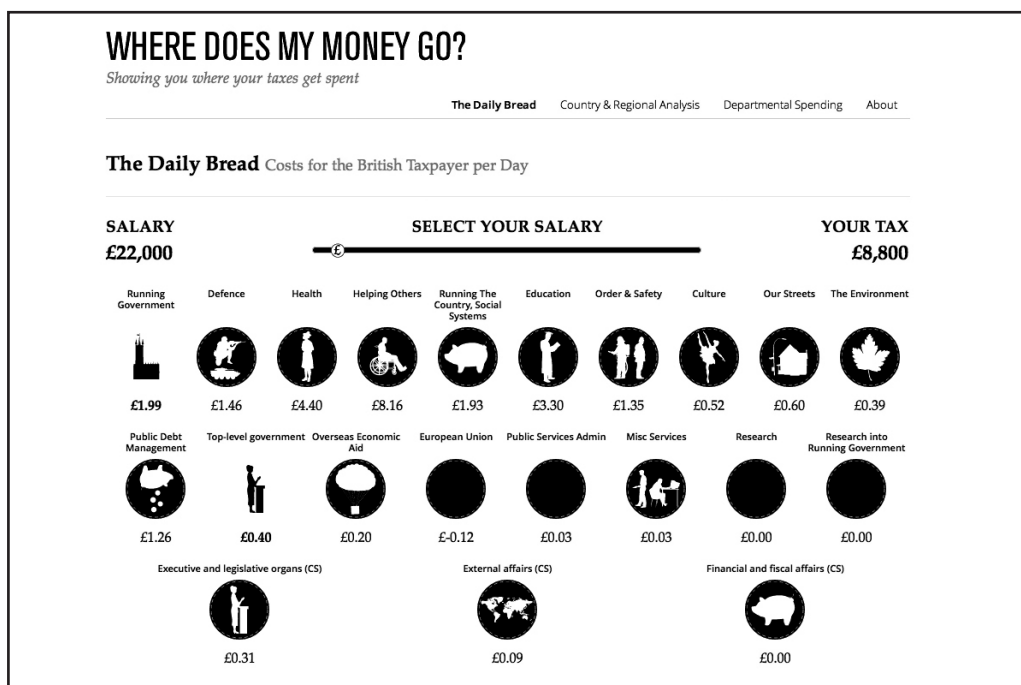
Известување за јавните буџети со OpenSpending.org

Во 2007 година, Џонатан дојде во Фондацијата „Отворено знаење“ (Open Knowledge Foundation) со краток предлог за проект наречен „Каде одат моите пари?“ (Where Does My Money Go?), кој имаше за цел да им помогне на граѓаните на Велика Британија да разберат како се трошат јавните фондови. Всушност, тоа требаше да биде проект за потврдување на концептот за еден поголем проект за визуелна презентација на јавните информации, врз основа на пионерските напори на „Институтот за изотипови“ (Isotype Institute) на Ото и Мари Нойрат (Otto and Marie Neurath) од 40-тите години на минатиот век.



Слика 3-8. Каде одат моите пари? (Фондација „Отворено знаење“)

Проектот „Каде одат моите пари?“ им овозможи на корисниците да ги прегледаат јавните податоци од широк дијапазон на извори со користење интуитивни алатки со отворен изворен код. Добивме една награда што ни помогна во развојот на прототипот на проектот, а подоцна ни беа доделени и фондови од програмата „4ИП“ (4IP) на „Канал 4“ (Channel 4) за тој прототип да прерасне во комплетна веб-апликација. Гуруто на информатичкиот дизајн Дејвид Мекендлес (David McCandless) од проектот „Информациите се прекрасни“ (Information is Beautiful; <http://www.informationisbeautiful.net/>) креираше неколку различни прикази на податоците што требаше да им помогнат на луѓето да ги разберат големите броеви – вклучувајќи го и приказот „Анализа на државно и регионално ниво“, кој покажува како се распределуваат парите во различни делови од државата, како и „Лебот наш насушен“ (Daily Bread), приказ на износите на даноците што граѓаните ги плаќаат секој ден, во фунти и пени.



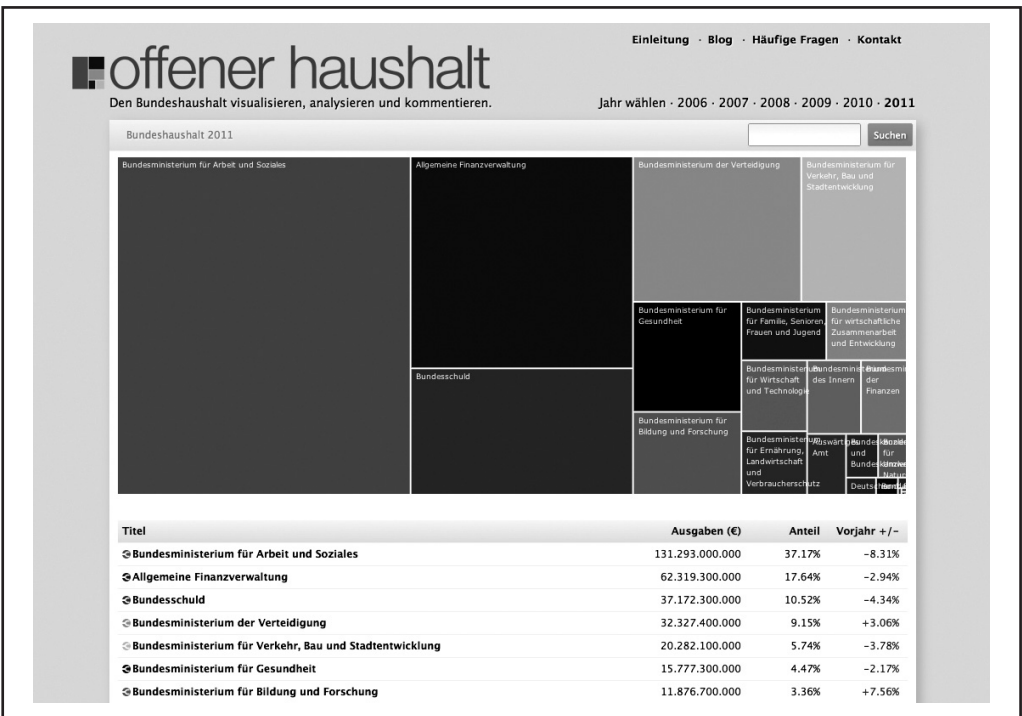
Слика 3-9. Калкулаторот на „Лебоџ наш насушен“ на проектот „Каге одаџ моиџе џари“ (Фондација „Ошворено знаење“)

Во тој стадиум, „светиот Грал“ на проектот беа податоците од системот „КОИНС“ (Combined Online Information System – COINS), најдеталната и најсеопфатна достапна база на податоци за финансиите на владата на Велика Британија. Во соработка со Лиза Еванс (Lisa Evans), пред таа да замине во „Датаблогот на Гардијан“, Џулијан Тод (Julian Todd) и Френсис Ирвинг (Francis Irving, денес познат по својата работа на проектот „Scraperwiki“), Мартин Розенбаум (Martin Rosenbaum) од Би-Би-Си и други, поднесовме голем број барања за пристап до податоците – многу од нив неуспешни (делумен приказ на целата таа сага дава Лиза во антрфилето „Користењето на слободниот пристап до информации за разбирање на трошењето“, во Глава 4 од овој прирачник).

Кога податоците најпосле беа објавени кон средината на 2010 година, тоа се сметаше за триумф на застапниците за транспарентност. Добивме можност податоците да ги добиеме однапред, за да можеме да ги внесеме во нашата веб-апликација, и тоа предизвика големо внимание на медиумите. На денот на објавувањето, дузини новинари се пријавија на нашиот канал на „ИРЦ“ (Internet relay chat IRC – буквалниот превод би гласел „разговор преку интернет“, заб.прев.) да дискутираат и да прашуваат за објавувањето, но и да поста-

ват прашање како да ги отворат и прегледаат податоците (датотеките беа со големина што се мери во десетици гигабајти). Иако некои експерти тврдеа дека обемното издание е толку комплицирано што транспарентноста всушност придонесува за опскурноста (obscurity through transparency), голем број храбри новинари нурнаа во податоците за да им понудат на своите читатели дотогаш невидена слика на трошењето на јавните фондови. „Гардијан“ организираше блогирање во живо за изданието, а голем број други медиуми известуваа за него и понудија свои анализи на наодите извлечени од податоците.

Не требаше да чекаме долго пред да почнеме да добиваме барања за можноста да спроведеме слични проекти и во други земји во светот. Непосредно по почетокот на работата на „Офенерхаусхалт“ (OffenerHaushalt) – верзијата на проектот за Државниот буџет на Германија креирана од Фридрих Линденберг (Friedrich Lindenberg) – ја отворивме и „Опенспендинг“ (OpenSpending, отворени трошоци), меѓународна верзија на проектот која имаше за цел на корисниците да им помогне во мапирањето на јавната потрошувачка во целиот свет, слично како што „Опен стрит мап“ (OpenStreetMap) им помогна во креирањето географски мапи. Со помош на талентираниот Грегор Ајш, применивме нов дизајн, делумно заснован на оригиналниот дизајн на Дејвид Мекендлес.



Слика 3-10. „OffenerHaushalt“, германската верзија на проектот „Kage ogai moite yari“ (Фондација „Отворено знаење“)

Во рамките на проектот „Опенспендинг“, соработувавме со новинари за да ги собереме, прикажеме, толкуваме и на јавноста да ѝ ги претставиме податоците за потрошувачката. „Опенспендинг“ е, пред сè и над сè, една огромна база на податоци за јавната потрошувачка што може да се пребарува – содржи информации за буџетот од повисок ред, како и податоци за реалните трошоци на ниво на индивидуални трансакции. На таа основа се изградени серија готови визуелизации како што се прикази со меури (bubble tree) и структурни прикази (treemap). Секој може да ги внесе податоците од локалната администрација и да ги визуелизира.

Иако на почетокот мислевме дека ќе имаме поголема побарувачка за некои од пософистицираните визуелизации, по разговорите на таа тема со неколку новински организации сфативме дека прво ќе треба да се задоволат некои основни потреби, како што е можноста за вградување динамички табели со податоци во нивните блогерски постови. Во желба да ги поттикнеме новинските организации, покрај нивните приказни, да обезбедат и јавен пристап до податоците, подготвивме и еден „виџет“ (widget, веб-уред, мала апликација за вршење на некоја специфична задача, заб.прев).

Првото издание на нашиот проект се совпадна со одржувањето на првиот Меѓународен фестивал на новинарството во Перуџа, Италија. Една група составена од девелопери, новинари и јавни службеници работеше на внесување на податоците за Италија во платформата „Опенспендинг“, што обезбеди визуелно богат приказ на јавните трошоци на државно, регионално и локално ниво. За тоа известија „Ил фато куотидијано“ (Il Fatto Quotidiano), „Ил пост“ (Il Post), „Ла stampa“ (La Stampa), „Република“ (Repubblica), италијанското издание на магазинот „Бајерд“ (Wired Italia), како и „Гардијан“ (Guardian).

Во 2011 година соработувавме со „Објавете што финансирате“ (Publish What You Fund) и „Институтот за меѓународен развој“ (Overseas Development Institute) на проект за мапирање на финансиската помош испратена во Уганда од 2003 до 2006 година. Тоа претставуваше новина затоа што, за прв пат, можевме да видиме како се движи помошта во споредба со националниот буџет – што ни овозможи да видиме во која мера се совпаѓаат приоритетите на донаторите и приоритетите на владите.

Дојдовме до неколку интересни заклучоци – на пример, и програмите за борба против ХИВ и програмите за планирање на семејството безмалку целосно се финансирани од надворешни донатори. „Гардијан“ известуваше за тој проект.

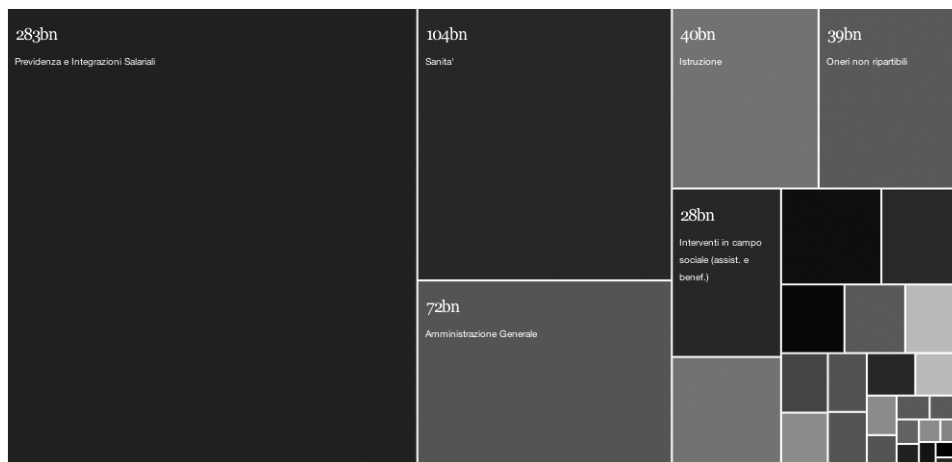
Исто така соработуваме со невладини организации и групи за застапување на вкрстена проверка на податоците за трошењето со други извори на информации. На пример, „Приваси интернешнел“ (Privacy International) ни пристапи со еден долг список на компании што произведуваат технологија за следење на комуникациите и надзор, како и список на агенции што присуствувале на еден познат меѓународен саем на технологија за следење и надзор, популарно познат како „бал на прислушувачите“. Преку систематско вкрстување на имињата

- "Dove vanno a finire i nostri soldi?" Iniziativa Open Data (dati trasparenti): clicca qui per saperne di più

LEGENDA:

Italian Regional Public Accounts = Conti pubblici delle regioni italiane;

Spending by Function = Spesa per funzioni; **Spending by Region** = Spesa per regioni; **unità di misura** billion (b) = miliardi



Слика 3-11. Италијанската верзија на проектот „Каде одајат моите пари“ („La stampa“)

на компаниите со податоците за трошењето, можевме да ги идентификуваме компаниите што добиле владини зделки – и потоа можевме за нив да поднесеме барања за пристап до информации. И за ова известуваше „Гардијан“.

Во моментот работиме на подобрување на фискалната писменост на новинарите и јавноста, како дел од еден проект наречен „Приказни за трошењето“ (Spending Stories), кој им овозможува на корисниците да ги поврзат податоците за трошењето со сториите за трошењето за да ги видат бројките зад вестите, како и вестите што произлегуваат од бројките.

Работата на тоа поле нè научи дека:

- Новинарите често не се навикнати да работат со сурови податоци, а многумина тоа не го сметаат за нужна основа на нивното известување. Извлекувањето приказни од суровите информации останува релативно нова идеја.
- Анализата и разбирањето на податоците бараат многу време, дури и кога ги поседувате потребните вештини. Нивното вклопување во циклусот на вести со краток рок на траење не е лесно, така што дата-новинарство најчесто се применува кај долгорочни истражувачки проекти.

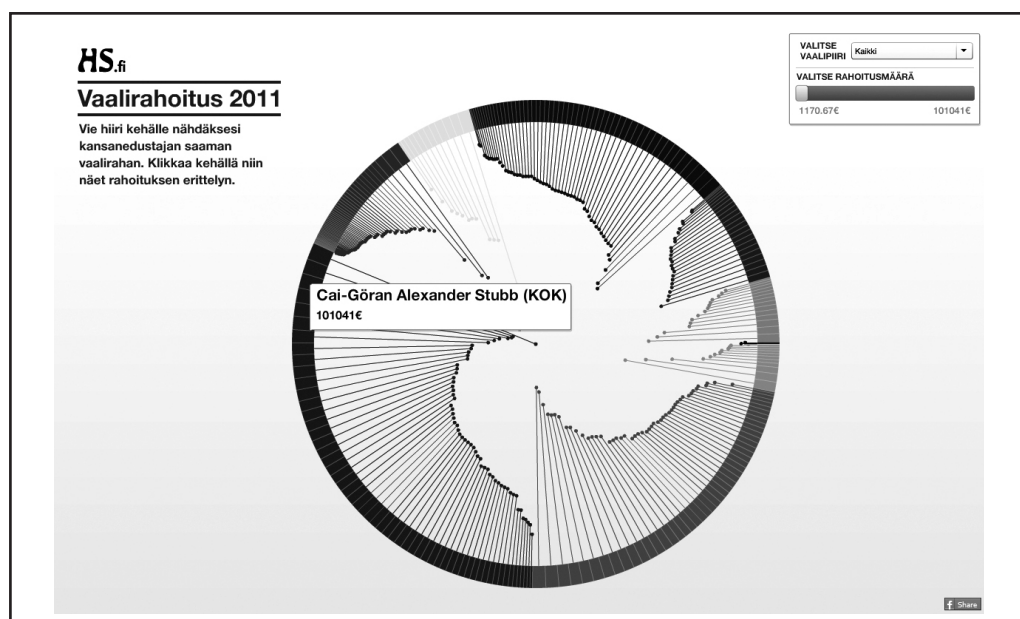
- Податоците објавени од властите често се некомплетни и застарени. Многу често се случува јавно достапните бази на податоци да не можат да се искористат за истражувачки цели без додавање поспецифични информации добиени преку барања за пристап до информации.
- Групите за застапување, научниците и истражувачите често имаат повеќе време од новинарите за спроведување пообемни истражувања засновани на податоци. Соработката со нив и тимската работа можат да се покажат многу корисни.

– Луси Чејмберс и Џонаџан Греј, Фондација „Отворено знаење“ (Lucy Chambers and Jonathan Gray, Open Knowledge Foundation)

Парламентарните избори и финансирањето на кампањите во Финска

Во последниве неколку месеци се одвиваат судски процеси поврзани со финансирањето на изборните кампањи за време на општите избори во Финска во 2007 година.

По изборите, медиумите открија дека законите за објавување на податоците за финансирањето на кампањите немале никаков ефект врз политичарите.



Слика 3-12. Финансирање на изборите („Хелсингин саномай“)

Во основа, фондовите за водење кампања биле користени за купување услуги од политичарите кои потоа не ги пријавиле фондовите со кои располагале, како што бараат финските закони.

Таквите случаи водеа до усвојување на построго законодавство. По општите избори од март 2011 година, весникот „Хелсингин саномат“ (Helsingin Sanomat) си постави задача внимателно да ги истражи сите достапни податоци за финансирањето на кампањите. Новото законодавство пропишува дека финансиските средства користени во кампања мора да се пријават, а само донаторите што дале помалку од 1.500 евра можат да останат анонимни.

1. Пронаоѓање податоци и девелопери

Од март 2011 наваму, „Хелсингин саномат“ организира таканаречени „ХС Опен“ (HS Open) хакатони. На нив ги каниме финските програмери, новинари и графички дизајнери да ни се приклучат на настанот што се одржува во подрумот на зградата во која е сместена редакцијата.

Учесниците се поделени на групи од по три члена и ги поттикнуваме да развиваат апликации и визуелизации. На секој од трите до сега организирани хакатони имавме по над 60 учесници. Решивме дека финансирањето на кампањите ќе биде во фокусот на „ХС Опен број 2“ (HS Open#2), во мај 2011 година.

Податоците и документите за финансирањето на кампањите ги чува Националното биро за ревизија на Финска. И тоа беше најлесниот дел од работата. Шефот на одделот за информатика на Бирото Јако Хамунен (Jaakko Hamunen) направи веб-страница што нуди пристап до нивната база на податоци за финансирањето на кампањите. Заводот за ревизија го спроведе сето тоа во само два месеци по поднесувањето на нашето барање.

Веб-страницата Vaalirahoitus.fi ќе им нуди на јавноста и на медиумите информации за финансирањето на кампањите на сите идни избори.

2. Разгледување идеи

Учесниците на „ХС Опен 2“ предложија 20 различни прототипови што да се прави со податоците. Сите прототипови можете да ги пронајдете на нашата интернет-страница (текстот е на фински јазик).

Јане Пелтола (Janne Peltola), биоинформатички истражувач, забележа дека податоците за финансирање на кампањите се слични на генетските податоци со кои се занимаваат во неговата област, во смисла дека содржат многу меѓузависни врски. Биоинформатичарите се користат со една алатка со отворен изворен код за мапирање на таквите меѓузависности. Ги процесиравме податоците со таа алатка, наречена „Цитоскејп“ (Cytoscape), и добивме еден многу интересен прототип.

3. Реализација на идејата на хартија и на интернет

Законот за финансирање на изборните кампањи вели дека избраните членови на парламентот мора да ги објават финансиските податоци најдоцна два месеци по изборите. Тоа практично значи дека реалните податоци ги добивме кон средината на јуни. На „ХС Опен“ ги имавме само податоците од оние пратеници што таа обврска ја исполниле пред законскиот рок.

Имавме проблем и со форматот на податоците. Заводот за ревизија податоците ги обезбеди во две „CSV“-датотеки. Едната ги содржеше вкупните буџети на кампањите, а другата имињата на сите донатори. Требаше да ги комбинираме двата документи за да создадеме трет со три колони: донатори, примачи и донираната сума. Во случаите кога политичарите трошеле сопствени пари, во нашиот формат тоа изгледаше вака: „Политичарот А донирал икс евра на Политичарот А“. Пристапот можеби не е најинтуитивен, но на „Cytoscape“ функционираше добро.

По прочистувањето и реформатирањето на податоците, само ги поминавме низ „Cytoscape“. Одделението за графичка обработка потоа од нив произведуваше графички прикази на цела печатена страница.

На крај подготвивме и една прекрасна визуелизација за нашата веб-страница. Не се работеше за графички приказ на мрежна анализа. Тоа што сакавме беше на луѓето да им овозможиме лесна проверка колку пари се трошат на изборните кампањи и кој ги дава тие пари. Првиот преглед ја прикажува распределбата на средства меѓу членовите на парламентот. Кога ќе кликнете на името на некој пратеник, добивате преглед на неговите или нејзините финансии.

Можете и да гласате дали определен донатор е добар или не. Визуелизацијата ја подготвија Јуха Рувинен (Juha Rouvinen) и Јука Кокко (Jukka Kokko) од рекламната агенција „Сатума“ (Satumaa).

Верзијата за интернет ги користи истите податоци како и мрежната анализа.

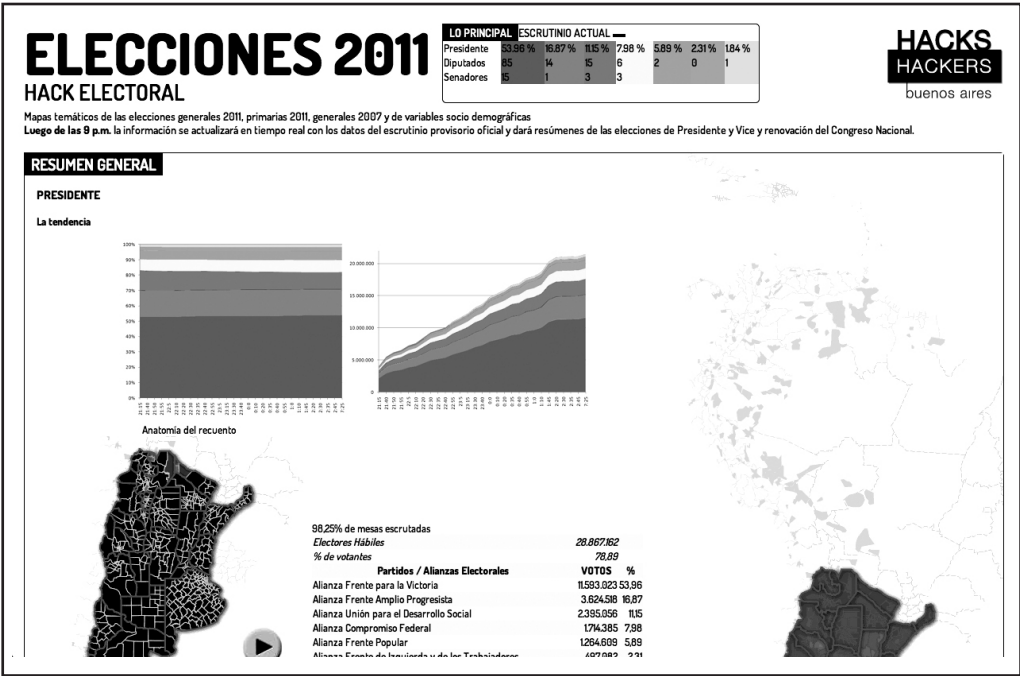
4. Објавување на податоците

Се разбира, Националниот завод за ревизија веќе ги имаше објавено податоците, така што немаше потреба и ние да го правиме тоа. Сепак решивме да ги објавиме затоа што нашите податоци беа прочистени и имаа подобра структура. Ние податоците ги објавуваме под лиценца „Криејтив комонс Наведи извор“ (Creative Commons Attribution). Неколку независни девелопери потоа ги искористија тие податоци за свои визуелизации, а неколку од нив објавивме и ние.

Во проектот користевме „Excel“ и „Google Refine“ за прочистување и анализа на податоците, „Cytoscape“ за мрежна анализа и „Illustrator“ и „Flash“ за визуелизација. Всушност, наместо „Flash“, сакавме да користиме „HTML5“, но немавме доволно време.

Што научивме? Најзначајната лекција за нас веројатно беше сознанието дека можеме да имаме проблем со начинот на структурирање на податоците. Ако изворните податоци не се ставени во соодветен формат, повторното пресметување и претворање може да одземе многу време.

Хакирање на избори во реално време (Хекс/хакерс Буенос Аирес)



Слика 3-13. Изборије во Арџентина во 2011 година (Хекс/хакерс Буенос Аирес)

„Електорал хек“ (Electoral Hack) е политичко-аналитички проект што ги визуелизира податоците од провизорните изборни резултати на изборите во Аргентина во октомври 2011 година. Системот нуди и информации од претходните избори и социодемографски статистики од целата земја. Проектот се ажурираше, во реално време, со информации од провизорните резултати од броењето на гласовите на изборите во 2011 година во Аргентина, за потоа да понуди сумирани извештаи за изборните резултати. Иницијатори на проектот беа групата „Хекс/хакерс Буенос Аирес“ и политичкиот аналитичар Енди Тоу (Andy Tow), а на него соработуваа новинари, девелопери, дизајнери, аналитичари, политиколози и други членови на локалниот огранок на „Хекс/хакерс“.

Какви податоци користевме?

Сите податоци потекнуваа од официјални извори: Државната изборна комисија обезбеди пристап до податоците од провизорното броење што го спроведе консултантската фирма „Индра“ (Indra); Министерството за внатрешни работи обезбеди информации за функциите за кои се бираа кандидати и за кандидатите на политичките партии; еден универзитетски проект обезбеди биографски податоци и информации за политичките платформи на сите претседателски кандидати; социодемографските податоци ги пронајдовме во Националниот попис на населението од 2001 година, Пописот од 2010 година, како и во податоците што ги добивме од Министерството за здравство.

Како ја развивме апликацијата?

Апликацијата беше развиена за време на Изборниот хакатон 2011 година (Election Hackathon 2011) на „Хекс/хакерс Буенос Аирес, еден ден пред изборите што се одржаа на 23 октомври 2011 година. На хакатонот учествуваа 30 волонтери што доаѓаа од различни области. „Хакирање на изборите“ беше развивано како отворена платформа што може понатаму да се развива и подобрува. Што се однесува до технолошките решенија, користевме „Гугл фјуџн тејблс“ (Google Fusion Tables), „Гугл мапс“ (Google Maps) и веќе постоечки библиотеки со векторски графики.

Работевме на изградбата на полигони за прикажување на географското мапирање и демографските податоци за изборите. Со комбинирање на полигоните изработени со софтверската апликација „ГИС“ (GIS) со геометарските решенија достапни во јавните табели обработени со „Google Fusion Tables“, подготвивме табели со клучеви што кореспондираат со базата на податоци за изборите на Министерството за внатрешни работи, „Индра“ и социодемографските податоци што ги добивме од „ИНДЕЦ“ (INDEC – Instituto Nacional de Estadística y Censos; Национален институт за статистика и пописи). Нив потоа ги искористивме за подготовка на визуелизациите во „Google Maps“.

Преку програмскиот интерфејс на „Google Maps“ објавивме неколку тематски мапи што ја прикажуваат просторната распределба на гласовите во различни нијанси на една боја, при што интензитетот на бојата го претставуваше процентот на гласови што индивидуалните претседателски кандидати ги добиле во различни административни окрузи и на различни гласачки места, со акцент на големите градски центри: Буенос Аирес, 24-те окрузи на територијата на Градот Буенос Аирес, градовите Кордоба и Росарио.

Истата техника ја користевме и за подготовка на тематски мапи од претходните избори, поточно примарните претседателски избори во 2011 година и изборите од 2007 година, како и распределбата на социодемографските податоци, како што се податоците за сиромаштијата, смртноста на децата и животните услови, за да овозможиме нивна натамошна анализа и споредба.

Проектот, исто така, ја прикажа просторната распределба на разликите во процентите од гласови што ги добија индивидуалните кандидати на општите избори во октомври 2011 година, споредено со примарните избори одржани во август истата година.

Подоцна, користејќи ги парцијалните податоци од провизорното броење на гласовите, изработивме анимирана мапа што ја прикажуваше анатомијата на броењето, со приказ на целиот тек на броењето од затворањето на гласачките места до наредното утро.

Позитивни страни

- Намерата ни беше да ги пронајдеме и претставиме податоците и успеавме во тоа. На располагање ја имавме базата на социодемографски податоци за децата на УНИЦЕФ, како и базата на податоци за кандидатите изработена од групата yoquierosaber.org од Универзитетот „Торкуато ди Тела“ (Universidad Torcuato Di Tella). На хакатонот собравме огромно количество дополнителни податоци кои, во крајна линија, не беа вклучени во проектот.
- Јасно е дека новинарската работа и програмирањето имаа огромна корист од вклученоста на научните работници. Проектот ќе беше невозможно да се спроведе без огромниот придонес што го дадоа Енди Тоу и Хиларио Морено Кампос (Hilario Moreno Campos).

Негативни страни

- Социодемографските податоци што ги имавме на располагање не беа ажурирани (најголемиот дел од нив беа собрани на Пописот од 2001 година) и не беа доволно детални. На пример, не вклучуваа детали за просечниот БДП на локално ниво, главните стопански активности, нивото на образование, бројот на училишта или лекари по глава на жител, како и многу други нешта што ќе беше супер да ги имаме.
- Изворно, системот беше замислен како алатка што може да се користи за комбинирање и прикажување на кои било податоци за да можат новинарите лесно да ги прикажат на интернет податоците што ги интересираат. Тоа, сепак, ќе мора да почека.
- Бидејќи проектот го изработија волонтери, во многу краток временски рок, не успеавме да направиме сè што сакавме. Сепак, постигнавме голем напредок во вистинската насока.
- Од истите причини, сиот труд на 30 луѓе на крај беше кондензиран во еден единствен програмер, во моментот кога почнаа да пристигаат податоците од владата, а ние се соочивме со проблеми во внесувањето на податоците во реално време. На програмерот му беа потребни неколку часа да ги реши проблемите.

Импликации

Платформата „Хакирање на изборите“ имаше големо влијание врз медиумите, врз известувањето на телевизија, на радио, во печатот и на интернет. Мапите изработени во рамките на проектот беа користени од неколку медиумски платформи за време на изборите и во деновите по нив. Со секој нов ден, со секое ново ажурирање на податоците, сообраќајот само растеше. Интернет-страницата, креирана токму на денот на изборите, имаше околу 20.000 уникатни посетители, а нејзините мапи беа репродуцирани на насловната страница на весникот „Пагина/12“ (Página/12) два дена по ред, а беа отпечатени и во статии објавени во „Ла насион“ (La Nación). Неколку мапи се појавија и во печатеното издание на весникот „Кларин“ (Clarín).

Тоа беше прв случај на користење интерактивни прикази на мапи создавани во реално време во историјата на аргентинското новинарство. Од носечките мапи јасно се гледаше надмоќната победа на Кристина Фернандез де Кирхнер (Cristina Fernandez de Kirchner) со освоени 54 проценти од гласовите, прикажана со интензивна боја. Тие, исто така, им послужија на корисниците да ги разберат специфичните случаи во кои локални кандидати остварија надмоќни победи во провинциите.

— Маријано Блехман, Маријана Беруецо, Серхио Сорин, Енги Тау и Марџин Сарсале од „Хекс/Хакерс Буенос Ајрес“ (Mariano Blejman, Mariana Berruezo, Sergio Sorín, Andy Tow, and Martín Sarsale, Hacks/Hackers Buenos Aires)

Податоците се вест: Викиликс

Сè почна кога еден член на истражувачкиот тим ме праша: „А тебе те бидува за работа со сметководствени табели, нели?“ А се работеше за „машала“ табела: 92.201 редови податоци од кои секој ред содржеше детален преглед на една воена акција во Авганистан. Тоа беа воените дневници на „Викиликс“ (WikiLeaks). То ест, тоа беше само првиот дел. Следеа уште две епизоди: Ирак, а потоа и дипломатските телеграми. Официјалното име за првите два дела беше „SIGACTS: the US military Significant Actions Database“ (База на податоци за значајните акции на воените сили на САД – СИГАКТС).

Авганистанските воени дневници беа пример за дата-новинарство на дело што го делиме со „Њујорк тајмс“ и „Дер шпигел“ (Der Spiegel). Ние сакавме да му овозможиме на нашиот тим од специјализирани новинари да ги пронајдат човечките приказни во информациите – и сакавме да ги анализираме информациите за да ја сфатиме пошироката слика, да покажеме како навистина се одвива војната.

Од клучно значење за тоа што сакавме да го постигнеме беше да не ја објавуваме целата база на податоци. „Викиликс“ веќе имаше намера да го стори

тоа, а ние сакавме да се обезбедиме дека нема да ги откриеме имињата на нашите информатори и да не ги доведеме во опасност единиците на НАТО. Во исто време, требаше да ги направиме податоците полесни за користење за нашиот тим од истражувачки новинари, предводен од Дејвид Ли (David Leigh) и Ник Дејвис (Nick Davies) [тие двајца преговараа со Џулијан Асанџ (Julian Assange) околу добивањето на податоците]. Исто така, сакавме да го поедноставиме пристапот до клучните информации, таму во реалниот свет, да биде што е можно појасен и поотворен.

Податоците ги добивме во форма на огромна „Ексел“-датотека; над 92.201 редови, од кои некои беа наполно празни или лошо форматирани. Документот во таа форма немаше да им биде од никаква полза на новинарите што бараат приказни во податоците, а беше преголем за да можеме на него да засноваме осмислени извештаи.

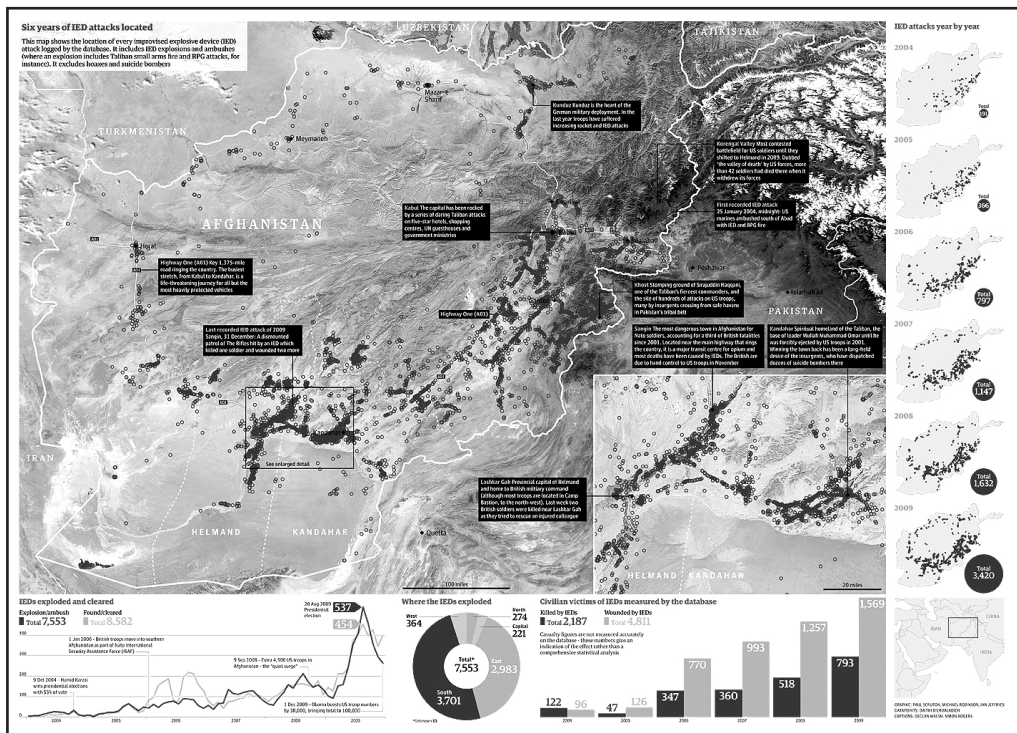
Нашиот тим изработи едноставна интерна база на податоци со користење „SQL“. Новинарите сега можеа да бараат приказни користејќи клучни зборови или настани. Наеднаш, целата група податоци стана достапна и тоа го олесни откривањето на приказните.

Податоците беа добро структурирани. Секој настан ги содржеше следните клучни податоци: точно време, датум, опис, број на ранети и загинати и, што е клучно, точни географски координати (географска должина и ширина на локацијата).

Ги филтриравме податоците за да ни помогнат да раскажеме една од клучните приказни од војната: порастот на бројот на напади со импровизирани експлозивни напади – ИЕН (improvised explosive device, IED), експлозивни напади склопени во нечија дневна соба што се оставаат покрај патиштата, непредвидливи и тешки за спречување. И понатаму се работеше за огромна група податоци, но ракувањето со неа беше многу полесно. Помеѓу 2004 и 2009 година, забележани се околу 7,500 експлозии на ИЕН или заседи (заседа е кога нападот е комбиниран со стрелба од огнено оружје или ракетни фрлачи). Уште околу 8.000 ИЕН бомби беа откриени на време и деактивирани. Сакавме да видиме како бројките се менуваат со текот на времето – а и да ги споредиме тие бројки. Податоците покажаа дека ситуацијата е најлоша на југот од земјата, каде што тогаш беа стационирани британските и канадските единици – потврда за она што нашите репортери што известуваа од војната веќе го знаеја.

Објавувањето на ирачките воени дневници во октомври 2010 година „истури“ уште 391.000 досиеја од војната во Ирак во јавната арена.

Споредено со протекувањето податоци од Авганистан, ова беше друга, повисока лига; може основано да се тврди дека тоа ја направи ирачката војна најдокументирана војна во историјата. Можевме да ги анализираме и расчлениме сите, дури и најситните детали. Еден фактор сепак отскокнува над другите: огромната бројка на загинати, огромното мнозинство од нив цивили.



Слика 3-14. Военише дневници на „Викиликс“ („Гардијан“)

Како и во случајот на Авганистан, „Гардијан“ реши да не ја преобјавува целата база на податоци, најмногу затоа што не можевме да бидеме сигурни дека во полето со сумирани податоци не се содржани и доверливи и тајни податоци за оние што ги доставиле информациите и слично.

Сепак, на нашите корисници им овозможивме да снимат една сметководствена табела со записи за сите инциденти во кои имало човечки жртви, вкупно околу 60.000 такви записи. Го отстранивме полето со сумирани податоци, така што останаа само чистите податоци: називот на единиците, бројот на загинали и географските координати.

Ги одвоивме сите инциденти во кои имало загинали и, со помош на „Google Fusion“, ги поставивме на една мапа. Мапата беше далеку од совршена, но претставуваше некаков почеток во обидите да ги мапираме обрасците на де-струкција што го разоруваа Ирак.

Дипломатските телеграми беа објавени во декември 2010 година. Тоа, повторно, беше сосема друга лига, огромна група податоци со официјални документи: 251.287 дописи од повеќе од 250 амбасади и конзулати на САД од

целиот свет. Тие даваат единствен приказ на дипломатскиот јазик што го користи САД – вклучувајќи и преку 50.000 документи од мандатот на сегашниот претседател Обама. Но, за какви податоци се работеше?

Телеграмите потекнуваат од огромната „Тајна мрежа за насочување на пораките со интернет-протокол“ (Secret Internet Protocol Router Network – SIPRNet). SIPRNet е глобалниот интернет-систем на воените сили на САД, одвоен од обичниот интернет наменет за цивилите и управуван од Министерството за одбрана во Вашингтон.

Од нападите во септември 2001 година наваму, во САД имаше серија обиди да се поврзат архивите каде што се чуваат владините информации, со надеж дека така ќе се избегне заробување клучни разузнавачки податоци во информациските складишта или „цевоводи“. Во изминатата деценија повеќе американски амбасади се поврзаа на мрежата SIPRNet, што овозможи размена на воените и дипломатските информации. До 2002 година, 124 амбасади беа врзани на мрежата: до 2005 година, бројката порасна на 180 амбасади, а денес на системот се поврзани огромно мнозинство од дипломатските мисии на САД во светот – и тоа е причината поради која мнозинството од телеграмите што ги имаме потекнуваат од 2008 и 2009 година. Како што пишува Дејвид Ли:

Телеграмата со ознака „SIPDIS“ автоматски се снима на заштитената интернет-страница со доверливи материјали на амбасадата од која потекнува. Од таму, до неа може да пристапат сите, не само вработените во Стејт департментот, туку и сите припадници на воените сили на САД што имаат овластување за пристап до доверливи информации со ниво „Доверливо“ („Secret“), лозинка и компјутер поврзан на мрежата SIPRNet.

... категорија која, верувале или не, опфаќа над три милиони лица. Овде постојат неколку нивоа на податоци, вклучително и нивото „SECRET NOFORN“, што означува информации што никогаш не можат да бидат покажани на лица што не се државјани на САД. Таквите информации се наменети за владините функционери во Вашингтон, сè до нивото на Државниот секретар на САД Хилари Клинтон (Hillary Clinton). Вообичаено, телеграмите ги подготвуваат амбасадорот или неговите подредени. Разузнавачките документите со ознака „Строго доверливо“ (Top Secret) или со повисока ознака за тајност не се достапни преку мрежата SIPRNet.

За разлика од претходните објави, овде се работеше претежно за текстуални документи, неквантификувани и без идентични типови податоци. Телеграмите ги содржат следните податоци:

Извор

Амбасадата или друго тело што ја испраќа телеграмата.

Список на ѝримачи

Вообичаено, телеграмите се испраќаат до неколку други амбасади и владини тела.

Поле за внесување на предметот на телеграмата

Кратко резиме на телеграмата.

Ознаки

Секоја телеграма е означена со неколку кратенки што служат како клучни зборови.

Текст на телеграмата

Овде е содржана целата порака што се испраќа, а ние одбравме да не ги објавуваме во целина од очигледни безбедносни причини. Еден интересен аспект на оваа приказна е начинот на кој телеграмите безмалку и самите предизвикуваа истекување информации „по барање“. Тие беа главна вест со недели по објавувањето; денес, секогаш кога ќе се појави приказна за некој корумпиран режим или некаков меѓународен скандал, пристапот до телеграмите ни дава и пристап до нови приказни.

Анализата на телеграмите е огромен потфат кој можеби никогаш нема да биде целосно завршен.

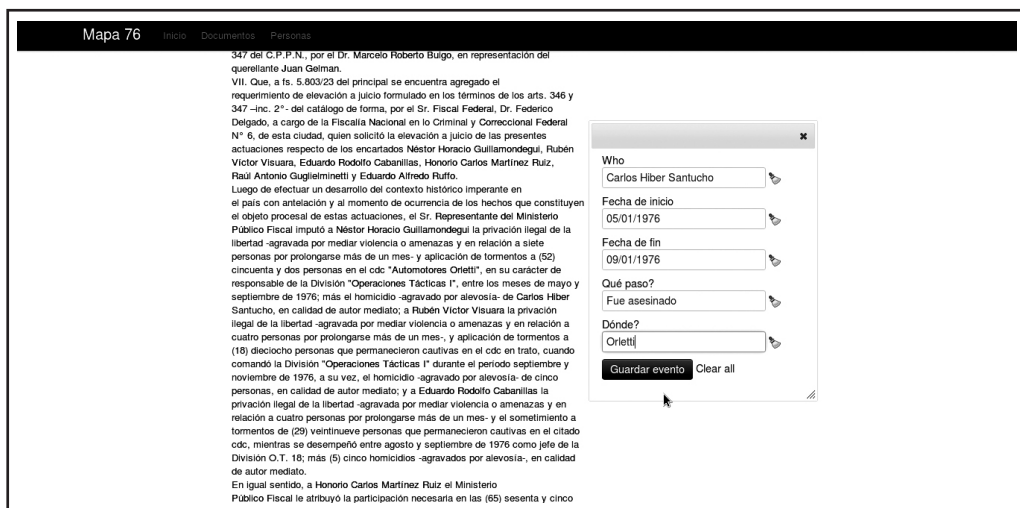
— Овој текст го приредивме од едно поглавје на книгата „Фактите се свети: Моќта на податоците“ (Facts are Sacred: The Power of Data) на Сајмон Роџерс, новинар на „Гардијан“ (објавена на „Kindle“)

Хакатонот „Мара76“

Огранокот на „Хекс/Хакерс“ во Буенос Аирес беше основан во април 2011 година. Организиравме две иницијални средби за да ѝ дадеме публицитет на идејата за поголема соработка меѓу новинарите и девелоперите на софтвер, и на секоја од нив учествуваа помеѓу 120 и 150 лица. Третата средба ја организиравме во форма на 30-часовен „хакатон“ со осум учесници, на конференцијата за дигитално новинарство во Розарио, град оддалечен 300 километри од Буенос Аирес.

Една тема, што се повторуваше на тие средби, беше желбата за „стругање“ големи количества податоци од интернет и нивна визуелна презентација. Проектот Мара76.info настана за да помогне во остварувањето на таа желба и им помага на корисниците во извлекувањето на податоците и нивното прикажување на мапи и хронолошки табели. Задачата воопшто не беше лесна.

Зошто ја нарековме „Мара76“? На 24 март 1976 година, во Аргентина беше изведен државен удар и тој режим остана на власт до 1983 година. Се проценува дека во тој период исчезнале 30.000 лица, илјадници се ликвидирани, а 500 деца родени во затвор биле присвоени од страна на воената диктатура. Триесет години подоцна, бројката на осудени за злосторства против човештвото извршени за време на диктатурата во Аргентина достигна 262 лица (заклучно со септември 2011 година).



Слика 3-15. Майа76 (Хекс/хакерс Буенос Аирес)

Во моментов се во тек 14 судски постапки, а во седум од нив е закажан почеток на главната расправа. Вкупно 802 лица се вклучени во различни отворени судски постапки. Процесуирањето на тие случаи генерира огромни количества податоци, а нивната обработка претставува голема тешкотија за истражувачите, новинарите, организациите за човекови права, судиите, обвинителите и другите заинтересирани лица.

Податоците се произведуваат на дистрибуиран начин, а истражувачите често не ги користат софтверските алатки што би им помогнале во нивното толкување. Во крајна линија, тоа значи дека често се случува некои факти да бидат занемарени, што води кон ограничени хипотези. „Мара76“ е истражувачка алатка што обезбедува отворен пристап до тие информации и нивно истражување за новинарски, правни, правосудни и историски цели.

Како подготовка за хакатонот, креиравме платформа што девелоперите и новинарите можеа да ја користат на настанот. Мартин Сарсале разви неколку основни алгоритми за извлекување структурирани податоци од едноставни текстуални документи. Користевме и неколку, но не многу, библиотеки од проектот DocumentCloud.org. Платформата требаше да врши автоматска анализа и извлекување имиња, датуми и локации од текстовите – и да им овозможи на корисниците прегледување на клучните факти во индивидуалните случаи (на пр., датумот на раѓање, каде е извршено апсењето, каде последен пат е видно исчезнатото лице итн.).

Целта ни беше да понудиме платформа за автоматско извлекување податоци за одлуките на воената диктатура во Аржентина. Баравме начин за авто-

матско (или барем полуавтоматско) прикажување на податоците поврзани со случаите од периодот 1976-1983 година, врз основа на пишани докази, тврдења и одлуки. Извлечените податоци (лични имиња, географски локации и датуми) се собираат, чуваат и можат да бидат анализирани и прочистени од страна на истражувачот што работи со нив, а можат да се прегледаат со помош на мапи, хронолошки табели и алатки за мрежна анализа.

Проектот ќе им овозможи на новинарите и истражувачите, обвинителите и сведоците да ја следат животната приказна на едно лице, вклучително и времето поминато во заточеништво и последователното исчезнување или пуштање на слобода. Таму каде што нема информации, корисниците можат да прочешлаат огромен број документи во потрага по информации што би можеле да бидат релевантни за случајот што го истражуваат.

Хакатонот го најавивме преку „Хекс/Хакерс Буенос Аирес“. Групата тогаш имаше околу 200 членови (во времето кога ги пишуваме овие редови, групата брои околу 540 членови). Контактиравме и голем број здруженија за човековите права. На средбата присуствуваа околу четириесет лица – новинари, организации што се занимаваат со застапување, девелопери и дизајнери.

За време на хакатонот ги идентификувавме задачите што различните учесници можеа да ги вршат самостојно и така да помогнат сè да помине мазно. На пример, од дизајнерите побаравме да работат на интерфејс што ќе комбинира мапи со хронолошки табели, од девелоперите побаравме да разгледаат можни начини за извлекување структурирани податоци и алгоритми за разликување на имињата, а од новинарите побаравме да проверат што се случило со определени лица, да ги споредат различните верзии на една иста приказна и да ги прочешлаат документите за да ги раскажат приказните за определени случаи.

Веројатно најголемиот проблем што се јави по завршувањето на хакатонот беше амбициозната природа на проектот, краткорочните цели беа сложени, а од друга страна, не е лесно да се координира лабаво поврзана мрежа на волонтери. Скоро сите луѓе вклучени во проектот имаа напорен ден, а многумина од нив учествуваа и на други настани и во други проекти. Групата „Хекс/Хакерс Буенос Аирес“ одржа девет средби во 2011 година.

Во моментот, активно се работи на развивањето на проектот. На таа задача работи основен тим од четири члена заедно со повеќе од дузина соработници. Проектот има и своја јавна мејлинг-листа и складиште за програмерски код, па сите што сакаат можат да се вклучат во проектот преку нив.

– Маријано Блехман, Хекс/Хакерс Буенос Аирес
(Mariano Blejman, Hacks/Hackers Buenos Aires)

Известувањето на „Дatabлогот на Гардијан“ за немирите во Велика Британија

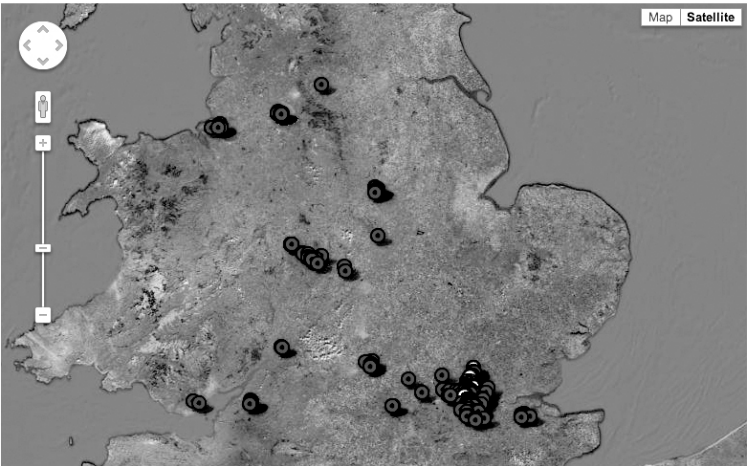
UK riots: every verified incident - interactive map

What has happened where as rioting spreads across England. Click on a dot to get details, or zoom to see more events

- Explore the data behind this map and discuss it
- Animation: see how the riots spread
- The riots mapped with poverty
- Download this map as a KML file
- Download it as a Google Fusion table

• NEW: list of riot-related cases before the courts

Simon Rogers, Ami Sedghi and Lisa Evans
guardian.co.uk, Thursday 11 August 2011 09:30 BST



Слика 3-16. Немирите во Велика Британија: сите потврдени инциденти („Гардијан“)

Летото 2011 година, Велика Британија беше погодена од бран улични немири. Политичарите тогаш категорично тврдеа дека таквите акции никако не се поврзани со сиромаштијата и дека оние што ги ограбуваат продавниците се најобични криминалци. Понатаму, премиерот и водечките конзервативни политичари ги обвинија социјалните медиуми за предизвикување и поттикнување на немирите, сугерирајќи и дека немирите биле организирани преку „Фејсбук“, „Твитер“ и „Блекбери месинџер“ [Blackberry Messenger (BBM)]. Се јавија и иницијативи за привремено затворање на социјалните медиуми. Бидејќи властите не спроведоа истрага за причините за немирите, „Гардијан“, во соработка со Лондонската школа за економија (London School of Economics) и за да одговори на тие прашања, го осмисли проетот „Читање на немирите“ („Reading the Riots“).

Весникот избилно се потпираше на дата-новинарството за да ѝ овозможи на јавноста подобро разбирање за тоа кој ограбуваше и зошто. Понатаму, заедно со една друга група претставници на академската заедница, предводена од професорот Роб Проктер (Rob Procter) од Универзитетот во Манчестер, беше разгледувана и улогата на социјалните медиуми што и „Гардијан“ во голема мера ги користеше во известувањето за време на немирите. Тимот на проектот „Читање на немирите“ беше предводен од Пол Луис (Paul Lewis), уредник на специјални проекти во „Гардијан“. За време на немирите, Пол известуваше од „првите борбени линии“ во градовите во цела Англија (особено преку „Твитер“, од профилот @paullewis). Вториот тим прегледа 2,6 милиони „твитови“ донирани од „Твитер“. Главната цел на работата со социјалните медиуми беше да се утврди начинот на кој озборувањата циркулираат на „Твитер“, улогата што различни корисници/актери ја имаат во пропагирањето и ширењето на тековите на информациите, да се види која платформа била користена за поттикнување и да се истражат другите форми на организирање.

Од гледна точка на користењето на дата-новинарството и визуелизацијата на податоците, сметам дека е полезно да се направи разлика помеѓу два клучни периоди: периодот во кој се одвиваа уличните немири и начините на кои податоците помагаа да се раскажат приказните додека траеја немирите; и вториот период на многу поинтензивно истражување на двата академски тима што соработуваа со „Гардијан“, задолжени за собирање на податоците, нивна анализа и пишување детални извештаи за наодите. Резултатите од првата фаза на проектот „Читање на немирите“ беа објавени во текот на една недела на интензивно известување, во првата половина на декември 2011 година. Овде ви нудиме неколку клучни примери за користење дата-новинарство во двата периоди.

Фаза 1: Немирите како што се одвиваа

Преку едноставни мапи, тимот за обработка на податоци на „Гардијан“ ги прикажа локациите на кои е потврдено избивање улични немири, а интегрираниот приказ на податоците за материјалната депривација и локациите на кои избиле немири почна да го подрива водечкиот политички наратив дека немирите не се поврзани со сиромаштијата. Во двата примери користевме алатки за мапирање кои веќе постоеја, додека во вториот пример имаме комбинација на податоците за географската локација со друга група податоци за да се извлечат други врски и линкови.

Во врска со користењето на социјалните медиуми за време на немирите (во нашиот пример тоа беше „Твитер“), весникот подготви визуелизација на „хаштаг“-ознаките што се користеа во тој период. Визуелизацијата покажува дека „Твитер“ главно е користен како одговор на немирите наместо за организација на ограбувањето. Притоа, хаштагот #riotcleanup на спонтаната кампања за чистење на улиците по завршувањето на немирите покажа најголема стапка на пораст на посетеноста и бројот на слобеници во периодот во кој траеја немирите.

Фаза два: Читање на немирите

Две визуелизации отскокнуваа и поттикнаа широка дебата по објавувањето на наодите од повеќемесечното интензивно истражување и од соработката со двата академски тима што ги спроведе весникот. Првата визуелизација, краток видеофилм, ги покажува резултатите од комбинирањето на познатите локации на кои избиле немири и адресите на лицата што учествувале во нив, како и таканаречената „патека за одење на немири“ (riot commute).

Во тој сегмент, весникот соработуваше со специјализираната компанија за мапирање транспортни прашања „ИТО Ворлд“ (ITO World) на моделирањето на најверојатните рути што учесниците во немирите ги користат за да стигнат до локациите каде што се ограбува, со акцент на различните обрасци во различните градови. Забележано е дека некои луѓе поминувале и поголеми растојанија за да стигнат на немирите.

Втората визуелизација се однесува на начините на ширење озборувања на „Твитер“. Во разговор со академскиот тим беше постигната согласност да се анализираат седум такви гласини. Академскиот тим потоа ги собра сите податоци поврзани со секоја од гласините и воспостави распоред за шифрирање на „Твитовите“ со користење четири основни шифри: Оние што само ја повторуваат гласината (шифра „тврдење“), што ја отфрлаат гласината (шифра „противставено тврдење“), што ја оспоруваат (шифра „прашање“) и оние што само коментираат (шифра „коментар“). Шифрарниците за сите твитови беа пополнети во три копии, а резултатите беа визуелизирани од Тимот за интерактивни проекти на „Гардијан“. Тимот на „Гардијан“, исто така, подготви статија за начинот на кој ја граделе визуелизацијата.

Посебно силен впечаток кај оваа визуелизација остава способноста да се прикаже, преку моќни слики, нешто што инаку тешко се опишува – виралната природа на гласините и озборувањата и начинот на кој тие се менуваат низ својот животен век. Мејнстрим-медиумите очигледно играат улога во некои од таквите гласини (на пример, преку нивно директно отфрлање или, од друга страна, потврдувајќи ги со тоа што ги претвораат во вести). Исто така, видлива е и корективната природа на „Твитер“ во смисла на справување со таквите озборувања. Визуелизацијата не само што помогна во раскажувањето на приказните, туку и обезбеди вистински увид во начинот на кој озборувањата се шират преку „Твитер“ и корисни информации што би можело да се направи во идни слични настани.

Од последниот пример станува јасно постоењето цврста синергија помеѓу весникот и академскиот тим способен за длабинска анализа на 2,6 милиони твитови што се однесуваат на немирите. Иако академскиот тим изгради комплет од претходно договорени алатки за потребите на нивната анализа, членовите на тимот сега работат на тоа што поскоро тие алатки да им станат достапни на сите на кои им се потребни, како работна површина за нивните

анализи. Во комбинација со упатствата за користење што ги дава тимот на „Гардијан“, тој комплет алатки ќе понуди корисна студија на случај за тоа како можат и другите да ја користат анализата на социјалните медиуми и визуелизацијата за да раскажат такви значајни приказни.

– *Farida Vis, Универзитетот на Лесџер (Farida Vis, University of Leicester)*

Свидетелства за успехот на училиштата во Илиноис

Државниот одбор за образование на Илиноис секоја година ги објавува „свидетелствата“ со демографски податоци и со податоци за успехот на сите јавни училишта во сојузната држава Илиноис. Се работи за огромна група податоци која оваа година содржеше табела со приближно 9.500 колони. Кога имаме толку многу податоци, проблемот е како да избереме што да претставиме. (Најтешката работа кај софтверските проекти не е како да се изгради софтверот, туку како да се обезбеди софтверот што го градиме да биде токму вистинскиот софтвер што ни треба.)

Во нашиот случај, изборот на тоа кои податоци се интересни беше задача на новинарите и на еден уредник од тимот што ја покрива областа на образованието. (Постојат многу податоци што на прв поглед можат да бидат интересни, но секој новинар веднаш ќе ви каже дека се работи за погрешни податоци или за податоци што водат во погрешна насока).

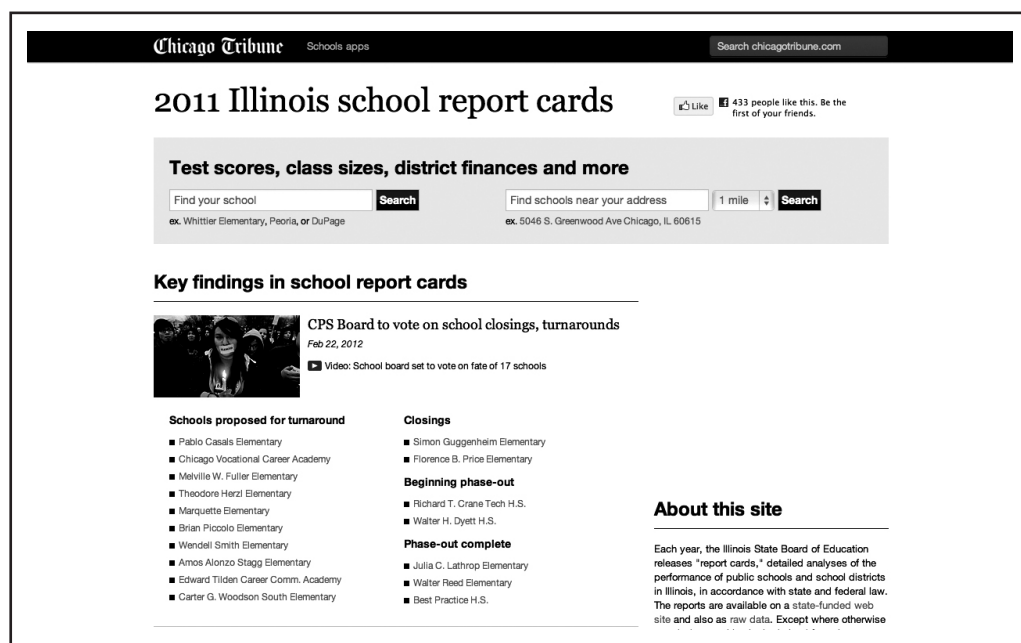
Исто така, редакцијата направи истражување и серија интервјуа со луѓе што имаат деца на училишна возраст. Интервјуата ги правевме во просторите на редакцијата, пред сè, со идеја да го пополниме недостигот на емпатија во редакцијата – никој од вработените во тимот за новински апликации нема деца на училишна возраст. Во текот на работата научивме многу за нашите корисници и за тоа колку лесна (или тешка!) за користење беше претходната верзија на нашата веб-страница посветена на училиштата.

Намерата ни беше дизајнот да го фокусираме на две посебни групи корисници:

- Родители кои имаат деца што одат на училиште и сакаат да знаат каков е квалитетот на нивното училиште.
- Родители во фаза на селидба што бараат каде да се преселат, знаејќи дека квалитетот на училиштата често има големо влијание врз изборот на новата адреса на живеење.

Во првиот случај, на проектот за изработка на веб-страницата посветена на училиштата работеа два девелопери во период од шест недели. На ажурирањето на веб-страницата во 2011 година работеа два девелопери на кои им

требаше четири недели да ја завршат работата. (Всушност, на вториот проект работеа три лица, но ниту едно од нив не беше вработено на полно работно време, така што збирно се работи за две работни места).



Chicago Tribune Schools apps Search [chicagotribune.com](#)

2011 Illinois school report cards

Like 433 people like this. Be the first of your friends.

Test scores, class sizes, district finances and more

Find your school Search Find schools near your address 1 mile Search

ex. Whittier Elementary, Peoria, or DuPage ex. 5046 S. Greenwood Ave Chicago, IL 60615

Key findings in school report cards

CPS Board to vote on school closings, turnarounds
Feb 22, 2012
Video: School board set to vote on fate of 17 schools

Schools proposed for turnaround

- Pablo Casals Elementary
- Chicago Vocational Career Academy
- Melville W. Fuller Elementary
- Theodore Herzl Elementary
- Marquette Elementary
- Brian Piccolo Elementary
- Wendell Smith Elementary
- Amos Alonzo Stagg Elementary
- Edward Tilden Career Comm. Academy
- Carter G. Woodson South Elementary

Closings

- Simon Guggenheim Elementary
- Florence B. Price Elementary

Beginning phase-out

- Richard T. Crane Tech H.S.
- Walter H. Dyett H.S.

Phase-out complete

- Julia C. Lathrop Elementary
- Walter Reed Elementary
- Best Practice H.S.

About this site

Each year, the Illinois State Board of Education releases "report cards," detailed analyses of the performance of public schools and school districts in Illinois, in accordance with state and federal law. The reports are available on a state-funded web site and also as raw data. Except where otherwise

Слика 3-17. Извештаи за успехот на училиштата во Илиноис во 2011 година („Чикаго трибјун“)

Дизајнот на информациите беше клучен сегмент на проектот. Иако пренесуваме многу мал дел од достапните податоци, тоа сепак се многу податоци и се соочивме со предизвикот како да ги направиме полесно сварливи. За среќа, можевме да позајиме дизајнер специјализиран за презентација на комплексни информации од нашиот графички оддел. Од него научивме многу за дизајнирањето графикони и, генерално, за патот до читлива презентација што сепак не ја потценува способноста или желбата на читателот да ги разбере бројките.

Сајтот го развиеме во апликациите „Питон“ (Python) и „Џанго“ (Django). Податоците ги чувавме во база на податоци од системот „МонгоДБ“ (MongoDB) – податоците за училиштата се хетерогени, со хиерархиска структура, што ги прави непогодни за складирање во релативна база на податоци. (Инаку најверојатно ќе користевме „PostgreSQL“.)

Во спроведувањето на проектот за прв пат експериментиравме со рамката за кориснички интерфејс „Бутстреп“ (Bootstrap) на „Твитер“ и мошне сме задоволни со резултатите. Графиконите ги исцртавме со користење на апликацијата „Флот“ (Flot).

Апликацијата, исто така, е дом за многуте приказни за успехот на училиштата што ги напишавме. Таа служи и како еден вид портал; кога имаме нова сторија за работата на училиштата, ја поставуваме најгоре на апликацијата, заедно со листа на училиштата што се релевантни за таа приказната. (По објавувањето нова сторија, читателите на www.chicagotribune.com се насочуваат кон апликацијата, а не кон конкретната сторија).

Првите информации што ги имаме велат дека на читателите им се допаѓа нашата апликација за училиштата. Повратните информации што ги добивме се главно позитивни (или барем конструктивни!), а посетите на страницата ги надминуваат сите очекувања. Како бонус, податоците ќе останат интересни цела година, па иако очекуваме посетите да се намалат како што сториите заминуваат од насловната страница, од претходните искуства знаеме дека читателите ја бараат апликацијата во текот на целата година.

Неколку клучни идеи што ги добивме од проектот се:

- Луѓето во одделот за графичка обработка ви се пријатели. Тие знаат како да ги направат комплексните информации полесно сварливи.
- Во небрано, побарајте помош од редакцијата. Ова е втор проект во кој спроведовме истражување и интервјуа на ниво на редакција, што се покажа како добар начин да се соберат ставовите на внимателни луѓе што, како и нашата читателска публика, имаат различно потекло и, општо земено, не се чувствуваат најслободно во работата со компјутери.
- Покажете го тоа што сте го сработиле! Голем дел од повратните информации што ги добивме всушност беа барања за пристап до податоците што беа користени во подготовката на апликацијата. Многу од податоците што ги имавме ги направивме достапни преку програмскиот интерфејс на апликацијата (API), а наскоро ќе го објавиме и делот што отпрво не мислевме да го објавуваме.

— Брајан Бојер, „Чикаго Трибјун“

Болнички сметки

Истражувачките новинари од „Калифорнија воч“ (CaliforniaWatch) дојдоа до информација дека еден голем синџир на болници во Калифорнија можеби се вклучени во систематска проневера на фондовите од федералната програма „Медикеар“ (Medicare) од која се покриваат трошоците за лечење на Американците постари од 65 години. Попрецизно, обвинувањата се однесуваа на измамата што се нарекува „upcoding“, односно пријавување дека примените пациенти се примени со покомплицирани здравствени проблеми од реалните – за што се добива поголем надоместок. Сепак, клучен извор на информации беше една синдикална организација што се бореше со управата на синџирот

За среќа, архивите на Секретаријатот за здравство на Калифорнија се јавни и нудат детални информации за секој поединечен пациент што бил лечен во болниците во државата. Вкупно 128 променливи вредности содржат до 25 дијагностички шифри наведени во прирачникот „Меѓународна статистичка класификација на болести и поврзани здравствени проблеми“ (познат и како МКБ-9) на Светската здравствена организација. Иако пациентите не се идентификувани по име, другите променливи податоци се однесуваат на возраста на пациентот, начинот на плаќање на трошоците и болницата во која бил лечен. Новинарите сфатија дека материјалите достапни од архивите ќе им помогнат да видат дали болниците во сопственост на синџирот пријавиле цени на лечењето на некои невообичаени болести и здравствени состојби што се повисоки од цените во другите болници.



Се работеше за огромен број податоци; безмалку 4 милиони досијеа годишно. Новинарите планираа да ги проверат документите за шест последователни години за да увидат на кој начин обрасците на однесување се менувале со текот на времето. Побараа податоци од државната агенција и податоците беа доставени на компјутерски компакт-дискови (CD-ROM) што лесно ги копиравме во компјутер. Новинарот што ги анализираше податоците го користеше системот „САС“ (софтвер за бизнис аналитика и разузнавање SAS). „САС“ е моќен систем (овозможува анализа на милиони документи) што го користат многу владини агенции, вклучувајќи го и Секретаријатот за здравство на Калифорнија, но е скап – таква анализа можеше да се направи и со користење на некоја од другите алатки за работа со бази на податоци, како што е „Microsoft Access“ или софтверот со отворен изворен код „MySQL“.

Откако ги обезбедивме податоците и ги напишавме програмите потребни за нивно проучување, пронаоѓањето сомнителни обрасци на однесување беше релативно едноставна работа. Едно обвинение тврдеше, на пример, дека синцирот на болници пријавува пациенти со различни нивоа на неухранетост во многу повеќе случаи од другите болници. Со САС, аналитичарот успеа да извади табели на зачестеност што ги покажуваа броевите на случаи на неухранетост пријавени секоја година во над 300 болници што нудат акутна нега во Калифорнија. Необработените табели на зачестеност ги внесовме во „Microsoft Excel“ за поблизу да ги испитае обрасците за секоја поединечна болница. Способноста на „Excel“ да подредува, филтрира и да врши пресметки на сурови бројки ја олесни идентификацијата на обрасците.

Особено впечатливи беа случаите на пациенти со „Квашиоркор“ (Kwashiorkor), синдром на протеинска дефициенција, која безмалку исклучиво се наоѓа кај изгладнети новоренчиња во земјите во развој погодени од глад. Сепак, синцирот на болници пријавуваше дека болниците во неговиот состав дијагностицираат многу висока стапка на стари Калифорнијци што имаат „Квашиоркор“, понекогаш и 70 пати повисока од просекот на сите болници во државата.

Други приказни беа произведени со анализа на пријавените случаи на септисемија, енцефалопатија, малигна хипертензија и нарушувања на автономниот нервен систем. Друга анализа се осврна на наводите дека синцирот испраќал на болничка нега, од единиците за итна помош, необично висок број пациенти што добиваат помош од програмата „Медикеар“, за кои постои многу поголема сигурност дека ќе ја покријат сметката за болничка нега од многу други пациенти примени од службите за итна помош.

Да сумираме, стории како горенаведените беа можни поради користењето на податоците за да се обезбедат докази и за независна проверка на наводите и обвинувањата што доаѓаат од извори што можеби имаат некаква своја скриена агенда. Овие приказни, исто така, претставуваат добар пример зошто се потребни строги закони за јавните архиви и документи; причината поради

која властите бараат од болниците да ги поднесуваат таквите податоци е токму поради потребата од ваков тип анализи, независно од тоа дали ќе ги врши владата, академската заедница, истражните органи или дури „граѓанските новинари“. Нивната тема е значајна затоа што истражува дали милиони долари од јавните фондови се трошат на правилен начин.

— *Стив Дојџ, Школа за новинарство „Волиџер Кронкајџ“, Државен универзитет на Аризона*

Криза во домовите за нега на стари лица

Истрагата на „Фајненшл тајмс“ за индустријата на приватни домови за нега на стари лица ги расветли начините на кои некои приватни инвеститори ја трансформираа негата на стари лица во машина за профит, со посебен осврт на цената во човечки животи што може да ја има бизнис-моделот што повеќе се грижи за поврат на инвестициите отколку за обезбедување добра нега.

Анализата не можеше да се погоди во подобро време затоа што се поклопи со откривањето на финансиските проблеми на „Садерн крос“ (Southern Cross), во тој момент најголемиот оператор на домови за нега на стари лица. Владата со децении ја поттикнуваше приватизацијата на секторот за нега на стари лица и лица со посебни потреби и продолжуваше да го фали приватниот сектор поради неговите разумни и разборити деловни практики.

Нашето истражување почна со анализа на податоците што ги добивме од регулаторното тело на Велика Британија овластено за инспекција на домовите за нега. Информациите беа јавни, но требаше упорност за да обезбедиме податоците да ги добиеме во форма што ќе овозможи нивно користење.

Податоците вклучуваа и ранг-листа (која денес не постои) за изведбата и резултатите на индивидуални домови за нега, како и преглед дали се во приватна сопственост, во сопственост на владата или непрофитни. Сè до јуни 2010 година, Комисијата за квалитет на негата (ККН) (Care Quality Commission (CQC)) го оценуваше квалитетот на услугите на домовите за нега (од 0 ѕвездички = лош квалитет, до 3 ѕвездички = одличен квалитет).

За нас, прв чекор беше детално прочистување на податоците, затоа што, на пример, податоците што ги добивме од Комисијата содржеа и невоедначени и неуниформирани категоризации. За таа задача примарно користевме „Excel“. Исто така откривме – преку деск и телефонско истражување – кои домови се во сопственост на приватни инвеститорски друштва. Пред финансиската криза, секторот на домови за нега на стари лица беше вистински магнет за приватните инвеститори и инвеститорите во недвижности. Со избивањето на кризата, неколку од нив – како што е „Садерн крос“ – се соочија со сериозни финансиски тешкотии. Ние сакавме да утврдиме какво е влијанието на управувањето од страна на приватни инвеститори врз квалитетот на негата, ако воопшто има влијание.

Еден релативно едноставен комплет пресметки во „Excel“ ни овозможи да утврдиме дека непрофитните и домовите со кои управува владата, во просек, обезбедуваат далеку подобар квалитет од оние управувани од приватниот сектор. Некои домови за неџа на стари лица во сопственост на приватни инвеститори работеа далеку над просекот, но други работеа многу под просекот.

Во комбинација со известување од лице место, студии на случаи на негрижа и длабински преглед на неуспешните регулаторни политики, како и други податоци за нивото на платите, стапките на поврат на инвестициите итн., нашата анализа успеа да ја наслика вистинската состојба на секторот за неџа на стари лица.

Неколку совети:

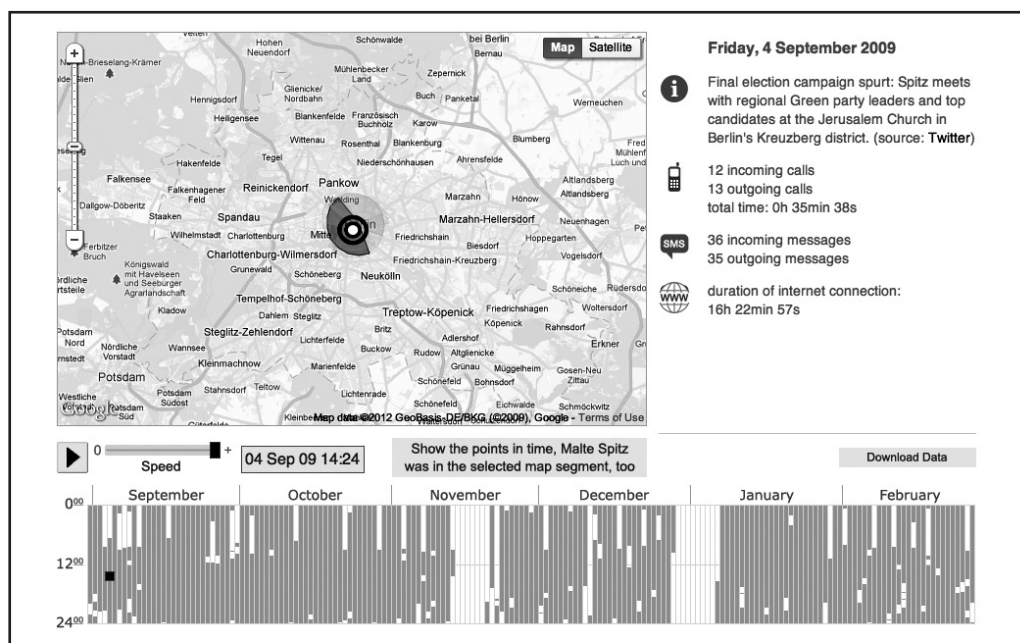
- Правете белешки за начинот на кој ракувате со оригиналните податоци.
- Чувајте копија од оригиналните податоци и никогаш не внесувајте промени во изворната копија.
- Проверете ги податоците и потоа проверете ги уште еднаш. Повторете ја анализата неколку пати (ако треба, почнете од почеток).
- Ако именуваат определени компании или индивидуални лица, обезбедете им можност да го искористат правото на одговор.

– Синџија О’Мурку, „Фајненшл тајмс“

Телефонот не чува тајни

Повеќето луѓе имаат само теоретско разбирање што може сè да се направи со податоците што наоколу ги испраќаат нашите мобилни телефони; до неодамна имаше мал број примери од реалниот свет. Токму поради тие причини, Малте Шпиц (Malte Spitz) од Зелената партија на Германија одлучи да ги објави своите податоци. За да добие пристап до информациите, тој мораше да поднесе тужба против телекомуникацискиот гигант „Дојче телеком“ (Deutsche Telekom). Податоците, собрани во еден огромен „Excel“ документ, послужија како основа на интерактивната мапа што ја подготви онлајн-изданието на „Цајт“ (Zeit Online). Секој од 35.831 редови со податоци во сметководствената табела претставува момент во кој мобилниот телефон на Шпиц пренесувал некакви информации во период од половина година.

Сами за себе, податоците се главно безопасни. Земени заедно, од друга страна, тие нудат нешто што истражителите го нарекуваат профил: јасна слика за навиките и вкусовите на едно лице, всушност, слика на неговиот живот. Профилот открива кога Шпиц прошетал по улица, кога се качил на воз, кога летал со авион. Покажува дека главно работи во Берлин, како и градовите што ги посетил. Покажува кога бил буден, а кога спие.



Слика 3-19. „Телефонот не чува тајни“ (Онлајн-издание на „Џајт“)

Податоците на „Дојче телеком“ веќе ја чуваа приватноста на еден дел од записите за активностите на Шпиц, имено, кого повикувал и кој го повикувал него. Таквите информации не само што би ја повредиле приватноста на многу луѓе од неговиот живот, тие исто така – дури и ако броевите се шифрирани – би откриле премногу информации за Шпиц (а владините агенти во реалниот свет би имале пристап до тие информации).

Ги замоливме Лоренц Мацат (Lorenz Matzat) и Михаел Крајл (Michael Kreil) од организацијата „ОпенДатаСити“ (Open Data City) да ги проверат податоците и да подготват решение за нивна визуелна презентација. „Прво користевме алатки какви што се „Excel“ и „Fusion Tables“ за и ние да ги сфатиме податоците. Потоа почнавме со развојот на мапа што ќе ѝ послужи на публиката да стапи во интеракција со податоците на еден нелинеарен начин“, ни кажа Мацат. За да илустрираме колку многу детали за нечиј живот можат да се ископаат од податоците што се чуваат, мапата ја проширивме со јавно достапни информации за неговата локација во даден момент (Твитер, блогови, информации за партиските активности наведени на јавно достапниот календар на неговата лична интернет-страница). Се работи за процес што секој добар истражител би го искористил за профилирање на следеното лице. Заедно со интерниот графички тим и тимот за истражување и развој (R&D) на онлајн-изданието на „Џајт“, тие финализираа еден лесен за снаоѓање интерфејс: со притисок на копчето тргнувате на пат низ животот на Малте Шпиц.

По успешната промоција на проектот во Германија, забележавме дека добиваме многу сообраќај од странство и решивме да направиме и верзија на апликацијата на англиски јазик. Откако ја доби германската интернет-награда „Гриме“ (Grimme Online Award), на проектот му беше доделена и наградата „ОНА“ (ONA Award) во септември 2011 година, како прв добитник на таа награда од Германија.

Сите користени податоци се достапни како сметководствена табела во „Google Docs“. Сторијата можете да ја прочитате на онлајн-изданието на „Цајт“ (Zeit Online).

– *Саша Фенор, „Цајт Онлајн“ (Sascha Venohr, Zeit Online)*

Кој модел на автомобил?

Стапката на паѓање на технички преглед

Во јануари 2010 година, Би-Би-Си ги доби податоците за стапката со која различни модели на автомобили од различни производители поминале или паднале на годишниот технички преглед (во оригинал „Ministry of Transport - MOT test“). На техничкиот преглед се проверува дали автомобилот ги задоволува безбедносните стандарди и дали може да учествува во сообраќајот; возилата постари од три години мора да поминат на технички преглед еднаш годишно.

Податоците ги добивме по барање за слободен пристап до информациите, но дури по долга битка со „ВОСА“ (VOSA), агенција во склоп на Министерство-то за транспорт што го надгледува системот за вршење технички прегледи. „ВОСА“ го отфрли нашето барање за пристап до информации со образложение дека тоа би значело повреда на трговската тајна. Тие тврдеа дека производителите чии возила имаат висока стапка на паѓање на технички преглед би претрпеле комерцијална штета. Поднесовме жалба до Комесарот за информации и тој реши дека откривањето на тие информации е прашање од јавен интерес. „ВОСА“ најпосле ни ги даде податоците, 18 месеци откако ги побаравме.

Во анализата на бројките се фокусиравме на најпопуларните модели и на споредба на автомобили со иста година на производство. Резултатите укажуваа на постоење големи разлики. На пример, кај возилата стари три години, 28 отсто од автомобилите од моделот „Меган“ на производителот „Рено“ (Renault Megane) паѓаа на техничкиот преглед, споредно со само 11 отсто кај „Тојота Корола“ (Toyota Corolla). Резултатите беа објавени на телевизија, на радио и на интернет.

Податоците ги добивме во форма на „PDF“=датотека од 1.200 страници, што мораше да ја конвертираме во сметководствена табела за да можеме да ја спроведеме анализата. Освен заклучоците, ја објавивме и сметководствената табела што ја подготвивме во „Excel“ (повеќе од 14.000 редови со податоци) на сајтот на „Би-Би-Си Њуз“ (BBC News) заедно со нашата сторија. Тоа на сите заинтересирани им овозможи пристап до податоците во форма во која може лесно да се искористат.

[Sign in](#)

[News](#)
[Sport](#)
[Weather](#)
[Travel](#)
[TV](#)
[Radio](#)
[More](#)

NEWS

Open Secrets

[« Previous](#)
[Main](#)
[Next »](#)

MOT failure rates released

Post categories: MOTs, transport
 Martin Rosenbaum | 06:00 UK time, Wednesday, 13 January 2010

The government agency which oversees the MOT system has backed down after 18 months and released data which shows how often different makes and models of cars and small vans fail MOTs.

This means that car and van buyers will now have access to the detailed MOT records of individual models, including reasons for failures. The figures show wide variation between different models, even when comparing vehicles of the same age.

The Vehicle and Operator Services Agency (VOSA), an arm of the Department for Transport, yesterday revealed 1,200 pages of detailed statistics on MOT failures following a freedom of information request made by the BBC in July 2008.

VOSA initially declined to supply the material, but last month the information commissioner ruled that disclosure is in the public interest and overturned VOSA's refusal.

About this blog

A blog about freedom of information, written by the BBC's Martin Rosenbaum.

For the latest updates across BBC blogs, visit the Blogs homepage.

Elsewhere at the BBC

You can read some of the stories the BBC has found using freedom of information here

Слика 3-20. Објавени сџајкиџе на џаѓање на џодиџниџи џехниџи џреѓлед (Би-Би-Си)

Како резултат, други луѓе и организации ги изработија податоците за сопствени анализи што ние немавме време да ги направиме затоа што брзавме да ја објавиме приказната (а во некои случаи се работеше за анализи што ги надминуваа нашите тогашни техниџки можности). Тие анализи вклучуваа истражување на стапките на паѓање на техниџки преглед за автомобили од други години на производство, споредба на резултатите за производителите во целина, а не само за индивидуални модели и креирање бази на податоџи со можност за пребарување на резултатите на индивидуални модели на автомобили. Во нашата онлајн-сторија додадовме линкови до таквите сајтови, така што нашите читатели можеа да имаат корист од нивниот труд.

Тука наведовме некои од предностите што со себе ги носи објавувањето на суровите податоџи врз кои се заснова некоја приказна. Може да има исклучоџи од тоа правило (на пример, ако планирате податоџите подоцна да ги искористите за дополнителни приказни, па поради тоа сакате да ги зачувате за себе), но, во целина, објавувањето на податоџите е корисно на повеќе начини:

- Вашата работа се состои од откривање на нештата и известување на луѓето за нив. Ако веќе сте се помачиле да ги добиете сите податоџи, дел од вашата работа е да ги проследите понатаму.

- Други луѓе можат да забележат некои значајни и интересни детали што вие сте ги превиделе, или напосто да ги забележат оние детали што ним им се важни, дури и ако вие не сте сметале дека се доволно значајни за да ги вклучите во вашата сторија.
- Други можат да го надградат вашиот труд со натамошна подетална анализа на податоците, со користење други техники за презентација и визуелизација на бројките, со користење на нивните сопствени идеи или технички вештини што би можеле да продрат во податоците на попродуктивен начин.
- Тоа е дел од процесот на вградување отчетност и транспареност во новинарскиот процес. На тој начин и други луѓе ќе можат да ја разберат вашата методологија или, ако сакаат, да ги проверат вашите резултати и наоди.

– *Марџин Розенбаум, Бу-Бу-Си (Martin Rosenbaum, BBC)*

Субвенции за јавниот автобуски превоз во Аргентина

Почнувајќи од 2002 година, субвенциите за јавниот автобуски транспортен систем во Аргентина растеа експоненцијално и секоја година достигнуваа ново рекордно ниво. Но, во 2011 година, по победата на изборите, новата влада на Аргентина најави намалување на субвенциите за јавните услуги, почнувајќи од декември таа година. Во исто време, владата одлучи да го пренесе управувањето со локалните автобуски и метро-линии на градските власти на Буенос Аирес. Градската управа на Буенос Аирес ја отфрли таа одлука поради некои нејаснотии поврзани со преносот на субвенциите на локалните власти и поради недостигот на доволно средства во локалниот буџет што би ја гарантирале безбедноста на транспортниот систем.

Додека траеше сето тоа, се состанав, за прв пат, со неколку колеги од „Ла насион“ (La Nacion) за да разговараме со што би можеле да ја почнеме нашата работа на полето на дата-новинарството. Уредникот на финансиската рубрика посочи дека податоците за субвенциите објавени од Секретаријатот за транспорт (<http://www.transporte.gov.ar/>) би можеле да бидат основа за првиот проект, а истовремено добар предизвик затоа што беа тешко разбирливи поради форматот и терминологијата со која се служеа.

Лошата состојба во која се наоѓа јавниот транспортен систем влијае на животите на повеќе од 5,8 милиони патници дневно. Доцнењата, штрајковите, дефектите, па дури и сообраќајните несреќи се честа појава. Поради тоа решивме да видиме каде завршуваат субвенциите наменети за јавниот транспортен систем и тие податоци да им ги направиме достапни на сите граѓани на Аргентина, преку нешто што го нарековме „Преглед на субвенци-

ите за јавниот транспорт“ (Transport Subsidies Explorer), што во моментот е во процес на изработка.

Почнавме со пресметки на тоа колкави суми добиваат автобуските компании од владата месечно. Со таа цел пред нас, ги прегледавме податоците објавени на интернет-страницата на Секретаријатот за транспорт – повеќе од 400 „PDF“-документи со месечните готовински исплати за повеќе од 1.300 компании во периодот по 2006 година.

Во тимот влезе еден виш програмер кој имаше задача да развие програма за „стругање податоци“, т.н. „скрејпер“ (scraper), за автоматизирање на редовното снимање и конверзија на тие PDF-документи во „Excel“ и „Database“ документи. Новата група податоци што содржи повеќе од 285.000 записи ја користиме за нашите истражувања и визуелизација, и во печатената и во онлајн-верзијата на весникот. Дополнително, податоците ги правиме достапни во машински читлив формат, така што сите Аргентинци ќе можат да ги користат и разменуваат.

Следниот чекор беше да идентификуваме колку, во просек, владата троши на месечното одржување едно возило на јавниот транспортен систем. За тоа да го сториме ја разгледавме интернет-страницата на една друга владина институција – Националната транспортна регулаторна комисија (ЦНРТ) (Comisión Nacional de Regulación del Transporte – CNRT; <http://www.cnrt.gov.ar/index2.htm>), тело одговорно за регулирање на транспортот во Аргентина. На интернет-страницата на комисијата пронајдовме список на автобуски компании со вкупно 9.000 возила. Развивме еден нормализатор за да можеме да ги разликуваме имињата на автобуските компании и да правиме вкрстена споредба на двете групи податоци.

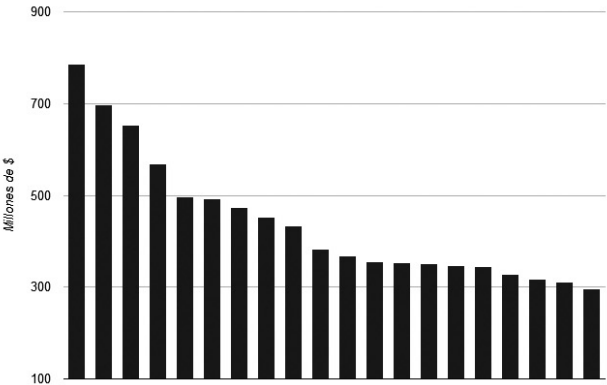
За да можеме да продолжиме понатаму, ни требаа регистарските броеви на сите возила. На интернет-страницата на „ЦНРТ“ пронајдовме списоци на возилата поделени по автобуска линија за секоја компанија посебно, заедно со регистарските таблички. Регистарските броеви за возилата во Аргентина се состојат од букви и бројки што ја опишуваат возраста на возилото. Мојот автомобил, на пример, има регистарски број „IDF234“, каде што „I“ го означува периодот март-април 2011 година. Со расчленување на регистарските таблички на автобусите на сите компании на листата, сакавме да ја пресметаме просечната возраст на возилата по компанија, за да покажеме колку пари одат во секоја од компаниите и да ги споредиме тие износи врз основа на просечната возраст на нивните возила.

Среде тој процес, содржината на „PDF“-документите објавени од владата мистериозно се промени, иако URL-адресите и имињата на документите останаа исти. Во некои PDF-документи недостигаа вертикалните „вкупни износи“, така што беше невозможно вкрстување на вкупните износи за целиот период од 2002 до 2011 година што беше предмет на нашето истражување.



Ranking de empresas subsidiadas

Cuáles son las empresas de colectivos que recibieron mayor cantidad de subsidios desde 2006.



Herramientas

Incrustar gráfico

- Descargar datos en .csv
- Descargar datos en .arff

Otros análisis

- Progresión de subsidios por mes
- Ranking de jurisdicciones subsidiadas
- Antigüedad de la flota
- Conversor de monedas

Datos

Ranking	Empresa	Total (en Millones de Pesos)
1	TRANSPORTE AUTOMOTOR PLAZA S.A.C.I.	785.45
2	EMPRESA LINEA 216 S.A.T.	698.29
3	CONSULTORES ASOCIADOS ECOTRANS S.A.	653.93
4	MICRO OMNIBUS NORTE S.A	568.38
5	GENERAL TOMAS GUIDO S.A.C.I.F.	497.87
6	MICRO OMNIBUS QUILMES S.A.C.I. Y F.	492.97
7	LA NUEVA METROPOL S.A.	474.58
8	SAN VICENTE S.A de TRANSPORTES	452.35
9	D.O.T.A. S.A. DE TRANSPORTE AUTOMOTOR	432.6
10	TRANSPORTE LARRAZABAL C.I.S.A.	383.28
11	AZUL S.A. DE TRANSPORTE AUTOMOTOR	367.07
12	LA CABAÑA S.A.	354.86
13	TRANSPORTES LA PERLITA S.A.	353.87
14	LA PRIMERA DE GRAND BOURG S.A.T.C.I.	350.62
15	TRANSPORTE IDEAL SAN JUSTO S.A.	346.04
16	EXPRESO GENERAL SARMIENTO S.A.	345.45
17	TRANSPORTES AUTOMOTORES RIACHUELO S.A.	327.18
18	EMPRESA DE TRANSP. TTE. GRAL. ROCA S.A	316.5
19	TRANSPORTE AUTOMOTORES LA PLATA S.A.	310.23
20	ALMAFUERTE EMP. DE TRANSP.S.A.C.I.E.I.	296.97

Anterior Siguiente

Слика 3-22. Рангирање на транспортните компании што добиле субвенции („Ла национ“)

Тој случај го претставивме на еден хакатон што „Хекс/Хакерс“ го организираа во Бостон, а таму девелоперот Мет Пери (Matt Perry) великодушно се понуди и изработи апликација што ја нарекуваме „PDF Spy“ („PDF-шпион“). Апликацијата ја доби наградата во категоријата „Најинтригантен проект“ на настанот. „PDF Spy“ се насочува кон некоја интернет-страница со многу „PDF“-документи и проверува дали содржината на тие документи е менувана. „Никогаш повеќе не дозволувајте да ве залаже фразата „транспарентност на владата“, пишува Мет Пери.

Кој го спроведуваше проектот?

На истражувањето што траеше 13 месеци работеше тим што се состоеше од седум новинари, програмери и еден дизајнер на интерактивни апликации.

За спроведување на проектот беа потребни:

- Новинари запознаени со начинот на функционирање на субвенциите за јавниот транспортент систем и постоечките ризици; познавање на пазарот на автобуски превоз.
- Програмер што поседува вештини на полето на „стругање“ на интернет, анализа и нормализација, како и извлекување податоци од „PDF“-документи и нивно внесување во сметководствени табели во „Excel“.
- Статистичар што ќе ги анализира податоците и ќе ги направи потребните пресметки.
- Дизајнер што ќе ги подготви интерактивните визуелизации на податоците.

Кои алатки ги користевме?

Апликацијата ја програмиравме во „VBasic“, а за „Истражувачот на субвенциите“ користевме „Excel Macros“, „Tableau Public“, платформата „Junar Open Data Platform“, како и „Ruby on Rails“, интерфејсот за програмирање апликации (API) на „Google charts“, и „MySQL“.

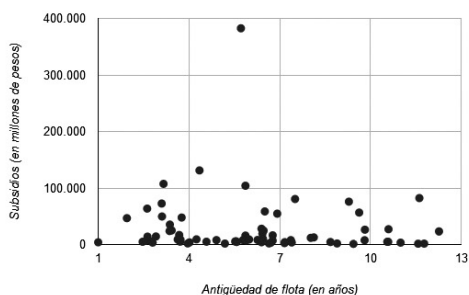
Проектот постигна огромен успех и имаше големо влијание. Апликацијата имаше илјадници посети, а истражувањето беше објавено на насловната страница на печатеното издание на „Ла насион“.

Успехот на тој прв проект на полето на дата-новинарството ни помогна внатре во редакцијата да поведеме иницијатива за основање единица за податоци што ќе го покрива истражувачкото новинарство и ќе дава услуги за јавноста. Како резултат ја добивме Data.lanacion.com.ar, платформа на која ги објавуваме податоците за различни теми од интерес за јавноста во формат што може да се чита машински.

— Анхелика Пералта Рамос, „Ла насион“ (Арџентина), (Angelica Peralta Ramos, La Nacion, Argentina)

Antigüedad de la flota

Cómo es la relación entre la antigüedad de la flota de cada empresa y los subsidios recibidos.



Herramientas

Incrustar gráfico

- Descargar datos en .csv
- Descargar datos en .arff

Otros análisis

- Progresión de subsidios por mes
- Ranking de empresas subsidiadas
- Ranking de jurisdicciones subsidiadas
- Conversor de monedas

Datos

Empresa	Antigüedad (en años)	Subsidios (en Millones de Pesos)
TRANSPORTE AUTOMOTOR PLAZA S.A.C.I.	5.72	382515.22
D.O.T.A. S.A. DE TRANSPORTE AUTOMOTOR	4.35	131078.01
TRANSPORTE LARRAZABAL C.I.S.A.	3.16	107317.63
CONSULTORES ASOCIADOS ECOTRANS S.A.	5.87	103975.56
GENERAL TOMAS GUIDO S.A.C.I.F.	11.62	82149.03
TRANSPORTES AUTOMOTORES RIACHUELO S.A.	7.51	80486.07
TRANSPORTE IDEAL SAN JUSTO S.A.	9.29	75783.63
EMPRESA DE TRANSP. TTE. GRAL. ROCA S.A.	3.1	72478.62
TRANSPORTES LOPE DE VEGA S.A.C.I.	2.62	63541.85
LOS CONSTITUYENTES S.A.T.	6.51	58496.65
EL NUEVO HALCON S.A.	9.63	56496.65
EXPRESO GENERAL SARMIENTO S.A.	6.92	54580.65
EMPRESA LINEA 216 S.A.T.	3.11	49578.33
EL PUENTE S.A.T.	3.76	47742.71
AZUL S.A. DE TRANSPORTE AUTOMOTOR	1.95	46617.68
TRANSPORTES RIO GRANDE S.A.C.I.F.	3.37	35292.58
NUDO S.A.	6.4	27932.39
EXPRESO QUILMES S.A.	10.6	27077.26
EXPRESO SAN ISIDRO S.A.T.C.I.F.I.	9.82	26240.67
EMPRESA TANDILENSE S.A.C.I.F.I.Y DE S.	3.43	24967.16

Anterior **Siguiente**

Слика 3-23. Сјоредба на просечнаи̑а возраси̑ на вознио̑и̑ њарк и износи̑е на средсџваи̑а добиени од владаи̑а („Ла насион“)

Граѓански дата-новинари

Големите редакции не се единствените што можат да работат на стории придвижувани од податоци. Вештините што ги користат дата-новинарите исто така може да им помогнат на граѓанските новинари да пристапат до податоци за нивните локални средини и да ги претворат тие податоци во приказни.

Тоа беше и основната мотивација на бразилскиот проект за граѓански медиуми „Пријателите на Жануарија“ (Friends of Januaria), кој доби грант од „Ражинг воисис“ (Rising Voices), огранокот на „Глобал воисис онлајн“ (Global Voices Online) за соработка со други организации, како и дополнителна поддршка од организацијата „Артикл 19“ (Article 19). Во септември и октомври 2011 година, група млади жители на еден помал град на северот на сојузната држава Минас Жераис, еден од најсиромашните региони во Бразил, поминаа низ обука за користење на основните новинарски техники и техниките на следење на извршувањето на буџетите. Научија и како се поднесуваат барања за слободен пристап до информации и како да пристапат до јавно достапните информации што се чуваат во официјални бази на податоци на интернет.



Слика 3-24. Проектиот за граѓански медиуми „Пријателите на Жануарија“ на граѓаните им ги пренесува клучните вештини и знаења што им се потребни за да станат граѓански новинари

Жануарија, град со околу 65.000 жители, е нашироко познат по катастрофалната политичка ситуација на локално ниво. Во три последователни четиригодишни мандати, на чело на градот се сменија дури седум различни градоначалници. Скоро сите беа отстранети од функцијата поради неправилности во работењето на јавната администрација, вклучувајќи и обвиненија за корупција.

Помалите градови како Жануарија често не добиваат никакво внимание на медиумите во Бразил, кои се фокусирани на поголемите градови и административните центри на сојузните држави. Од друга страна, жителите на помалите градови се можни сојузници во следењето на работата на јавната администрација, затоа што најдобро ги познаваат дневните предизвици со кои се соочуваат нивните локални заедници. Со интернет како сојузник, локалното население сега може да добие многу подобар пристап до информациите за локалните буџети и други податоци.

По учеството во 12 работилници, неколку од новите граѓански новинари од Жануарија почнаа да демонстрираат како таквиот концепт за пристап до јавно достапни податоци може да се користи и во помалите градови. На пример, Сораја Аморим (Soraia Amorim), 22-годишна граѓанска новинарка, напиша статија за бројот на доктори на платниот список на градот, според податоците на сојузната влада. Таа откри дека официјалната бројка не соодветствува со реалната ситуација во градот. Во пишувањето на статијата, Сораја го користеше пристапот до податоците достапни на интернет на страницата на „Единствениот здравствен систем – ЕЗС“ (Sistema Único de Saúde, оригиналната кратенка е „SUS“), сојузна програма за обезбедување бесплатна здравствена заштита за населението на Бразил. Според податоците достапни од ЕЗС, во Жануарија требаше да работат 71 доктори и специјалисти во разни области на медицината.

Бројот на доктори наведен во податоците на ЕЗС се разликуваше од она што Сораја го знаеше за бројот на доктори во регионот: нејзините сограѓани постојано се жалеа поради недостигот на доктори, а некои пациенти мораа да патуваат во соседните градови за да бидат прегледани. Таа потоа интервјуираше жена што имала сообраќајна незгода возејќи моторцикл и не можела да добие медицинска помош во болницата во Жануарија затоа што немало ниту еден доктор на располагање. Сораја зборуваеше и со градскиот секретар за здравство, кој призна дека во градот работат помалку доктори од бројката објавена од ЕЗС.

Почетните наоди отвораат многу прашања за причините за таквите разлики помеѓу официјалните информации објавени на интернет и реалната состојба во градот. Едно од прашањата е дали податоците на сојузно ниво се погрешни, што би значело дека во Бразил постои сериозен проблем со недостигот на информации од областа на здравството. Можно е и локалните власти на Жануарија да доставуваат погрешни информации до ЕЗС. И едната и другата можност би требало да водат до детална истрага за да се добие дефинитивен одговор на тоа прашање. Приказната на Сораја сепак е значајна алка во тој синцир затоа што става акцент на недоследноста на податоците, а може, исто така, и да охрабри и други луѓе посериозно да се зафатат со тоа прашање.

„Јас живеам на село и едвај завршив средно училиште“, вели Сораја. „Кога некој ќе ме прашаше што сакам да правам во животот, секогаш велев дека сакам да бидам новинарка. Знаејќи од каде доаѓам, сепак мислев дека тоа е скоро невозможно“. По учеството во обуката организирана од „Пријателите на Жануарија“, Сораја верува дека пристапот до податоците е значајна алатка за промена на состојбите во нејзиниот град. „Сметам дека можам да помогнам да постигнеме позитивни промени во мојот град, во мојата земја, во светот“, додава таа.

Друг граѓански новинар што учествуваше во проектот е 20-годишниот Алисон Монтиеритон (Alysson Montiériton). И тој користеше податоци во подготовката на една статија. Токму на првиот час одржан во рамките на проектот, додека граѓанските новинари шетаа низ градот барајќи теми што би биле интересни за известување, Алисон реши да пишува за расипаниот семафор на една важна крстосница што бил расипан цела година. Откако научи како да бара податоци на интернет, тој ја побара точната бројка на регистрирани возила во градот и вкупната сума на даноци што ги платиле сопствениците на автомобили. Тој пишува:

Ситуацијата во Жануарија се влошува поради големиот број возила во градот. Според „ИБГЕ“ (IBGE – водечки институт за статистички истражувања во Бразил), во 2010 година во Жануарија имало 13.771 возила (7.979 од нив мотоцикли)... Локалното население верува дека оддолжувањето на поправката на семафорот не се должи на недостиг на ресурси. Според Секретарот на трезорот (министер за финансии) на сојузната држава Минас Жераис, во 2010 година градот од сопствениците на возила собрал 470 илјади реи на име на разни давачки.

Пристапот до податоците му овозможи на Алисон да покаже дека во Жануарија има многу автомобили (безмалку еден автомобил на пет жители) и дека расипаниот семафор претставува опасност за многу луѓе. Понатаму, тој можеше на своите читатели да им покаже колку вкупно средства градот собрал на име на разни давачки од сопствениците на возила и, врз основа на тој износ, да праша како е можно да нема пари за поправка на семафорот за да се подобри безбедноста на возачите и пешаците.

Иако во двата случаи се работи за мошне едноставни стории, Сораја и Алисон покажуваат дека и граѓанските новинари можат да ги користат податоците во пишувањето стории. Не е нужно да бидете дел од голема редакција, со многу специјалисти, за да користите сурови податоци во своите статии. По само 12 работилници, Сораја и Алисон, без какво било претходно новинарско искуство, беа способни да работат на стории што произлегуваат од сурови податоци и да напишат интересни статии за проблеми во нивната локална заедница. Освен тоа, нивните статии покажуваат дека податоците можат да бидат корисни и кога се работи за мали проекти. Со други зборови, и малите групи на податоци и табели, не само огромните бази на податоци, содржат полезни и вредни информации.

– Аманда Роси, „Пријателите на Жануарија“ (Amanda Rossi, *Friends of Januária*)

Големата табла со изборни резултати

Иако изборните резултати нудат извонредни можности за визуелно прикажување приказни, долго време ја испуштавме таа можност. Во 2008 година, заедно со графичкиот оддел, решивме дека тоа мора да се промени.

Сакавме да пронајдеме начин за прикажување на резултатите што раскажуваат некоја приказна, без да изгледаат само како замрсен збир на броеви во табела или поставени на мапа. А токму тоа го правевме со резултатите претходните избори (<http://nyti.ms/senate-1>; <http://nyti.ms/senate-2>; <http://nyti.ms/senate-3>).

Тоа, секако, не значи дека нешто е нужно погрешно во прикажувањето купишта броеви, она што јас го нарекувам „моделот на Си-Ен-Ен“ („CNN model“, кој се потпира на прикажувањето на постојано нови и нови табели). Таквиот пристап функционира затоа што на читателите им го дава токму тоа што тие сакаат да го знаат: кој победил?

Нормално, обидите да „чепкате“ во нешто што во основа функционира сосема добро со себе носат голема опасност. Обидот да понудиме нешто што е радикално поразлично и што отстапува од очекувањата на луѓето, наместо да ги направи нештата поедноставни, можеше да ги направи покомплицирани.

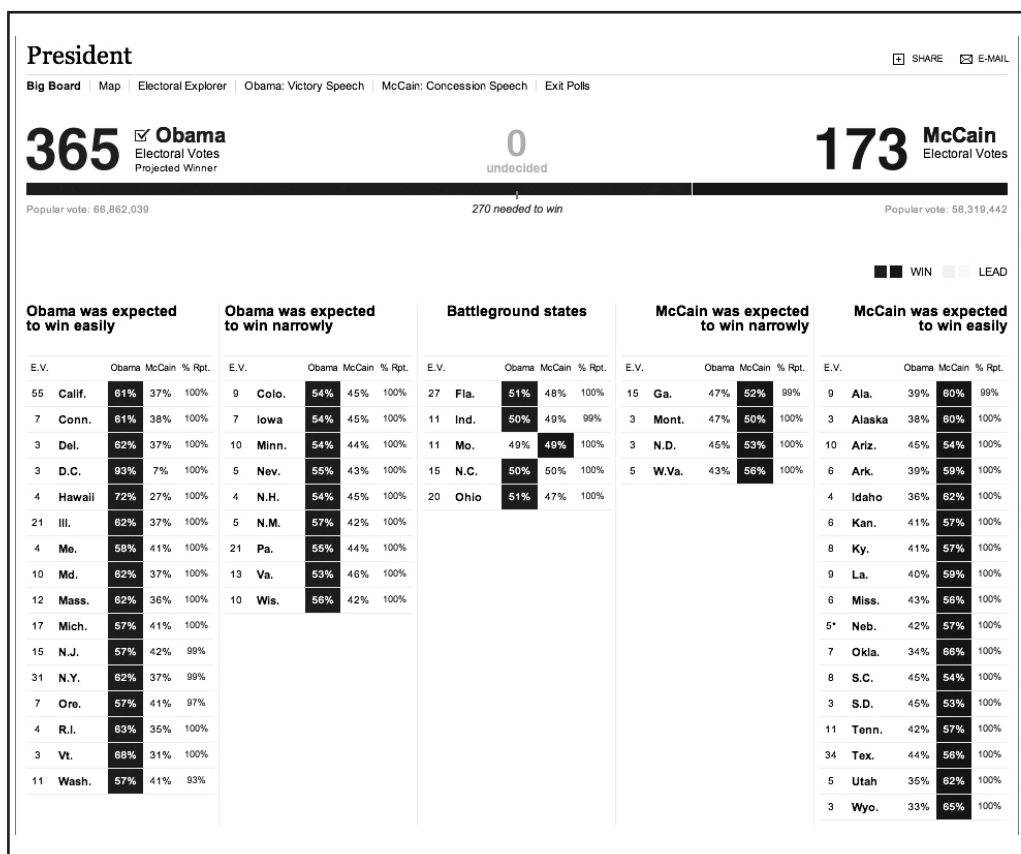
На крај, вистинскиот одговор на нашата дилема го понуди Шен Картер (Shan Carter) од графичкиот оддел, нешто што на крај го нарековме „големата табла“. Кога прв пат ги видов скиците, буквално застанав за да проверам дали сум буден или сонувам.

Го добивме токму тоа што ни требаше.

Зошто сметам дека се работи за извонреден пример за визуелно новинарство? Како прво, погледот на читателот веднаш се задржува на големата хоризонтална „прачка“ („bar“) на врвот на приказот на која се прикажани резултатите од изборите, она што во контекст на новинарството се нарекува „лид“ (lede, „насловница“). Таа му го кажува на читателот токму тоа што тој или таа сака да го знае, и тоа го прави веднаш, на едноставен начин и без каков било „визуелен шум“.

Понатаму, читателот преминува на државите групирани во пет колони, според прогнозата на „Тајмс“ (The Times) за тоа кој кандидат ќе победи во која држава. Во средишната од петте колони се наоѓа она што во новинарството би го нарекле антрфиле со носечката вест (nut graph), и таму објаснуваме зошто победата му припадна на Обама. Интерактивниот сегмент е кристално јасен: Обама победи во сите држави каде што неговата победа беше очекувана, како и во четири од петте држави кои можеа да појдат и на едната и на другата страна.

Сметам дека токму таквата конструкција од пет колони е пример за разликите помеѓу визуелното новинарство и другите форми на дизајн. Кога сè е беспрекорно, добрата визуелна историја истовремено изгледа прекрасно и ин-



Слика 3-25. Големата табла со изборните резултати („Њујорк тајмс“)

формира. Но, ако треба да одлучува помеѓу приказната и естетиката, првенствената грижа на новинарот треба да биде приказната. Иако таквиот приказ можеби не би бил првиот избор на некој чистокрвен дизајнер, тој мошне добро ја раскажува приказната.

Најпосле, како и секој добар интерактивен приказ на интернет, и овој го повикува читателот да продолжи понатаму. Приказот нуди детали како што се процентите на освојените гласови на граѓаните по држава, бројот освоени „избирачки гласови“, при што известувањето намерно е резервирано за да не им конкурира по/во интересот на главните поенти на приказната.

Поради сето тоа, „Големата табла“ е извонреден пример за визуелно новинарство што речиси совршено се поклопува со проверениот образец на „превртена пирамида“.

– Арон Пилхофер, „Њујорк тајмс“

„Краудсорсинг“ на цените на водата

Почнувајќи од март 2011 година, информациите за цените на водата за пиење од градските водоводи во Франција се собираат во еден експеримент за собирање податоци делегирано на поголема група граѓани – т.н. „краудсорсинг“ (надворешна набавка од мноштво добавувачи заб.прев.). За само четири месеци, над 5.000 граѓани на кои им беше дојдено преку глава од контролата што големите корпорации ја имаат врз пазарот на вода за пиење, одвоија малку време да ги побараат своите сметки за вода, да ги скенираат и да ги постават на интернет-страницата на проектот „Цената на водата“ (Prix l'Eau, <http://www.prixdeleau.fr/>). Резултат на нивниот труд е едно дотогаш невидено истражување што на едно место собра многу „бубалици“, невладини организации и традиционални медиуми, здружени во намерата да ја подобрат транспарентноста на проектите за водоснабдување.



Слика 3-26. Цената на водата (Фондација „Франс либерти“)

Францускиот пазар на водоснабдување се состои од повеќе од 10.000 потрошувачи (градови и општини што купуваат вода и ја дистрибуираат до даночните обврзници) и водоснабдувачки компании што можат да се избројат на прстите на едната рака. Рамнотежата на моќ во тој олигопол е нарушена во полза на корпорациите, до таа мера што понекогаш тие наметнуваат различни цени за водата во два соседни града!

Француската невладина организација „Франс либерте“ (France Libertés, „Француски слободи“) веќе четврт век работи на прашањата поврзани со водоснабдувањето во целиот свет. Денес таа е фокусирана на подобрување на транспарентноста на францускиот пазар и на оспособување на граѓаните и градоначалниците што се вклучени во преговори за склучување договори за снабдување со вода. Француската влада одлучи да се зафати со тој проблем пред две години, со национален попис на цените и квалитетот на водата. До сега се собрани само околу 3 отсто од податоците. За да ја забрза работата, „Франс либерте“ сакаше да обезбеди директно вклучување на граѓаните во тој процес.

Со помош на тимот на француската информативна веб-страница „ОВНИ“ (OWNI; <http://www.owni.fr>) дизајнирав еден интерфејс за „краудсорсинг“ (<http://www.prixdeleau.fr/>) на кој корисниците можеа да ги постават скенираните сметки за вода и да ја внесат цената што ја плаќаат за водата од чешма. Во последните четири месеци, 8.500 граѓани се регистрираа и беа поднесени и потврдени над 5.000 индивидуални сметки за вода.

Иако не дозволува совршена оцена на ситуацијата на пазарот, таа им покажа на сите инволвирани, како што се националните надзорни тела за водоснабдување, дека во општеството постои реална загриженост за цената на водата од чешма што доаѓа одоздола, од граѓаните. Граѓаните отпрвин беа скептични околу целата таа приказна за транспарентноста, но како што се одвиваше операцијата, го променија мислењето и се придружија на борбата на „Франс либерте“ против нетранспарентноста и корпоративните малверзации. Што можат медиумските организации да научат од овој пример?

Прифаќајќи ги невладините организации како партнери

На невладините организации (НВО) им требаат огромни количества податоци за да можат да креираат документи за јавни политики. Тие ќе бидат поподготвени да ги покријат трошоците на некоја операција за собирање податоци од управата на еден весник.

Корисниците можат да ви обезбедат сурови податоци

„Краудсорсингот“ е најуспешен тогаш кога собирањето или прочистувањето на податоците им е препуштено на корисниците.

Побарајќи на извори

Размислувавме дали да побараме од корисниците да ги скенираат и да ни ги достават оригиналните сметки, плашејќи се дека тоа ќе одбие некои од нив (особено поради тоа што нашата целна публика се состоеше од лица што се постари од просекот). Иако можеби таквото барање одби некои корисници, тоа, од друга страна, ја зголеми веродостојноста на податоците.

Воспоставејте инструмент за потврда на веродостојноста

Осмисливме систем за бодирање и инструмент за проверка на податоците од страна на другите корисници, за потврда на поднесените податоци. Системот се покажа како премногу замрсен и на корисниците не им нудеше

мотив повторно да ја посетат интернет-страницата. Тимот на „Франс либерте“, од друга страна, го прифати тој систем, а десеттемина вработени во организацијата пронајдоа дополнителна мотивација во системот за бодирање.

Не комплицирајте

Воспоставивме механизам за автоматско испраќање електронска пошта што на корисниците им овозможуваше да поднесат барање за пристап до информации за ценовната политика во водоснабдувањето со само неколку кликови на глумчето. Иако се работеше за иновативна алатка со добри дизајнерски решенија, таа не резултираше со значаен број поднесени барања (беа испратени само околу 100 барања).

Образете ѝ се на вашата публика

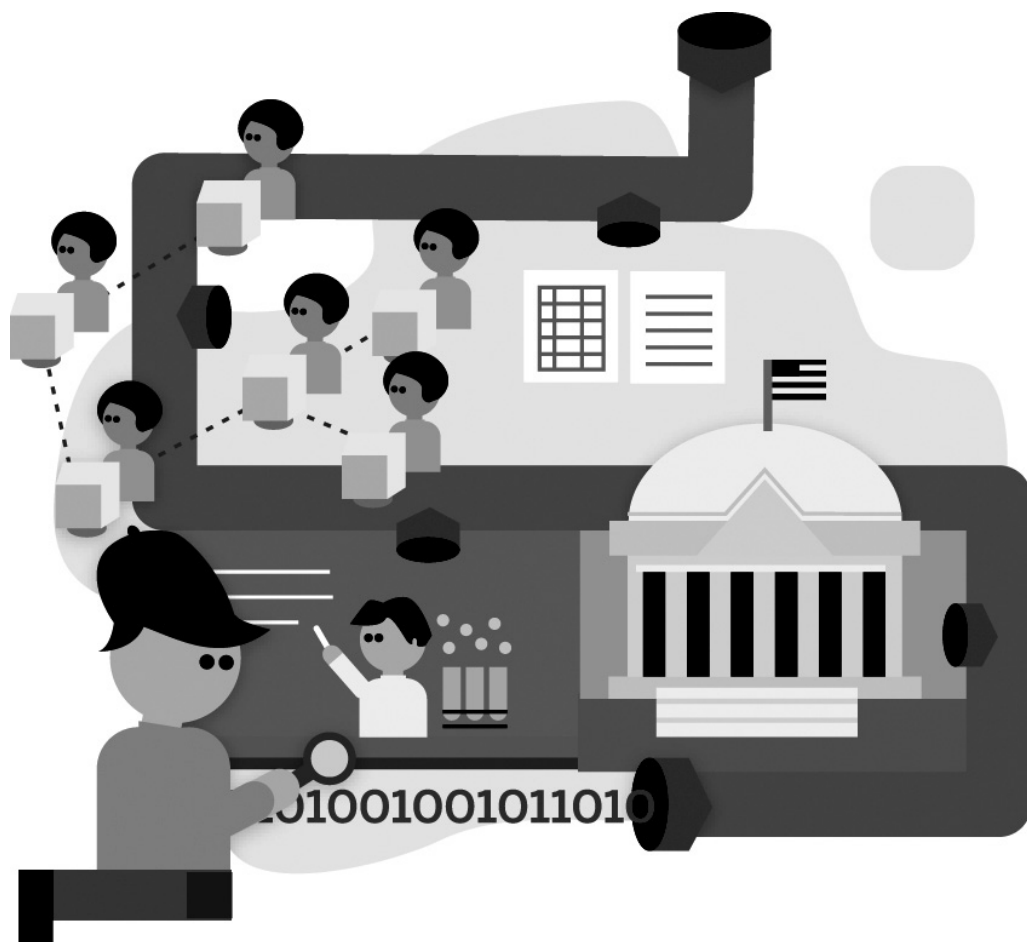
„Франс либерте“ стапи во партнерство со џуз-магазинот за потрошувачки права „60 Millions de Consommateurs“ („60 Милиони потрошувачи“), кој обезбеди целосно вклучување на своите читатели. Тоа партнерство се покажа совршено за тековната операција.

Внимателно бирајте ги клучните показатели за успехот

Проектот имаше само 45.000 посетители за четири месеци, што е сообраќај што сајтот на „Њујорк тајмс“, <http://www.nytimes.com/>, го добива за 15 минути. Она што навистина е важно е дека секој петти посетител се регистрира на сајтот, а секој десетти посетител или посетителка одвои време да ја скенира и постави својата сметка за вода.

— *Николас Кајзер-Брил (Nicolas Kayser-Bril), Journalism++*

Собирање на податоците



Значи, подготвени сте да почнете со реализација на вашиот прв проект во областа на дата-новинарството. Што понатаму? Првата работа што ви е потребна се некакви податоци и во ова поглавје ќе ви покажеме каде можете да ги пронајдете. Ќе научите како да побарате податоци на интернет, со повикување на законите за слободен пристап до информации, како се користи „стругањето на екрани“ (screen-scraping) за собирање податоци од неструктурирани извори и како се користи „краудсорсингот“ за да добиете податоци од вашите читатели. На крај, ќе видиме и што велат законите за повторно објавување групи податоци и ќе разгледаме неколку едноставни правни алатки што овозможуваат да им дозволиме и на други луѓе да ги користат нашите податоци.

Краток преглед на темата на работа

Барате податоци за определена тема или прашање? Не сте сигурни што сè постои на таа тема или каде да барате? Не знаете од каде би почнале? Во ова поглавје ќе се посветиме на прашањето од каде да почнеме во пронаоѓањето на изворите на податоци што се достапни за јавноста на интернет.

Стеснете го пребарувањето

Иако понекогаш се тешки за наоѓање, пребарувачите индексираат голем број од базите на податоци што постојат на интернет, без оглед дали тие што ги објавиле го сакале тоа или не. Следат неколку корисни совети:

- Кога барате податоци, осигурајте се дека во полето за пребарување ќе внесете термини што се поврзани со содржината на податоците што ги барате, но и со форматот во кој сметате дека се чуваат или со изворот каде што очекувате дека ќе ги пронајдете. „Google“ и другите пребарувачи дозволуваат пребарување по вид на датотека. Можете, на пример, да барате само сметководствени табели (со додавање на фразите „filetype:XLS filetype:CSV“ на вашите критериуми за пребарување), геодетски податоци („filetype:shp“) или извадоци од бази на податоци („filetype:MDB, filetype:SQL, filetype:DB“). Ако сакате, можете да барате дури и „PDF“-документи („filetype:pdf“).
- Можете да пребарувате и по сегменти од „URL“-адреси. Пребарувањето на „Google“ со критериумот „inurl:downloads filetype:xls“ ќе ги прикаже сите „Excel“-датотеки што во адресата го содржат зборот „downloads“ (ако добиете само еден резултат, често вреди да се провери кои други резултати се јавуваат во папката на веб-серверот каде што се чува таа датотека). Пребарувањето можете и да го ограничите само на резултатите што се наоѓаат на едно име на домен, преку внесување, на пример, на критериумот „site:agency.gov“.
- Еден друг популарен трик е да не ја барате содржината директно, туку да ги барате местата каде што се чуваат големи количества податоци. На

Одете директно на изворот

Првиот трик што го користам за да ги добијам податоците од некое јавно тело или институција е да се обидам да стигнам директно до тој што ги чува податоците, а не до лицето за односи со јавноста и не преку барање за пристап до информации. Се разбира дека можам да направам барање за пристап до информации или јавни документи, но тоа ја забавува работата. Голема е веројатноста дека ќе добијам одговор дека податоците не се чуваат во форматот што мене ми треба или (и тоа навистина се случи неколку пати) дека владиното тело користи комерцијален софтвер и не може да ги извлече податоците во бараниот формат. Но, ако успеам прво да дојдам до лицето задолжено за ракување со податоците во таа организација, можам да поставам прашања дали имаат податоци за темата што ме интересира и како ги чуваат. Можам да откријам во кој формат се чуваат. Можам да говорам со јазикот на податоците и да дознаам сè што ми треба за да биде барањето успешно. Прашувате дали има пречки за таквиот приод? Често е многу тешко да стигнете до тие луѓе.

Службеникот за јавни информации (СИ) ќе сака да се обратам до неговата канцеларија. Открив дека во такви случаи најдобро е да се обидам да организирам конференциски телефонски разговор или, уште подобро, директна средба на која ќе присуствуваме службеникот за јавни информации, нивниот „гуру“ за податоци и јас. На состанокот можам работите да ги поставам така што нема да можат да кажат не. Им кажувам дека не сакам да им отворам премногу работа. Им велам дека не сакам да подготвам непотребно тешко или прешироко барање, така што средбата би ми помогнала да разберам што точно поседуваат и како на најдобар начин да го побарам точно тоа што ми е потребно.

Ако таквиот пристап се покаже неуспешен, резервна варијанта ми е прво да побарам пристап до информациите за систематизацијата на архивите и појмовник на податоците што ги чуваат. Дури потоа го поднесувам барањето. Понекогаш знам и прво да прашам како ги чуваат податоците и во кој систем. На тој начин, можам да ги проверам можните начини за извлекување на податоците и пред да го напишам барањето.

За крај, најголемиот успех што сум го постигнал е од времето кога работев во еден мал локален весник во Монтана. Ми требаа некои податоци од окружните власти за кои ми беше кажано дека не можат да се експортираат надвор од главниот компјутерски терминал. Истражував малку и се понудив да појдам кај нив и да им помогнам. Заедно со службеникот за податоци напишавме кратка скрипта и ги испечативме податоците на компјутерска дискета (ова се случуваше многу одамна). Јас ги добив саканите податоци, а окружната администрација беше оспособена да одговара на барањата за пристап до податоците. Иако тие немаа таква намера, сепак и тие понекогаш требаше да извлекуваат податоци, а не го разбираа целосно системот што го користеа, така што си помогнавме едни на други.

-- *Черил Филијс, „Сиеџл џајмс“ (Cheryl Phillips, The Seattle Times)*

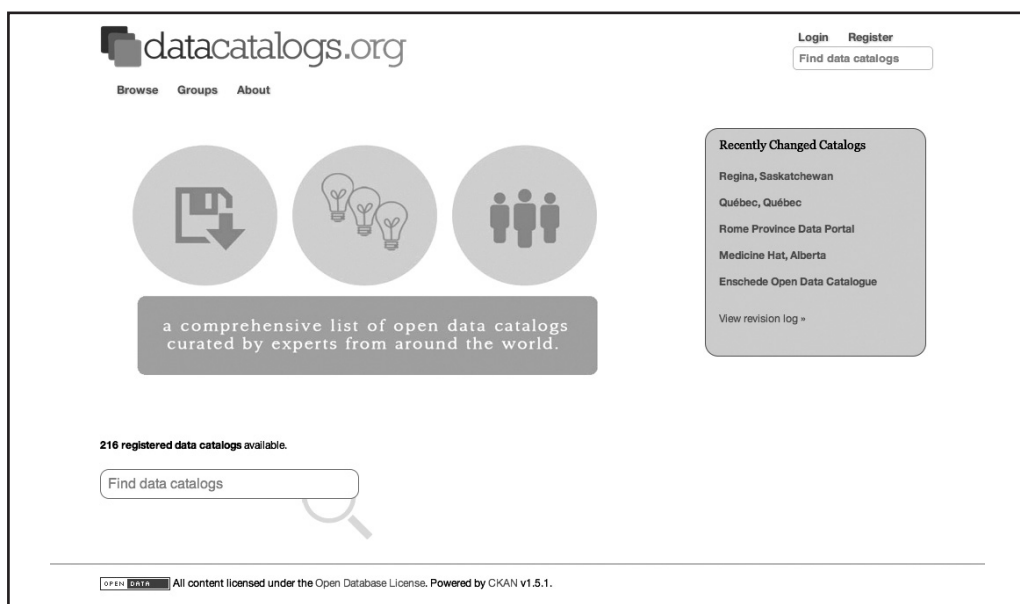
пример, пребарувањето со критериум „site:agency.gov Directory Listing“ ќе ви понуди листинзи генерирани од веб-серверот што овозможуваат лесен пристап до суровите датотеки, додека критериумот „site:agency.gov Database Download“ ќе ги бара само листинзите што се креирани наменски.

Прегледување на онлајн-сервисите за чување и давање податоци

Во последниве неколку години, на интернет се појавија повеќе специјализирани портали, серверски центри и други сајтови за чување податоци. Од нив можете да дознаете повеќе за тоа какви видови податоци постојат на интернет. За почеток, би можеле да ги посетите:

Официјалниите портали за чување податоци

Подготвеноста на владата за објавување определени податоци варира од една држава до друга. Сè поголем број држави основаат „дата-портали“ (инспирирани од примерите на data.gov во САД и data.gov.uk во Велика Британија) за да ја поттикнат граѓанската и комерцијалната употреба на владините информации. На адресата <http://datacatalogs.org/> можете да пронајдете еден одличен и редовно ажуриран список на такви веб-портали. Друга корисна веб-страница е „Податоците од владите во светот на Гардијан“ (Guardian World Government Data), мета-пребарувач што нуди бројни меѓународни каталози со владини податоци.



Слика 4-1. datacatalogs.org (Фондација „Отворено знаење“)

„Дајна хаб“ (Data Hub)

Ресурс ориентиран кон заедниците, со кој управува Фондацијата „Отворено знаење“ (Open Knowledge Foundation). Ги олеснува пронаоѓањето, размената и повторната употреба на јавно достапните извори на податоци и работи на автоматизација на процесот.

„Скрејпервики“ (ScraperWiki)

„Скрејпервики“ е онлајн-алатка што го „олеснува процесот на извлекување корисни податоци за да можат да бидат повторно искористени во нови апликации, или низ нив да „ријат“ новинари и истражувачи“. Повеќето „стругачи“ и нивните бази на податоци се јавни и можат повторно да се користат.

Порталиите со податоци на Свешката банка и на Обединетите нации

Овие сервиси нудат показатели со највисоко ниво на квалитет за сите држави, често со години наназад.

„Базата“, „Инфочимс“ и „Датамаркет“ (Buzzdata, Infochimps, DataMarket)

Се работи за нови „старт-ап“-потфати што имаат за цел да изградат заедници околу прашањата поврзани со размената и препродажбата на податоци.

„Датাকাуч“ (DataCouch)

Место каде што можете да ги снимите, прочистите, разменуваат и визуелизирате вашите податоци.

„Фрибејс“ (Freebase)

Интересна филијала на „Google“ што нуди „збир на меѓусебно поврзани луѓе, локации и предмети, изградена од заедница вљубена во отворените податоци“.

Истражувачки податоци

Постојат бројни агрегатори на истражувачки податоци, како што е „UK Data Archive“, организирани по држава или по научна дисциплина. Иако постојат многу податоци со слободен пристап, постојат и податоци што се достапни само со претплата или податоци за чие натамошно користење или редистрибуција мора да добиете дозвола.

Собирање податоци од архиви што се чуваат на хартија

Веднаш по објавувањето на документите на воените сили на САД од Авганистан и Ирак од страна на „Викиликс“, решивме да го приспособиме тој концепт за одбележување на 50-тата годишнина од Војната во Алжир со објавување на воените дневници од таа војна. Се зафативме со собирање и дигитализација на архивите на француските воени сили за војната во Алжир. Тие документи се достапни во Архивот на Министерството за војна во Париз. Впрочем, документите се достапни печатени на хартија. Испративме новинари и студенти да ги фотографираат документите. Се обидовме да ги скенираме со преносен скенер „Canon P-150“, но обидот не успеа, пред сè, затоа што повеќето од документите се поврзани со хефталици.

На крај, успеавме за неколку недели да собереме околу 10.000 страници разни документи. Ги обработивме со софтвер за препознавање текст (ABBY FineReader), но резултатите беа лоши. Уште полошо, министерството произволно реши да не ни дозволи пристап до кутиите со најинтересните документи во архивата. Пред сè, министерството го забранува повторното објавување документи што инаку можат слободно да се фотографираат во неговите простории. Поради тоа заклучивме дека не вреди да се ризикува и целиот проект го ставивме на чекање.

– Николас Кајзер-Брил (Nicolas Kayser-Bril), Journalism++

Поставете прашање на некој форум

Пребарајте во архивата на одговори или поставете прашање на интернет-страниците „Get the Data“ или „Quora“. „GetTheData“ е веб-страница на која можете да поставувате прашања поврзани со податоци. Каде можете да пронајдете податоци за определено прашање? Како да пребарувате или како да повлечете податоци од определен извор? Кои алатки да ги користите за визуелен преглед на некоја група податоци? Како да ги прочистите податоците или да ги префрлите во формат со кој можете да работите?

Прашајте преку „мејлинг-листа“

Мејлинг-листите (список на сите примачи на електронска пошта на определена тема) ја собираат сета мудрост на една заедница собрана околу определена тема. За новинари што работат со податоци, одлични почетни точки се листите „Data-Driven Journalism List“ и „NICAR-L“. Двете листи се полни со дата-новинари и „гикови“ (geek, лице што се занимава со определена област, вообичаено интелектуална или комплексна, и само затоа што е интелектуална или комплексна) во областа на известувањето со помош на компјутер (УПК), што работат на секакви можни проекти. Голема е веројатноста некој веќе да работел на слична сторија и може да ви даде идеја од каде да почнете, а можеби и линк до податоците што ви требаат. Можете да го испробате и проектот „Вомбат“ (Project Wombat (<http://project-wombat.org/>)) „дискусиона листа за тешки референтни прашања“, голем број мејлинг-листи на Фондацијата „Отворено знаење“, мејлинг-листите на „theInfo“ или да побарате кои мејлинг-листи постојат за темата или регионот за кој сте заинтересирани.

Придружете ѝ се на групата „Хекс/хакерс“

„Хекс/хакерс“ е брзорастечка меѓународна организација за „грасрутс“ (grassroots, активност што произлегува од поширока граѓанска база, наспроти активностите наметнати одозгора заб.прев.) новинарство, со дузини ограноци и илјадници членови на четири континенти. Нејзината мисија е создавање мрежа од новинари („hack“, жаргонски термин за новинар, најсоодветен превод би било „пискарало“) и технолози („hackers“, хакери) кои ја преосмислуваат иднината на вестите и информациите. Во таква широка мрежа секогаш постои голема веројатност некој да знае каде можете да го побарате тоа што ви е потребно.

Прашајте експерт

Универзитетските професори, јавните службеници и претставниците на индустријата често знаат каде треба да се побара нешто. Побарајте ги по телефон. Напишете им електронска порака. Пријдете им на некој настан. Посетете ги во нивната канцеларија. Прашајте љубезно. „Пишувам сторија на тема „икс“. Каде можам да го пронајдам ова? Дали знаете кој го поседува ова?“

Дознајте повеќе за информатичките технологии што ги користи владата

Разбирањето на техничкиот и административниот контекст во кој владите ги чуваат своите информации често може да помогне кога се обидуваме да пристапиме до податоци. Без оглед дали се работи за „CORDIS“, „COINS“ или „THOMAS“, големите бази на податоци именувани со акроними често стануваат најкорисни дури откако ќе научите нешто за целта за која се наменети.

Пронајдете ги владините организациски дијаграми и побарајте ги оние оддели/единици што имаат повеќе полиња на одговорност (на пример, известување, ИТ-услуги) и прегледајте ги нивните веб-страници. Многу податоци се чуваат во повеќе различни оддели и додека некои од нив потребната база на податоци ја чуваат како најголема скапоценост, друг оддел може да ви овозможи слободен пристап.

Побарајте ги динамичните инфографики на владините интернет-страници. Тие често се придвижувани од структурирани извори на податоци/АПИ-интерфејси што можат да се користат и независно (на пример, мали апликации за следење авионски летови, апликации за временска прогноза изработени во програмскиот јазик „Java“).

Повторете го пребарувањето

Кога ќе дознаете повеќе за тоа што го барате, пребарајте повторно користејќи фрази и помалку веројатни комбинации од зборови што сте ги забележале по претходното пребарување. Можеби ќе имате повеќе среќа со пребарувачите!

Прегледување на телефонските листинзи

Пред неколку месеци сакав да ги анализирам телефонските листинзи на гувернерот на Тексас Рик Пери (Rick Perry), кој тогаш беше претседателски кандидат. Листинзите ги добив со барање за пристап до јавни документи, но по подолго чекање. Податоците пристигнаа во форма на 120 и кусур страници документи со квалитет на факсимил. Целиот потфат бараше внесување и чистење на податоците, по што следеше проверка преку програмскиот интерфејс на WhitePages.com (онлајн-телефонски именик) за да откријам чии се телефонските броеви наведени на листата на повици.

Преклопувањето на имињата со изборните податоци на државно и сојузно ниво (податоците на „FEC“, Федералната изборна комисија) откри дека Пери контактирал со донаторите на кампањата и т.н. „супер ПАК-ови“ (Political Action Committee – ПАС, Политички акционен комитет - ПАК) преку службените телефони, практика што се среќава со неодобрување и која отвори повеќе прашања за врските на Пери со еден „супер ПАК“ што работеше во негова полза.

— Џек Гилам, *Асошијетед џепс* (Jack Gillum, Associated Press)

Напишете барање за слободен пристап до информации

Ако сте уверени дека некое владино тело ги поседува информациите што ви се потребни, веројатно најдобрата алатка што ви стои на располагање е барањето за слободен пристап до информации. Повеќе за начинот како се поднесува такво барање имате во следната секција.

– Брајан Бојер, „Чикаџо трибјун“ (Brian Boyer, Chicago Tribune), Џон Кеф, Рајо „WNYC“ (John Keefe), Фридрих Линденберг, Фондација „Опен нолиџ“ (Friedrich Lindenberg, Open Knowledge Foundation), Џејн Парк, „Creative Commons“ (Jane Park), Крис Ву, „Хекс/Хакерс“ (Chrys Wu, Hacks/Hackers)

Кога постапувањето според законот не дава резултати

Откако прочитав една стручна статија во која пишуваше дека објавувањето на наодите од хигиенските инспекции во рестораните довело до намалување на бројот на болести поврзани со исхраната во Лос Анџелес, побарав од хигиенските служби во Париз да ми дадат листа на извршените инспекции. Според пропишаната процедура во францускиот Закон за слободен пристап до информации, чекав 30 дена да го добијам одбивањето да ми дадат одговор, а потоа се обрativ до Комисијата за пристап до јавни податоци (француската комисија е позната по кратенката „CADA“) која одлучува за легитимноста на барањата за пристап до информации. Комисијата го потврди моето барање и ѝ нареди на администрацијата да ми ги достави бараните податоци. Администрацијата побара дополнително време од два месеци и „CADA“ го прифати тоа барање. Два месеци подоцна, администрацијата сè уште не беше сторила ништо по моето барање.

Се обидов да дојдам до некои познати (и со длабок џеб) застапници за отвореност на податоците кои би ми помогнале случајот да го дадам на суд (со вкупни трошоци од 5.000 евра и сигурна победа поради поддршката од „CADA“), но тие се плашеа дека тоа ќе им ги наруши односите со официјалните програми за отворени податоци. Ова е еден од неколкуте примери кога француската администрација едноставно го игнорира законот, а официјалните иницијативи не прават ништо за да ги поддржат барањата за податоци што доаѓаат од базата.

– Николас Кајзер-Брил (Nicolas Kayser-Bril), Journalism++

Вашето право на пристап до податоците

Пред да поднесете барање за слободен пристап до информации [Freedom of Information (FOI)], проверете дали податоците што ги барате се веќе достапни – или претходно биле побарани од некој друг. Претходното поглавје понуди неколку сугестии каде би можеле да барате. Ако сте ги разгледале можните опции и сè уште ги немате потребните податоци, тогаш можеби ќе сакате да поднесете формално барање. Следат неколку совети што ќе ви помогнат барањето да биде успешно:

Планирајте однапред за да заштедите време

Размислете за можноста да поднесете формално барање за пристап секогаш кога почнувате да барате информации. Најдобро е да не чекате да ги исцрпите сите други можности. Ќе заштедите време ако барањето го поднесете на почетокот од вашето истражување и ако паралелно спроведувате и други истраги. Бидете подготвени за одолжување на процесот: понекогаш на јавните тела им е потребно повеќе време за процесирање на барањата, и добро е да бидете подготвени за тоа.

Проверете ги правилата за можните трошоци и надоместоци

Пред да го поднесете барањето, проверете ги правилата за можни надоместоци што се бараат за поднесување на барањето или за добивање на информациите. На тој начин ќе се осигурате дека ги знаете вашите права во случај некој јавен службеник одненадеж да ви побара пари. Побарајте ги документите во електронска форма за да ги избегнете трошоците за копирање и поштарина. Во барањето наведете дека претпочитате информациите да ви бидат доставени во електронски формат. Така ќе го избегнете плаќањето надоместок, освен, се разбира, во случаите кога информациите не се достапни во електронска форма, иако денес вообичаено е можно да се скенираат документите што не се дигитализирани и потоа скеновите да се испратат како прилог по електронска пошта.

Познавајте ги своите права

Дознајте кои се вашите права пред да почнете за да знаете што можат и што не можат да прават јавните службеници. На пример, повеќето закони за слободен пристап до информациите пропишуваат временски рокови во кои властите се обврзани да одговорат на барањето. Генерално, тие рокови во повеќето закони се движат од неколку дена до еден месец. Дознајте однапред кои рокови се пропишани во вашата земја и направете белешка за датумот на кој сте го поднеле барањето.

Владите не се обврзани да ги обработуваат податоците за вас, но би требало да ви ги достават сите податоци што ги имаат, а ако се работи за податоци што мора да ги имаат за да ги остваруваат своите законски надлежности, тогаш сигурно мора и да ви ги дадат.

Јасно кажејте дека ги знаете своите права

Вообичаено, законот не налага во барањето јасно да се повикате на законот за слободен пристап до информации, но сепак се препорачува тоа да го сторите, затоа што така покажувате дека ги знаете своите законски права, а поголема е веројатноста дека тоа ќе поттикне правилно процесирање на барањето, во согласност со законот. Треба да се знае дека во барањата што се поднесуваат до ЕУ, важно е да се спомне дека се работи за барање за пристап до документи, а најдобро е јасно да се спомне Уредбата 1049/2001 како основа за барањето.

Не комплицирајте

За сите држави важи правилото дека е најдобро да почнете со едноставно барање за информации и потоа да поставувате дополнителни прашања, откако ќе ја добиете првичната информација. Така ќе го избегнете ризикот јавната институција да побара продолжување на роковите поради „сложеност на барањето“.

Остатите фокусирајте

Барањето за информации што ги поседува еден оддел од јавното тело веројатно ќе биде одговорено побрзо од барањето за кое е потребно да се пребарува низ целото јавно тело.

Одговорот на барањето кај кое има потреба јавното тело да се консултира со трети лица (на пример, приватна компанија што ја доставила информацијата, друга влада на која тоа може да има влијание) може да одземе повеќе време. Бидете упорни.

Размислувајте пошироко од сиговиите на архивското одделение

Обидете се да дознаете какви податоци се собираат и чуваат. На пример, ако дојдете во посед на празен примерок од формуларот што полицијата го пополнува за сообраќајни незгоди, можете да видите кои информации за автомобилските несреќи се бележат, а кои не.

Барањето нека биде прецизно

Размислете пред да го поднесете барањето: дали можеби е двосмислено на еден или друг начин? Тоа е особено важно ако планирате да ги споредите податоците што сте ги добиле од неколку различни јавни тела.

На пример, ако барате бројки за последните три години, некои тела ќе ви испратат информации за последните три календарски години, други ќе ви испратат информации за последните три фискални години, па нема да можете да направите директна споредба. Ако сакате вистинското барање да го скриете во едно поопшто формулирано барање, тогаш барањето треба да биде доволно широко поставено за да ги опфати информациите што сакате да ги добиете, но не толку широко што ќе биде нејасно или ќе го обесхрабри одговорот. Прецизните и јасни барања обично добиваат побрзо и поквалитетен одговор.

Поднесете повеќе барања одеднаш

Ако не сте сигурни каде точно треба да го поднесете барањето, ништо не ве спречува истото барање да го поднесете истовремено до две, три или повеќе тела. Понекогаш различни тела ќе ви дадат различни одговори, но тоа, всушност, може да ви помогне така што ќе ви даде поцелосна слика за тоа какви информации се достапни за темата што ја истражувате.

Поднесете меѓународно барање

Сè почесто е можно барањата да се поднесуваат по електронски пат, така што воопшто не е важно каде живеете. Друга можност е, ако не живеете во државата во која сакате да го поднесете барањето, барањето да го испратите до амбасадата на таа држава, а нивна задача е да го проследат до надлежното јавно тело. Најпрво ќе треба да проверите во амбасадата дали се подготвени да ви излезат во пресрет – може да се случи вработените во амбасадата да не добиле обука на полето на правото на пристап до информации, па во таков случај, подобро е барањето директно да го испратите до надлежното јавно тело.

Најравејте ѝ пробен обиг

Ако планирате истото барање да го испратите до повеќе јавни тела, почнете така што ќе испратите еден нацрт од барањето до неколку од нив, како проба. Пробата ќе ви открие дали ја користите вистинската терминологија за материјалите што ви се потребни, дали воопшто е можно да добиете одговор на барањето, па ќе можете, по потреба, да го ревидирате барањето пред да го испратите до сите тела.

Предвидете ги можните исклучоци

Ако мислите дека вашето барање подлежи на исклучоци од обврската за давање пристап до информациите, во подготовката на барањето издвојте го прашањето што се однесува на потенцијално чувствителни информации од другите информации за кои е разумно да се претпостави дека нема да подлежат на некаков исклучок. Поделете го барањето на две засебни барања.

Побарајте физички ѝрисиџ до документиџ

Ако живеете во близина на местото каде што се чуваат информациите (на пр., во главниот град каде што се чуваат информациите), можете да побарате и увид во оригиналните документи. Тоа може да ви помогне кога истражувате информации кои можеби се содржани во поголем број документи што би сакале да ги прелистате. Таквиот увид треба да е бесплатен и да се закаже во разумно време кое вам ви одговара.

Чувајте ги белешките за барањата!

Поднесете писмено барање и зачувајте една копија или запис за себе, за да можете да покажете дека барањето било испратено, во случај да мора да поднесете жалба ако не ви било одговорено на барањето. Таквите записи и белешки ќе ви послужат како доказ дека сте поднеле барање, ако планирате да пишувате за процесот на барање информации.

Објавете дека сѐ поднеле барање

Одговорот на вашето барање можете да го забрзате ако ја известите јавноста дека сте го поднеле барањето: ако напишете или емитувате статија во која ќе објавите дека барањето е поднесено, тоа ќе изврши притисок врз јавната институција да го разгледа барањето и да одговори на него. Статијата можете да ја ажурирате во текот на целиот процес или откако ќе го добиете одговорот – ако нема одговор до истекот на законскиот рок, можете и тоа да го претворите во вест. Дополнителна корист од објавувањето е што ќе ги образувате припадниците на јавноста за правото на пристап до информации и начинот на кој тоа функционира во пракса.



Постојат неколку одлични сервиси што можете да ги искористите во подготовката на барањето, и на последователната кореспонденција, јавно достапни на интернет, какви што се „What Do They Know?“ (Што знаат тие?) за јавните тела во Велика Британија, „Frag den Staat“ (Прашајте ја државата) за јавните тела во Германија и „Ask the EU“ (Прашајте ја ЕУ) за институциите на Европската Унија. Проектот „Alaveteli“ работи на обезбедување слични услуги во десетици држави во светот.

WhatDoTheyKnow.com

Sign in or sign up

another really handy site by mysociety.org

Home Make a request View requests View authorities Read blog Help

Make a new Freedom of Information request

Start now »

Search over **108988 requests** and **5741 authorities**

Search

First, type in the name of the UK public authority you'd like information from. By law, they have to respond (why?).

Who can I request information from?

WhatDoTheyKnow covers requests to 5741 authorities, including:

- Department of Health 791 requests
- Kent County Council 518 requests
- Department for Work and Pensions 1421 requests
- Scottish Natural Heritage 17 requests
- Ministry of Defence 887 requests
- British Broadcasting Corporation 875 requests
- Royal Mail Group Limited 353 requests
- Wirral Metropolitan Borough Council 591 requests

What information has been released?

WhatDoTheyKnow users have made 108988 requests, including:

Vale of Glamorgan Council answered a request about IT Support Services about 2 hours ago

“ John Wicker: Please find inserted information provided to this unit although I would inform you the service area did have concerns disclosing the detail of the information requested down to the L... **”**

Слика 4-2. „Што знаат тие?“ („Мај соцајети“ (My Society))

Користење на слободниот пристап до информациите за разбирање на потрошувачката

Правото на слобода на информации (Freedom of Information – FOI) го имам користено на неколку различни начини во известувањето за „COINS“, најголемата база со податоци за буџетската потрошувачка, буџетот и други финансиски информации на Владата на Велика Британија. Во почетокот на 2010 година, Џорџ Озборн (George Osborne) најавуваше дека, ако стане „Канцелар на благајната“ (официјално име за британскиот министер за финансии заб. прев.), ќе ја отвори за јавноста базата на податоци „COINS“ за да овозможи поголема транспарентност на државната благајна. Тогаш изгледаше како добра идеја да проверам какви податоци се чуваат во „COINS“ и како се структурирани, па испратив неколку барања за слободен пристап до информации – едно барање за шематскиот приказ на базата на податоци, едно барање за упатствата што вработените во Трезорот ги добиваат за употреба на „COINS“ и едно барање за договорот што Трезорот го склучил со производителот на базата. Сето горенаведено водеше до објавување многу корисни податоци. Исто така ги побарав, а подоцна и ги објавив, сите шифри за трошоците наведени во базата. Така веќе можевме да знаеме што е „COINS“ и пред Џорџ Озборн да биде именуван за Канцелар на благајната во мај 2010 година, а наредниот месец, во јуни 2010, да ја отвори базата на податоци за јавноста. Податоците од „COINS“ беа искористени од неколку веб-страници што ѝ нудеа на јавноста да ги истражува тие податоци – два такви потфати се OpenSpending.org и страницата „Coins Data Explorer“ на „Гардијан“ (Прегледувач на податоците од COINS).

Дополнителните истражувања укажуваа дека во базата недостига „Единицата за владини сметки (ЕВС)“ [Whole of Government Accounts (WGA)], збирка од 1.500 комплетни сметки на телата што се финансираат од јавни фондови. Побарав пристап до податоците од ЕВС за 2008 и 2009 година, но без успех. Го побарав и извештајот од ревизијата на ЕВС со надеж дека таму ќе има објаснување зошто ЕВС не била сметана за погодна за објавување. И тоа барање ми беше одбиено.

Во декември 2011 година, ЕВС беше дел од објавените податоци од „COINS“. Сепак, сакав да се осигурам дека постојат доволно упатства како да се подготви комплетното сметководство за секое од 1.500-те тела опфатени со вежбата ЕВС. Вториот начин на кој го користев слободниот пристап до информации беше да се осигурам дека податоците објавени во рамките на агендата за поголема транспарентност во Велика Британија се добро објаснети и дека ги содржат оние податоци што треба да ги содржат. Доставив барање за пристап до информации за комплетното сметководство на сите јавни тела опфатени со ЕВС.

-- Лиза Еванс, „Гардијан“ (Lisa Evans, the Guardian)

Вклучете ги колежите

Ако вашите колеги се сомневаат во вредноста на барањата за пристап до информации, можеби најдобар начин да ги убедите е да напишете статија врз основа на информациите што сте ги добиле по пат на барање за пристап до информации. Препорачуваме во финалната верзија на статијата што била објавена или емитувана да спомнете дека сте се повикале на законот за пристап до информации, што ќе обезбеди зацврстување на неговата вредност и ќе ја подигне јавната свест за правото на пристап до информации.

Побарајте сурови податоци

Ако сакате бараните податоци да ги прегледувате, анализирате и да ги обработувате со помош на компјутер, тогаш треба да побарате податоците да ви бидат дадени во електронски формат, погоден за машинско читање. Изразете го на јасен и недвосмислен начин, така што ќе прецизирате, на пример, дека ви требаат информации за буџетот во формат „погоден за анализа со софтвер за водење сметководство“. Можете, во барањето, изрично да наведете дека сакате информациите да ви бидат доставени несумирани или во грануларна форма. За повеќе информации за оваа тема прочитајте го извештајот достапен на линкот <http://bit.ly/access-report>.

Прашања што се однесуваат на организации изземени од законите за слободен пристап до информации

Можеби барате информации за невладини организации, приватни компании, верски и/или други организации што се изземени од законите за слободен пристап до информации. Сепак, можно е да побарате информации за таквите субјекти од јавните тела што подлежат на обврските според законите за слободен пристап до информации. На пример, можете да побарате од некоја владина служба или министерство да ви одговори дали финансирала или соработувала со определена приватна компанија или невладина организација и да ги побарате документите што се однесуваат на таквата соработка. Ако ви е потребна дополнителна помош во подготвувањето барање за пристап до информации, може да ја побарате во „Прирачникот за новинари за пристап до информации од јавен карактер“ на организацијата „Legal Leaks“.

– Хелен Дарбишер (Helen Darbishire), Access Info Europe, Ѓорџе Пагејски (Djordje Padejski) од програмата за новинарство „Најџ“ на Универзитетот Стенфорд (Knight Journalism Fellow, Stanford University), Мартин Росембаум, Бу-Бу-Си (Martin Rosenbaum, BBC), и Фабрицио Скролини, Лондонска школа за економија и политички науки (Fabrizio Scrollini, London School of Economics and Political Science)

Пристапот до информации функционира.

Користете го!

Користењето на законите за слобода на информациите – понекогаш наречено „вобинг“ (wobbing, терминот изворно е дел од професионалниот жаргон на холандските новинари заб. прев.) – е одлична алатка. „Вобингот“ бара методичен пристап и многу упорност. Следат три примери што ги илустрираат добрите и лошите страни на „вобингот“ кои произлегуваат од моите искуства како истражувачки новинар.

Студија на случај 1: Субвенции за земјоделците

Европската Унија секоја година исплаќа безмалку 60 милиони евра во субвенции за земјоделството и агроиндустрискиот сектор. Програмата за субвенции е воспоставена во втората половина на 1950-тите години, и тогаш политичарите објаснуваа дека тие ќе им помогнат на најсиромашните фармери. Сепак, првиот вистински успех, благодарение на едно барање за пристап до информации во Данска во 2004 година, укажуваше дека таквото објаснување е приказна за мали деца. Малите фармери се наоѓаа во проблеми до гуша и често се жалеа на таквата состојба и јавно и во приватни разговори, а во реалноста, најголемиот дел од парите одеше кај големите земјопоседници и агроиндустриските компании. Јас сакав да знам дали постои некаков образец што се повторува во цела Европа?

Во летото 2004 година ги побарав податоците од Европската комисија. Комисијата ги добива податоците од државите членки во февруари секоја година. Податоците покажуваат кој аплицирал за субвенции од фондовите на ЕУ, колкави суми се доделени на корисниците и дали фондовите се наменети за обработка на земјата, развој на локалниот регион или како поддршка за извозот на млеко во прав. Комисијата тогаш ги добиваше податоците во форма на „CSV“-датотеки снимени на компакт-диск (CD). Се работеше за голем број податоци кои, начелно, беа лесни за обработка. Секако, ако успеевте да ги добиете.

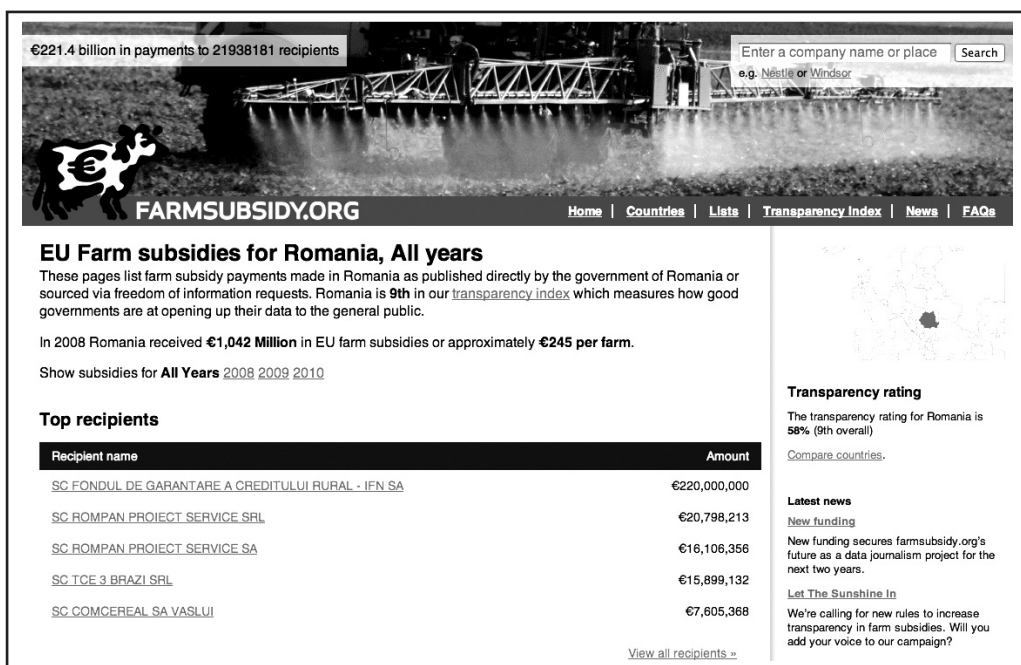
Во 2004 година, Комисијата одби да ги достави податоците, со изговор дека податоците се внесени во база и нивното извлекување ќе бара многу труд и напор. Европскиот омбудсман го карактеризираше тој изговор како „лошо водење на администрацијата“. Сите документи поврзани со тој случај се достапни на интернет-страницата wobbing.eu. Во 2004-тата година немавме многу време да изигруваме правни сладokusци. Не интересираа само податоците.

Затоа се здруживме со повеќе партнерски организации од цела Европа за да ги обезбедиме податоците за секоја држава поединечно. Колегите од Англија, Шведска и Холандија ги добија податоците за нивните земји во 2005 година. Наскоро потоа се отворија и Финска, Полска, Португалија, неколку региони во Шпанија,

Словенија и други држави. Дури и во Германија, традиционално проблематична за „вобингот“, успеав да добијам определени податоци од покраината Северна Рајна-Вестфалија во 2007 година. Навистина, морав да одам на суд за да ги добијам податоците – но од нив подоцна произлегоа неколку одлични статии во магазинот „Штерн“ и неговото издание на интернет („Stern“ и „Stern online“).

Дали беше случајно тоа што Данска и Велика Британија први го отворија пристапот до податоците? Мислам дека не е случајно. Во тој период, на субвенциите за фармерите мораше да се гледа во контекст на преговорите што се одвиваа во Светската трговска организација (СТО) и притисокот врз субвенциите во тие преговори. Данска и Велика Британија спаѓаат во групата најлиберални држави во Европа, па реално е да се претпостави дека политичките ветрови во тие две земји дуваа во правец на поголема транспарентност.

Приказната не завршува таму, и за повеќе епизоди од таа серија, како и за преглед на податоците, посетете ја интернет-страницата <http://farmsubsidy.org/>.



€221.4 billion in payments to 21938181 recipients

Enter a company name or place

e.g. Nestle or Windsor

FARMSUBSIDY.ORG [Home](#) | [Countries](#) | [Lists](#) | [Transparency Index](#) | [News](#) | [FAQs](#)

EU Farm subsidies for Romania, All years

These pages list farm subsidy payments made in Romania as published directly by the government of Romania or sourced via freedom of information requests. Romania is **9th** in our [transparency index](#) which measures how good governments are at opening up their data to the general public.

In 2008 Romania received **€1,042 Million** in EU farm subsidies or approximately **€245 per farm**.

Show subsidies for **All Years** [2008](#) [2009](#) [2010](#)

Top recipients

Recipient name	Amount
SC FONDUL DE GARANTARE A CREDITULUI RURAL - IFN SA	€220,000,000
SC ROMPAN PROIECT SERVICE SRL	€20,798,213
SC ROMPAN PROIECT SERVICE SA	€16,106,356
SC TCE 3 BRAZI SRL	€15,899,132
SC.COMCEREAL.SA VASLUI	€7,805,368

[View all recipients >](#)

Transparency rating

The transparency rating for Romania is **58%** (9th overall)

[Compare countries.](#)

Latest news

[New funding](#)

New funding secures farmsubsidy.org's future as a data journalism project for the next two years.

[Let The Sunshine In](#)

We're calling for new rules to increase transparency in farm subsidies. Will you add your voice to our campaign?

Слика 4-3. Веб-страницата на проектот „Субвенции за земјоделците“ (Farmsubsidy.org)

Лекцијата што ја научивме е дека треба да се бара пристап до податоците. Ситуацијата во Европа е таква што имаме многу различни закони за пристап до информации, а политичките интереси на државите постојано се менуваат. Искористете ја таквата ситуација.

Знајте ги своите права

Дали треба да внимавате на авторските права или други права кога објавувате сурови податоци? Иако е најдобро секогаш да се консултирате со правната служба на вашата организација или медиум, основното правило е дека не мора да барате дозвола за објавување ако се работи за податоци што ги објавила владата. Не грижете се премногу за тоа ако податоците се објавени од организација што не заработува од продавање податоци, но ако се објавени од организација што заработува од продажба на податоци, тогаш дефинитивно треба да побарате дозвола.

— Сајмон Роџерс, „Гардијан“

Студија на случај 2: Несакани дејства

На полето на употреба на лекарства, сите сме лабораториски глупчиња. Лекарствата можат да имаат и несакани дејства. Тоа сите го знаеме и се обидуваме да балансираме помеѓу потенцијалните придобивки и ризици кога ја носиме одлуката. Несреќна околност е што често таа одлука не можеме да ја носиме врз основа на добри информации.

Кога тинејџерите пијат апчиња против акни и мозолчиња, тие се надеваат дека ќе добијат мазна кожа, а не лошо расположение. А токму тоа се случи со едно лекарство што кај младите луѓе предизвикуваше депресија, па дури и склоност кон самоубиство. Немаше доволно информации за опасностите од таквите несакани дејства, што очигледно е добар мотив за новинарите да се заинтересираат за приказната.

Постојат собрани податоци за несаканите дејства. Производителите на лекови се обврзани редовно да им доставуваат на здравствените власти информации за забележаните несакани дејства. Откако некој лек е одобрен за продажба, таквите податоци се чуваат кај националните или европските тела.

Првиот пробив повторно го имавме во Данска. И Холандија ги отвори своите бази на податоци за несаканите ефекти за време на едно прекугранично истражување што го спроведуваше комбиниран данско-холандско-белгиски тим. Уште еден пример за „вобинг“ на слепо беше можноста, во истражувањето на нашиот случај, на холандските власти да им посочиме дека таквите податоци се слободно достапни во Данска.

А приказната дека, како последица од земањето на лекот, во Европа имавме суицидални млади луѓе и, за жал, случаи на самоубиства во неколку држави беше вистинита. Новинарите, истражувачите и семејството на една млада жртва извршија жесток притисок за да добијат пристап до тие информации. Европскиот омбудсман помогна во барањето транспарентност од Европската агенција за лекови (European Medicines Agency) и изгледа дека неговото вклучување во процесот не

беше бадијала. Сега останува новинарите да ја завршат задачата, да ги извлекат податоците и да спроведат детална анализа на добиените материјали. Дали сите сме само обични лабораториски глупчиња, како што рече еден истражувач, или дали контролните механизми навистина функционираат?

Лекцијата што ја научивме е дека, кога се работи за транспарентност, „не“ не е прифатлив одговор. Бидете упорни и следете ја приказната подолго време. Можеби нешто ќе се промени и подоцна ќе добиете можност за поквалитетно известување засновано на подобар пристап до информации.

Студија на случај 3: Шверцување на смртта

Неодамнешната историја може да биде болна за цели групи население, особено во повоени периоди и во време на општествена транзиција. Како, тогаш, новинарите да дојдат до валидните податоци потребни за истражувањата ако, на пример, оние што биле воени профитери минатата деценија денес се на власт? Со таква задача се зафати еден тим од словенски, хрватски и босански новинари.

Тимот ја истражуваше трговијата со оружје во поранешна Југославија за време на ембаргото на ОН во првата половина на 1990-тите години. Основа за тој потфат беа документите собрани од парламентарните анкетни комисии што ја истражуваа таа тема. За да се утврдат и документираат правците за транспорт и да се разбере структурата на трговијата со оружје, беше потребно да се следат индивидуалните пратки во пристаништата според регистрацискиот бројот на бродот, а потоа по регистарските таблички на камионите.

Неколку парламентарни комисии во Словенија спроведоа истрага за прашањето на профитирањето од балканските војни, но не стигнаа до никаков заклучок. Сепак, постоеше една извонредна вредна и значајна трага од декласифицирани документи и податоци, вклучувајќи и 6.000 страници документи што новинарскиот тим од Словенија ги доби преку барање за пристап до информации.

Во овој случај, податоците требаше да бидат извлечени од документите и распоредени во различни бази на податоци. По додавањето дополнителни податоци, нивна анализа и натамошни истражувања, новинарите можеа да ја исцртаат мапата на различните транспортни правци користени во илегалната трговија со оружје.

Тимот успеа и прикажа уникатни и единствени резултати и успеа да ја освои првата од серијата награди за својата работа. Од поголема важност е фактот што таа приказна е значајна за целиот регион и можат да ја искористат и новинарите од другите држави низ кои поминувал смртоносниот товар.

Лекцијата што ја научивме е дека треба да „ископаеме“ добри сурови материјали, дури и ако на нив налетаат неочекувани места, и да ги комбинираме со постојните јавно достапни податоци.

- Бриџит Алфтер (Brigitte Alfter), *Journalismfund.eu*

Поднесете барање за пристап до информации заедно со партнерите

Проблеми со владината корупција има во сите држави на Балканот. Корупцијата често е дури и поголема ако зборуваме за отчетноста и одговорноста на локалните власти во тие држави. Во 2009 година, група новинари од Србија, собрани околу белградскиот „Центар за истражувачко новинарство“ (Centre for Investigative Reporting), потрошија неколку месеци на проверка на различни видови документи добиени со барање за пристап до информации од повеќе од 30 општини во Србија. Претходно, јавноста немаше речиси никаков пристап. Идејата беше да се добијат оригиналните документи од владините архиви и податоците да се префрлат во сметководствени табели, да се спроведат основните проверки и споредби меѓу општините за да се стигне до максималните и минималните бројки. Буџетските податоци, редовните и вонредните трошоци, платите на функционерите, патните трошоци, бројот на вработени, трошоците за мобилна телефонија, дневниците исплатени за службено патување, износите на јавните набавки и други податоци беа користени како основни показатели за проектот. Тоа беше прв случај новинарите да побараат такви информации во земјата.

Од активностите произлезе една сеопфатна база на податоци што открива многубројни случаи на лажни трошоци за репрезентација, злоупотреби и корупција. Листата на најдобро платени градоначалници посочуваше дека некои од нив примаат поголема плата дури и од претседателот на Србија. Многу други функционери беа преплатени и земаа огромни дневници и надоместоци за патни трошоци. Тешко добиените податоци за јавните набавки помогнаа во расветлувањето на она што и официјално претставуваше хаотична ситуација. Од базата на податоци произлегоа повеќе од 150 новинарски стории, а поголемиот број од нив беа објавени од локалните и националните медиуми во Србија.

Научивме дека споредбата на документите со сличните податоци добиени од слични владини тела и институции може да укаже на отстапувањата и девијациите и да фрли светло на можната корупција. Прекумерните и необично високите трошоци можат да се откријат само преку споредба.

— Ѓорѓе Падејски од Програмата за новинарство „Најџ“ на Универзитетот Стенфорд (Djordje Padejski, Knight Journalism Fellow, Stanford University)

Собирање податоци од интернет

Пробавте сè што можеше да се проба и не успеавте да ги обезбедите саканите податоци. Ги пронајдовте податоците на интернет, но нема можност за снимање, а опцијата „копи-пејст“ (copy-paste) не работи. Без паника, можеби сè уште постои начин да се извлечат податоците. Можете, на пример:

- Податоците да ги добиете од програмските интерфејси на апликациите (API), на пример, интерфејсите на базите на податоци достапни на интернет и повеќето современи веб-апликации (вклучувајќи ги и „Твитер“, „Фејсбук“ и многу други). Тоа е одличен начин да се дојде до владини или комерцијални податоци, како и податоци од интернет-страниците на социјалните мрежи.
- Да извлечете податоци од PDF-датотеки. Тоа е навистина тешко, затоа што PDF е јазик за комуникација со печатачите и не задржува многу информации за структурата на податоците прикажани на некој документ.
- Извлекувањето на информации од PDF-датотеки не е дел од темата што ја обработуваме во оваа книга, но постојат алатки и упатства како се изведува.
- Да ги стругате податоците од екраните на веб-страниците (Screen scrape). Кога „стругате“ веб-страници, всушност ја извлекувате структурираната содржина од една нормална интернет-страница со помош на алатка за стругање или со програмирање на неколку линии компјутерски код. Иако се работи за моќен метод што може да се користи на многу места, потребно е извесно разбирање на начинот на кој функционира интернетот.

Покрај сите тие извонредни технички решенија, не заборувајте ги едноставните опции: често е исплатливо да потрошите малку повеќе време и да побарате документ што содржи машински читливи податоци или да појдете до институцијата што ги поседува податоците што ви се потребни.

Во ова поглавје ќе ве спроведеме низ основите на „стругањето“ податоци од една HTML веб-страница.

Што се машински читливи податоци?

Крајната цел на повеќето од овие методи е да се добие пристап до машински читливи податоци. Машински читливите податоци се наменети за компјутерска обработка, а не за нивна презентација на живите корисници. Структурата на таквите податоци се однесува на информациите што ги содржат, а не на начинот на кој ќе бидат прикажани. Примери за формати погодни за машинско читање се датотеките изработени во „CSV“, „XML“, „JSON“ и „Excel“, додека форматите како што се „Word“-документи, „HTML“-страници и „PDF“-датотеки се грижат за визуелниот изглед на информациите. „PDF“, на пример, е јазик за комуникација директно со печатачот; негова главна преокупација се позициите на линиите и точките на една страница, а не буквите или графичките прикази како целини.

„Стругање“ на веб-страници: со каква цел?

Сите ние барем еднаш сме го правеле следново: ќе појдеме на некоја веб-страница, ќе видиме интересна табела и ќе се обидеме да ја копираме во „Excel“ за да можеме да направиме некакви пресметки или да ја зачуваме табелата за подоцна. Тоа понекогаш не е можно, или информациите што ни требаат се расфрлени на повеќе веб-страници.

Копирањето на рака брзо станува заморно, па вреди да размислиме за можноста тоа за нас да го направи специјално напишан компјутерски код.

Предноста на „стругањето“ е што работи буквално на сите веб-страници, од временската прогноза до владините трошоци, дури и ако определена веб-страница не поседува интерфејс за програмирање што овозможува пристап до суровите податоци.

Што може, а што не може да се „струга“

Се разбира, постојат некои ограничувања во поглед на тоа што може да се струга. Следните фактори го отежнуваат стругањето:

- Лошо форматиран HTML-код што содржи малку, или воопшто нема структурни информации (на пример, постарите владини веб-страници).
- Системи за автентификација наменети за спречување автоматски пристап (на пример, заштита со системот „CAPTCHA“ или барањето да се плати за пристапот).
- Системи што работат на принцип на корисничка сесија, со користење „колачиња“ во прегледувачот за следење на активностите на корисникот.
- Непостоење целосен список на достапни предмети или можност за пребарување со „џокер“ (wildcard search).
- Администраторите на серверите го блокираат пристапот „на големо“.

Постојат и определени законски ограничувања: некои држави ги признаваат правата на сопственост и располагање со базите на податоци, што може да го ограничи вашето право на натамошно користење на информациите што биле објавени на интернет. Понекогаш можете и да решите да ја игнорирате лиценцата и да ги искористите таквите информации – некои законодавства им нудат посебни права на новинарите на тоа поле. „Стругањето“ на слободно достапните владини податоци би требало да е дозволено, но можеби ќе треба да проверите што можете да правите со нив пред да преминете на објавување.

Трговските друштва, а и некои невладини организации, имаат многу помало разбирање за тоа и можат да ве обвинат за „саботажа“ на нивните системи. Објавувањето некои информации може да се смета за повреда на приватноста на личноста, а со тоа и да претставува повреда на законите за приватност на личните податоци или на професионалната етика.

Крпење, стругање, компилирање, чистење

Главниот предизвик поврзан со многу видови податоци во Велика Британија не е како да се дојде до нив – туку како да ги добиете во форма во која можат да се искористат понатаму. Податоците за репрезентативните трошоци, другите деловни интереси на членовите на Парламентот, лобирањето и многу други се објавуваат редовно, но на начин што ги прави мошне тешки за анализирање.

Кај некои видови информации, единствената опција е да се засукаат ракавите. Крпењето и спојувањето десетици датотеки изработени во „Excel“, од кои секоја содржи само десет или дваесет внесови, беше единствениот начин да се направи целосна листа на состаноците одржани од министрите во владата. Од друга страна, „стругањето“ на веб-локации се покажа извонредно корисно кај други видови информации.

Користењето сервиси каков што е „СкрејперВики“ (ScraperWiki) за да побараме програмер што ќе ни направи „стругач“ за информации каков што е Регистарот на деловните интереси на членовите на Парламентот, ни заврши половина од работата. Ги добивме сите информации за членовите на Парламентот на една сметководствена табела, подготвена за (вистина долготрајна) работа на анализирање и чистење на податоците.

Таквите сервиси [или алатките каква што е „Аутвит хаб“ (Outwit Hub)] се огромна помош за новинарите што се обидуваат да компилираат замрсени податоци за кои не можат самите да напишат компјутерски код што би ја завршил работата.

— Џејмс Бол, „Гардијан“

Алатки што ќе ви помогнат во „стругањето“

Постојат повеќе програми што можат да се употребат за извлекување податоци од некоја веб-локација, а во нив спаѓаат и дополнителните елементи што се додаваат на прегледувачите, како и неколку веб-сервиси. Во зависност од прегледувачот што го користите, алатките како што е „Ридабилити“ (Readability, <http://www.readability.com/>, алатка за извлекување на текстот од некоја веб-страница) или „Даун дем ол“ (DownThemAll, <http://www.downthemall.net>, дозволува снимање на повеќе датотеки одеднаш) ќе ви помогнат во автоматизацијата на неколку напорни и здодевни задачи, додека додавката за „стругање“ на прегледувачот „Chrome“ (Chrome Scraper extension) е наменета токму за извлекување табели од веб-локации. Отпосле развиените додавки за определени апликации како што е „Фајрбар“ [FireBug, <http://getfirebug.com/>, за прегледувачот „Фајрфокс“ (Firefox) - таа веќе е интегрирана во другите прегледувачи како што се „Chrome“, „Safari“ и „Internet Explorer“] ви овозможуваат да ја следите структурата на некоја веб-страница и да видите кои комуникации се разменуваат помеѓу вашиот прегледувач и серверот на веб-страницата.

„СкрејперВики“ (ScraperWiki, <https://scraperwiki.com/>) е веб-страница што ви овозможува да програмирате „стругачи“ на неколку програмски јазици како што се „Python“, „Ruby“ и „PHP“. „СкрејперВики“ е вистинскиот избор ако сакате да се занимавате со „стругање“ без да морате да се занимавате со програмирање. Други веб-сервиси, како што се „Google Spreadsheets“ и „Yahoo! Pipes“, исто така овозможуваат определено ниво на извлекување информации од други веб-страници.

Како функционираат „веб-стругачите“?

„Веб-стругачите“ (web scrapers) се кратки компјутерски програми напишани во „Python“, „Ruby“, „PHP“ или некој сличен програмски јазик. Изборот на програмскиот јазик најчесто зависи од тоа до која програмерска заедница имате пристап: ако некој во вашата редакција, или во вашиот град, веќе користи некој од наведените програмски јазици, изгледа логично и вие да го прифатите тој јазик.

Додека некои од едноставните „кликни и посочи“-стругачи што ги спомнавме погоре се добра појдовна точка, вистински сложениот дел од „стругањето“ на една веб-локација лежи во изборот на вистинските страници и елементите во нив за да се дојде до саканите информации.

Тука воопшто не се работи за програмирање, туку за разбирање на структурата на веб-страницата или на базата на податоци.

Вашиот прегледувач, во прикажувањето на веб-страниците, скоро секогаш користи една од следните две технологии: „HTTP“ за комуникација со серверот и за барање пристап до специфични ресурси (документи, илустрации или видео) и „HTML“, јазикот на кој се програмирани веб-страниците.

Анатомијата на веб-страниците

Сите HTML-страници се структурирани како хиерархија од „фиоки“ (box) [дефинирани со HTML-ознаки, таканаречени „тагови“ (tag)]. Секоја голема „фиока“ содржи многу помали „фиоки“ – на пример, табела поделена на многу помали единици: редови и ќелии. Постојат многу видови „тагови“ што имаат различни задолженија – некои од нив произведуваат нови „фиоки“, други произведуваат табели, илустрации или линкови (врски до други веб-локации). „Таговите“ можат да имаат и други својства (на пример, можат да служат како уникатни идентификатори) и можат да бидат поделени во групи наречени „класи“, со што овозможуваат насочување и „фаќање“ индивидуални елементи во еден документ. Избирањето на соодветните елементи и извлекувањето на нивната содржина со помош на „таговите“ е клучно во пишувањето на „стругачот“.

Ако ги погледнеме елементите на една веб-страница, сè може да се разложи на помали фиоки сместени во поголеми фиоки.

За да можете да ги „стругате“ веб-страниците, мора да научите нешто за различните видови елементи што можат да се најдат во еден HTML-документ. На пример, елементот `<table>` опфаќа цела една табела, а внатре користи елементи `<tr>` (скратено од „table row“, ред во табелата) за означување на редовите, кои, од своја страна, за секоја одделна ќелија ги содржат елементите означени со `<td>` (скратено од „table data“, податоци во табелата). Најчест елемент со кој ќе се сретнете е `<div>`, што во основа означува секаков можен блок со содржина. Најлесен начин да разберете за што се работи е да го активирате менито со девелоперски алатки во вашиот прегледувач: тоа ќе ви овозможи да го видите основниот код на секој дел од некоја веб-страница со поминување со глумчето преку него.

„Таговите“ ја вршат истата работа што ја вршат потпирачите за книги и ги означуваат почетокот и крајот на една единица. На пример, `<em>` го означува почетокот на сегмент од текстот што е напишан во курзив или почетокот на нагласен сегмент од текстот, а `</em>` го означува крајот на тој сегмент. Навистина е лесно.

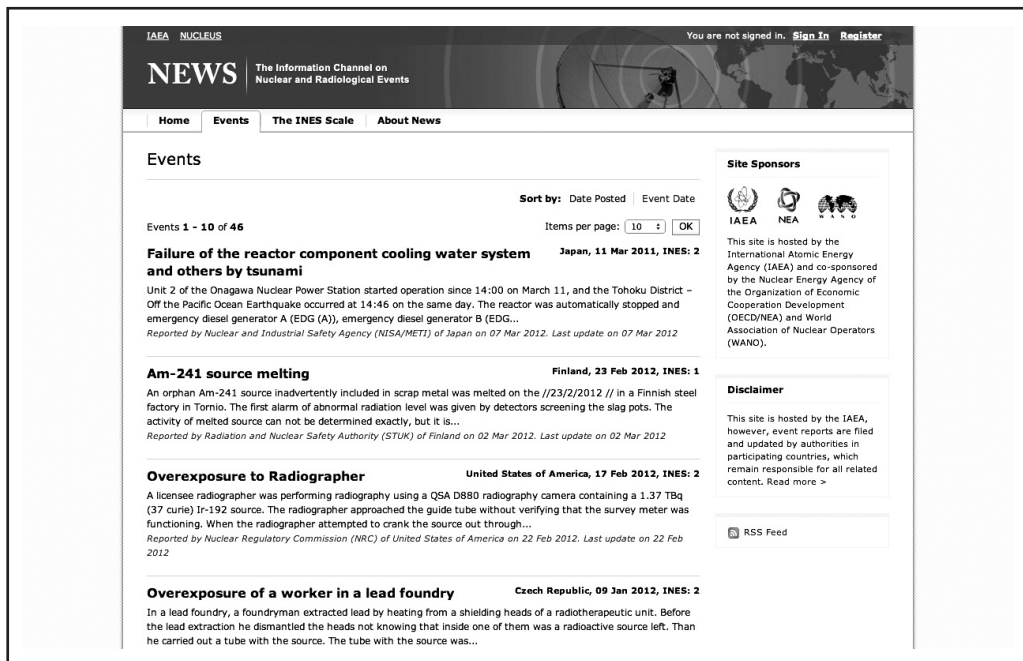
Пример: „Стругање“ на податоците за нуклеарните инциденти со помош на програмскиот јазик „Python“

NEWS (<http://www-news.iaea.org/EventList.aspx>) е порталот на Меѓународната агенција за атомска енергија (ИАЕА) што известува за инцидентите со истекување радијација во светот (истовремено е одличен кандидат за членство во „Клубот на институции со интересни имиња!“). Веб-страницата нуди листа на инциденти во едноставна форма што личи на „блогерска“ веб-страница и лесно може да се „иструга“.

За почеток, креирајте нов „стругач“ во Python на „ScraperWiki“ и ќе добиете поле за внесување текст што главно е празно и содржи само неколку линии код за зацврстување на структурата. Во друг прозорец на прегледувачот отворете ја веб-страницата на ИАЕА и отворете го менито со девелоперски алатки на вашиот прегледувач. Со опцијата „Елементи“ (Elements), обидете се да го пронајдете HTML-елементот на еден од насловите на објавените вести. Менито со девелоперски алатки ви помага да ги поврзете елементите на веб-страницата со основниот HTML-код.

Подетално истражување ќе открие дека насловите се означени со елементи `<h4>` во една табела `<table>`. Секој одделен настан е ред со ознака `<tr>`, а во него се содржани и описот на настанот и датумот кога се случил. Ако сакаме да ги извлечеме насловите на сите настани, ќе треба да пронајдеме начин да ги извлечеме сите редови во секвенца, истовремено повлекувајќи го целиот текст што е содржан внатре во елементите на насловот.

За тој процес да го преточиме во компјутерски код, потребно е да бидеме свесни за сите чекори што треба да се преземат. Затоа предлагаме една еднос-



Слика 4-4. Порталот на Меѓународната агенција за атомска енергија (ИАЕА) (news.iaea.org)

тавна игра: Во прозорчето на „ScraperWiki“, обидете се самите да ги напишете индивидуалните упатства за сè што треба да направите додека го програмирате овој „стругач“, како постапката во некој рецепт за готвење (пред секоја линија внесете „тараба“ (#) за да му кажете на „Python“ дека не се работи за вистински компјутерски код). На пример:

```
# Look for all rows in the table
# Unicorn must not overflow on left side.
```

Обидете се да бидете што е можно попрецизни и не претпоставувајте дека програмата треба да знае нешто за страницата што се обидувате да ја „истругате“.

Откако пишавте неколку реда псевдокод, да го споредиме напишаното со суштинскиот код за нашиот прв „стругач“.

```
import scraperwiki
from lxml import html
```

Тоа е првиот сегмент и овде ги импортираме постојните функции од библиотеките што содржат парчиња однапред напишан компјутерски код. Функцијата scraperwiki ќе ни овозможи преземање цели веб-страници, додека lxml е алатка за структурна анализа на „HTML“-документи. Добрата вест овде е што

ако пишувате „стругач“ во „Python“ со користење „ScraperWiki“, овие две линии код секогаш остануваат исти.

```
url = "http://www-news.iaea.org/EventList.aspx"
doc_text = scraperwiki.scrape(url)
doc = html.fromstring(doc_text)
```

Потоа кодот креира име (променливо): `url`, и тука како вредност ја внесува URL-адресата на страницата на ИАЕА. Тоа му кажува на „стругачот“ дека тоа постои некаде и дека треба да внимаваме на него.

Ќе забележите дека URL-адресата е затворена во наводници, затоа што самата не е дел од програмскиот код туку е низа, секвенца од карактери.

Потоа ја користиме променливата вредност `url` како влезна вредност за функцијата `scraperwiki.scrape`. Функцијата извршува определена задача – во нашиов случај, таа ќе „спушти“ веб-страница на нашиот компјутер. Кога ќе заврши, таа ќе му додели друга променлива излезна вредност, `doc_text`. Вредноста `doc_text` ќе го содржи текстот спуштен од веб-страницата; не визуелната форма што ја гледате во прегледувачот, туку изворниот код со сите „тагови“. Знаејќи дека таквите формулари не се лесни за разложување, ќе користиме друга функција, `html.fromstring`, за генерирање специјален приказ во кој лесно можеме да ги разгледаме индивидуалните елементи, таканаречен „Модел на документ-објектот“ (Document object model - DOM).

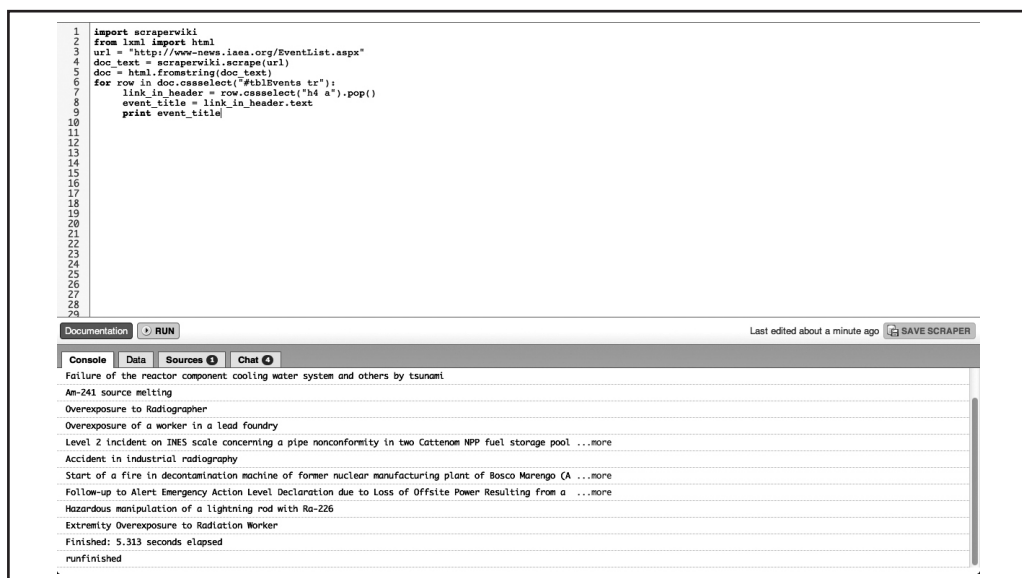
```
for row in doc.cssselect( "#tblEvents tr" ):
    link_in_header = row.cssselect( "h4 a" ).pop()
    event_title = link_in_header.text
    print event_title
```

Во завршниот чекор го користиме приказот DOM за да ги пронајдеме сите индивидуални редови во табелата и да го извлечеме насловот на настанот од „насловницата“. Овде користиме два нови концепти: „for loop“ (наредба за повторливо извршување на кодот) и избор на елементите (`cssselect`). Наредбата „for loop“ во основа го прави токму она што го сугерира и името; наредбата ќе помине низ листа од предмети, на секој од нив ќе му додели привремено име (во нашиот случај, „row“) и ќе ги изврши сите потребни дејства за секој од нив.

Вториот концепт, изборот на елемент, користи специјален јазик за пронаоѓање елементи во документот. Критериумите за избор „CSS“ вообичаено се користат за додавање информации за поставеноста на HTML-елементите и може да се користат за прецизно извлекување некој елемент од една страница. Во нашиот пример (линија 6) го бираме `#tblEvents tr`, кој ќе ги спореди сите тагови `<tr>` во елементот за табели со идентификациски број (ID) `tblEvents` (табелата е ознака за идентификацискиот број). Ќе забележите дека како резултат ќе добиете листа од `<tr>` елементи.

Тоа можеме да го видиме во следната линија (линија 7), каде применуваме друг селектор за да ги пронајдеме сите тагови <a> (што означуваат хиперлинк) во насловот <h4>. Овде сакаме да видиме само еден единствен елемент (секој ред има само еден наслов) и затоа треба да го отстраниме („por off“) од врвот на листата што ја добивме од нашиот селектор со функцијата .por().

Ќе забележите дека некои елементи во приказот DOM содржат вистински текст [т.е. текст што не е дел од некој јазик за означување (markup language)], до кој можеме да пристапиме со користење на синтаксата [element].text од линијата 8. Најпосле, во линијата 9, го печатиме тој текст на конзолата на „ScraperWiki“. Ако го стартувате „стругачот“, во помалиот прозор треба да ви се прикаже листа со имињата на настаните од веб-страницата на ИАЕА.



Слика 4-5. Еген „сџурач“ на дело (ScraperWiki)

Сега гледате еден најосновен стругач на дело: ја спушта веб-страницата, ја трансформира во DOM-приказ и потоа ви дозволува да изберете и извлечете определени содржини. Сега кога го имате костурот, можете да се обидете и сами да решите некој од долунаведените проблеми со користење на „ScraperWiki“ и документацијата на „Python“:

- Дали можете да ја пронајдете адресата на линкот кој води до секој наслов на настан?
- Дали можете да го селектирате малото поле во кое се наведени датумот и местото со користење на CSS-класата на која ѝ припаѓа и да го извлечете текстот од тој елементот?

- ScraperWiki нуди мала база на податоци за секој „стругач“ во која можете да ги чувате резултатите; копирајте го потребниот примерок од нивната документација и приспособете го за да ги зачува насловот на настанот, линковите и датумите.
- Листата на настани се протега на повеќе веб-страници; можете ли да „остругате“ повеќе страници одеднаш за да ги соберете и историските податоци?

Додека се обидувате да ги решите тие задачи, разгледајте наоколу по „ScraperWiki“: Во постоечките „стругачи“ можете да пронајдете многу корисни примери; често можете да најдете и мошне интересни податоци. Така нема да морате да го подготвувате вашиот „стругач“ од нула, само изберете некој што наликува на она што вам ви е потребно и прилагодете го на проблемот со кој се занимавате.

– Фридрих Линденберџ, Фондација „Отворено знаење“

„Стругање“ на јавна база на податоци

На некои лекари во Франција им е дозволено сами да ги определуваат цените на своите услуги, што значи дека 30-минутен преглед кај онколог може да ве чини помеѓу 70 и 500 евра. Според законот, податоците за цените на прегледите се јавни, но надлежната администрација обезбедува само база на податоци на интернет која е тешка за снаоѓање. За да обезбедам квалитетен преглед на цените на медицинските услуги за потребите на „Ле Монд“, решив да ја „истругам“ целата база на податоци.

Тогаш нештата станаа интересни. Формуларот за пребарување на базата на податоци беше изработен во „Flash“ и водеше до HTML-страница со резултати, преку „POST“-барање (барање во кое резултатите се добиваат преку електронска пошта). Со помош на Николас Кајзер-Брил, сфативме дека апликацијата користи и трета страница како „скриен“ чекор помеѓу формуларот за пребарување и страницата со резултати. На таа страница се чуваше „колаче“ со вредностите внесени во формуларот за пребарување кон кои потоа пристапуваше страницата со резултатите. Тешко е да се замисли позамрсен процес, но опциите на библиотеката „cURL“ во PHP помага тие пречки да се надминат со леснотија, секако, откако ќе дознаете каде се лоцирани. Иако ни беа потребни 10 часа работа да дојдеме до базата на податоци, трудот се исплатеше.

– Александар Лешене, „Ле Монд“

Интернет како извор на податоци

Како можете да пронајдете повеќе информации за нешто што постои само на интернет? Без оглед дали пред вас гледате адреса на електронска пошта (email address), веб-страница, илустрација или статија на „Википедија“ (Wikipedia), во ова поглавје ви ги нудиме алатките што ќе ви обезбедат повеќе информации за нив.

Веб-алатки

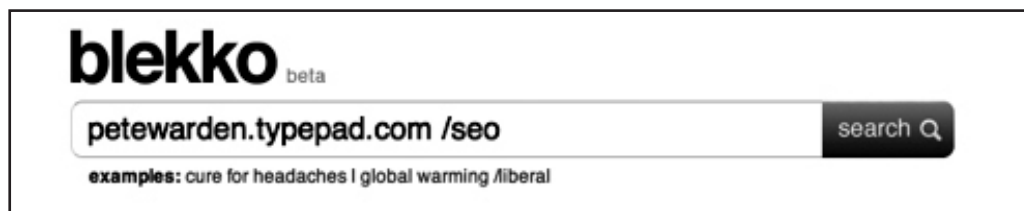
Најнапред да видиме неколку различни сервиси што можете да ги користите за да откриете повеќе за една цела веб-страница, а не само за определена страница од таа веб-страница:

„Whois“ („Koj e“)

Ако појдете на страницата <http://whois.domaintools.com/> (или едноставно внесете „whois www.example.com“ во апликацијата „Terminal.app“ на компјутер со оперативен систем „Мак“ („Mac“), со тоа што вие ќе ја внесете саканата URL-адреса наместо www.example.com) можете да ги добиете основните регистрациски информации за која било веб-страница. Последниве години, некои сопственици на веб-страници избираат да ја заштитат приватноста на податоците во регистрацијата, но во поголемиот број случаи ќе добиете име, физичка адреса, адреса на електронска пошта и телефонски број на лицето на чие име е регистрирана веб-страницата. Можете да ја внесете и нумеричката ИП (IP) адреса и да ги добиете податоците за организацијата или лицето што е сопственик на серверот. Ова е особено ползна алатка кога се обидувате да пронајдете повеќе информации за злонамерните корисници на некоја услуга, корисници што користат навредлив јазик, имајќи предвид дека повеќето веб-страници ги бележат и чуваат IP-адресите од кои е пристапено до нив.

„Блеко“ (Blekko)

Пребарувачот „Блеко“ (Blekko) нуди невообичаено голем увид во статистичките податоци што ги собира за веб-страниците додека „ползи“ по Мрежата. Ако во пребарувачот внесете некое име на домен, заедно со критериумот „/seo“, ќе добиете страница на која се прикажани информациите за таа URL-адреса.



Слика 4-6. Пребарувачот „Блеко“ (Blekko.com)

Првиот табулар на Слика 4-7 покажува кои други веб-страници се поврзуваат со тој домен, подредени според популарност. Тоа може да се покаже особено полезно кога се обидувате да сфатите какво покривање добива одредена веб-страница, или ако сакате да разберете зошто таа страница е високо рангирана на резултатите од пребарувањата на „Google“, бидејќи тоа рангирање се заснова на таквите влезни врски. На Сликата 4-8 можете да видите кои други веб-страници се отвораат од истата машина. Често измамниците и „спамерите“ (spammers, лица или субјекти што испраќаат, најчесто промотивни или злонамерни пораки до голем број адреси, при што пораките не се очекувани и посакувани од примачите заб. прев.) ги покриваат своите траги на патот до легитимност со градење повеќе сајтови што меѓусебно се поврзани и се реферираат едни на други. Тие изгледаат како независни и самостојни домени, податоците од регистрацијата можат и да им се разликуваат, но тие често се сместени на ист сервер, затоа што така е многу поевтино. Овие статистики ви даваат увид во скриената бизнис-структура на сајтот што го истражувате.

Inbound links: 6,050 from 302 domains:					
#	from host	host rank	links	last	actions
1	twitter.com	12,366.4	1		  
2	www.guardian.co.uk	6,481.2	1		  
3	www.forbes.com	3,699.8	1	41d ago	  
4	www.newscientist.com	3,678.4	2		  
5	code.google.com	3,451.1	1		  
6	www.huffingtonpost.com	3,238.2	1		  
7	news.cnet.com	3,185.8	2		  
8	gizmodo.com	2,119.3	6	39d ago	  

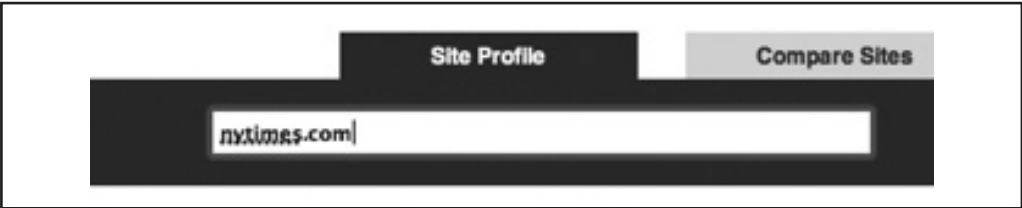
Слика 4-7. Разбирање на популарноста на интернет: кој линкува до кого? Друг „згоде“ табулар е „Крол стаџ“ („Crawl stats“), одобено секцијата „Cohosted with“. (Blekk.com)

Cohosted With:			
host	whois	view	
thelongtail.com	whois		
codinghorror.com	whois		
longtail.com	whois		
cityofsound.com	whois		
hypebot.com	whois		
therestisnoise.com	whois		
stevenberlinjohnson.com	whois		
planetout.com	whois		
riehworldview.com	whois		

Слика 4-8. Окривање на „спамерите“ и измамниците на интернет (Blekk.com)

Compete.com

Преку истражувања и анкетање на просечните американски потрошувачи, <http://www.compete.com/> собира детални статистики за користењето на повеќето веб-страници, а некои од најосновите податоци се достапни без надомест. На менито, изберете го табуларот „Site Profile“ и внесете го саканиот домен (Слика 4-9). Ќе се прикаже графикон со сообраќајот кон таа страница за претходната година, заедно со статистики за бројот на посетители и колку често тие доаѓале на веб-страницата (Слика 4-10). Бидејќи се засновани на анкети и истражувања, броевите се приближни, но открива дека се доволно прецизни кога добив можност да ги споредим со интерните аналитички податоци на неколку веб-страници. Тие се особено добар извор за споредба на две индивидуални веб-страници. Иако апсолутните бројки можеби не се сосема точни, тој сепак претставува добра споредба на релативната разлика во популарноста на двете веб-страници. Од друга страна, тие ги истражуваат само навиките на потрошувачите во САД, па можеби нема да има толку податоци за сајтовите што се наменети претежно за интернационалната публика.



Слика 4-9. Услугаа за профили на веб-локации на Compete.com



Слика 4-10. Што е модерно? Што е барано? – жешките точки на интернет (Compete.com)

Google's Site Search (Пребарувачој на веб-страници на „Google“)

Клучниот збор за пребарување „site“: може да се покаже како особено полезна опција кога се обидуваме да ги истражите сите содржини што постојат на определен домен на интернет. Ако во полето за пребарување го внесете критериумот „site:example.com“, „Google“ ќе ви ги прикаже само резултатите за посочената веб-страница. Можете и дополнително да го стесните пребарувањето ако го вклучите и префиксот на страниците за кои сте заинтересирани, на пример, „site:example.com/pages/“, и ќе ги добиете само резултатите што одговараат на тој образец. Тоа е особено полезно кога се обидуваме да ги пронајдете информациите што сопствениците на домените ги направиле јавно достапни, но не се волни на тоа да му дадат повеќе публицитет, па изборот на вистински клучни зборови може да ве доведе до материјали што откриваат многу работи.

Веб-страници, илустрации и видео

Понекогаш може да се интересирате за активностите околу специфична приказна наместо околу цела една веб-страница. Алатките наведени подолу ви обезбедуваат различни погледи на начините на кои луѓето читаат, копираат, разменуваат и одговараат на содржини на интернет.

Bit.ly

Секогаш се навраќам на bit.ly (систем за кретење на долгите „УРЛ“, адреси) кога сакам да знам на кој начин луѓето разменуваат определен „линк“ (интернет-врска) меѓу себе. Bit.ly се користи со внесување на URL-адресата што ве интересира. Потоа кликнете на линкот „Info Page+“. Тој ќе ве однесе на страницата со комплетните статистики (внимание, можеби ќе мора прво да ја одберете опцијата „aggregate bit.ly link“ ако сте регистрирани како корисник за таа услуга). Таму ќе добиете претстава колку е популарна бараната веб-страница, вклучувајќи ги и активностите на „Фејсбук“ и „Твитер“, а под тие податоци можете да ги проследите и јавните дискусии за тој линк обезбедени од веб-страницата backtype.com. За мене таа комбинација од податоци за сообраќајот и дискусиите е од голема помош кога се обидувам да разберам каде лежи тајната на популарноста на некоја веб-страница или веб-локација, и кои се нејзините следбеници. На пример, таму пронајдов цврсти докази дека доминантниот наратив за „грасрутс“ (grassroots) споделувањето и Сара Пејлин (Sarah Palin) е сосема погрешен.

Твиџер (Twitter)

Како што расте бројот на неговите корисници, веб-страницата за микроблогирање станува сè покорисна како мерач на начините на кои луѓето ги разменуваат или зборуваат за индивидуалните содржини. Леснотијата со која може да ги пронајдете јавните дискусии за некој линк е изненадувачка. Копирајте ја URL-адресата за која сте заинтересирани во полето за пребарување и потоа можете да ја изберете опцијата „повеќе твитови“ (more tweets) за да ги добиете сите резултати.

„Кеш“-опцијајата на Google

Кога некоја веб-страница ќе стане контроверзна, издавачите можат да ја тргнат од интернет или да ги променат нејзината содржина и нејзиниот изглед без најава или соодветна изјава дека тоа го сториле. Ако се сомневате дека се соочувате со токму таков проблем, првото место на кое што можете да проверите е „кешираната“ верзија на таа страница на Google (Google's Cache), односно изгледот на страницата при последното ажурирање од страна на пребарувачот. Фреквенцијата на ажурирањето постојано се зголемува, па најдобро ќе биде ако проверката ја направите во првите неколку часа откако ќе се посомневате дека се случила некаква промена. Внесете ја целната URL-адреса во полето за пребарување на Google и потоа кликнете на тројната стрелка од десната страна на резултатот што се однесува на саканата веб-страница.

Треба да се покаже графички приказ и, ако сте имале среќа, во горниот дел од прегледот ќе има и мал линк со назив „Cache“. Кликнете на тој линк за да ја погледнете зачуваната „слика на екранот“ (snapshot) од страницата што ја чува Google. Ако имате проблеми со „подигањето“ на таа страница, можете да преминете на попримитивната страница со чист текст преку друг линк на врвот на „кешираната“ страница. Најдобро ќе биде да направите „кадар на екранот“ (screenshot) или да ги копирате сите релевантни содржини што ќе ги пронајдете, затоа што може да ги сними по наредната „обиколка“ на интернет.

Уредбата „Вејбек“ на Архивата на интернет (Internet Archive's Wayback Machine)

Ако сакате да дознаете како определена веб-страница се менувала во подолг временски период, во текот на неколку месеци или години, „Архивата на интернет“ (Internet Archive) има сервис што се нарекува „The Wayback Machine“ (буквалниот превод би гласел „Машина што води во дамнешни времиња“, заб.прев.) кој периодично прави кадри на екраните на најпопуларните веб-страници на мрежата. Појдете на сајтот, внесете го линкот што сакате да го проверите и, ако има постари копии, ќе добиете приказ на календар и можете да го изберете точното време во кое сакате да видите како изгледала веб-страницата. Ќе добиете приближна верзија на изгледот на веб-страниците во саканиот момент. Често се случува да недостигаат стилските елементи и илустрациите, но вообичаено приказот е сосема доволен за разбирање на фокусот на содржината на таа страница во тоа време.

View Source (Погледнете го изворниот код)

Иако веројатноста е мала, сепак, девелоперите често оставаат коментари или други траги во HTML-кодот што ја дава подлогата за некоја веб-страница. Во зависност од прегледувачот што го користите, нејзината позиција во мениот може да се разликува, но секогаш постои опцијата „View Source“ (Погледни го изворниот код) што ви дозволува да го прегледате чистиот HTML-код. Нема потреба да го знаете точното значење на машински читливите делови, само внимавајте да не ги пропуштите исечоците текст што се расфрлени меѓу нив. Дури и ако се работи за белешки за авторските права или наведување на името на авторот (или авторите), таму често можете да пронајдете траги за процесот на создавање и намената на веб-страницата.

TinEye („ТинАј“)

Понекогаш навистина сакате да го знаете изворот на некоја илустрација или фотографија, но без јасна легенда што го наведува изворот или авторот, не постои очигледен начин тоа да го сторите со традиционалните пребарувачи каков што е Google. „ТинАј“ (TinEye, „око од калај“, заб.прев.) нуди процес на пребарување на илустрации „наназад“ (“reverse image search” process), каде што ја внесувате илустрацијата што ја имате, а TinEye пронаоѓа други слики на мрежата кои личат на вашата. Поради тоа што користи програма за препознавање слики за споредбата, TinEye работи дури и кога оригиналната илустрација е дисторзирана, сечена или компресирана. Тоа особено е ефективно кога се сомневате дека некоја слика лажно била претставена како оригинална или нова, затоа што може да ве одведе до нејзиниот вистински извор.

YouTube („Јуџјуб“)

Ако кликнете на копчето за статистики (Statistics) во долниот десен агол на видео-приказот, ќе добиете повеќе информации за гледачите што го погледнале тоа видео во подолг временски период. Иако не се комплетни, податоците се корисни за грубо разбирање на тоа кој и кога го погледал видеото и од каде доаѓаат гледачите.

Електронска пошта

Често, кога истражувате некои пораки што сте ги добиле по електронска пошта, сакате да имате повеќе детали за идентитетот и локацијата на испраќачот. Не постои добра „готова“ алатка што може да ви помогне да го сторите тоа, но може да ви помогне ако ги знаете основите на скриените „заглавија“ („header“, „заглавие“, делот од пораката во која се наведени податоците за испраќачот и примачот, заб.прев.) во секоја порака испратена по електронска пошта. Тие функционираат како поштенски жигови и можат да откријат изненадувачки многу податоци за испраќачот. Тоа особено се однесува на фактот дека често ја содржат и IP-адресата на уредот од кој е испратена пораката, слично на идентификацијата на телефонските повици. Потоа можете тој IP-број да го проверите со опцијата „whois“ (за неа говоревме погоре) за да откриете која организација е сопственик на тој уред. Ако се покаже дека уредот припаѓа на организација каква што се телекомуникациските компании „Comcast“ или „AT&T“ кои обезбедуваат интернет-услуги за потрошувачите, тогаш можете да ја посетите веб-страницата „MaxMind“ (<http://www.maxmind.com/en/home>, сервис за географско лоцирање на IP-адреси) за да ја добиете приближната локација.

За гледање на таквите „заглавија“ во „Gmail“ (сервисот за електронска пошта на Google), отворете ја пораката и потоа отворете го менито веднаш до копчето за одговор (Reply) горе десно и потоа изберете ја опцијата „Show original“ (Прикажи ја оригиналната порака).

Ќе се отвори нова страница на која се прикажани скриените содржини. На почетокот ќе има неколку десетици редови текст проследени со точка-завирка (:). IP-адресата што ја барате може да биде наведена во тој дел, но начи-

нот на кој е именувана ќе зависи од начинот на испраќање на електронската порака. Ако е испратена преку „Hotmail“ (сервисот за електронска пошта на „Microsoft“), ќе биде наречена X-Originating-IP, но ако е испратена преку апликацијата „Outlook“ или од сервисот за електронска пошта на „Yahoo“, тогаш ќе биде наведена во првиот ред кој почнува со зборот Received:

Проверката на адресата преку „Whois“ покажува дека таа ѝ била доделена на „Вирџин медија“ (Virgin Media), провајдер на интернет-услуги од Велика Британија, па ја проверив и преку сервисот за геолокација на „MaxMind“ за да откријам дека потекнува од мојот роден град Кембриџ. Тоа значи дека можам да бидам прилично сигурен дека пораката ја испратиле моите родители, а не некакви измамници што се претставуваат дека се моите родители!

Трендови

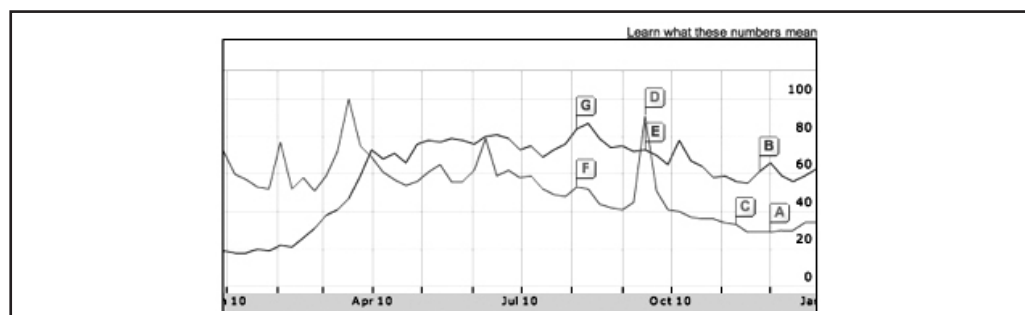
Ако истражувате некоја поширока тема, а не точно определена веб-страница или предмет, следниве алатки можат да ви помогнат во барањето информации:

Сообраќајот кон статиите на „Википедија“ (Wikipedia Article Traffic)

Кога сакате да дознаете како се менувал интересот на јавноста за некоја тема или личност со текот на времето, можете да ги погледнете дневните статистики за посети и прегледувања за која било страница на „Википедија“ на веб-страницата <http://stats.grok.se/>. Сајтот изгледа како да е нафрлан на брзина, но ќе ви овозможи да ги откриете потребните информации со малку копање наоколу. Внесете го името што ве интересира за да го добиете месечниот преглед на сообраќајот кон таа страница. Ќе се прикаже графикон кој покажува колку пати страницата била посетена и прегледана за секој ден од наведениот месец. Проблем е што не можат да се видат податоците за повеќе месеци одеднаш, што значи дека повторно ќе морате да го внесете месецот и повторно да пребарувате за да ги видите промените на подолг рок.

„Гуџл инсајтс“ (Google Insights)

Добра идеја за навиките за пребарување на јавноста може да добиете преку сервисот „Insights“ (Увиди) на Google, (<http://www.google.com/insights/search/>; види Слика 4-11). Внесете неколку чести фрази за пребарување, на



Слика 4-11. Google Insights (Google)

пример, „Justin Bieber vs Lady Gaga“ („Џастин Бибер против Лејди Гага“) и ќе добиете графички приказ на историјата на пребарување за секој од нив посебно. Постојат повеќе опции за рафинирање на податоците што ќе се прикажат, од стеснување на географската област до повеќе хронолошки податоци. Единствено разочарување е недостигот на апсолутни вредности – добивате само релативни проценти што понекогаш се тешки за интерпретација и толкување.

– *Пит Ворден, независен девелопер и аналитичар на податоци*

„Краудсорсинг“ на податоци за „Даталогот на Гардијан“

Според „Википедија“, „краудсорсингот“ е дистрибуиран пристап кон решавањето проблеми и процесот на производство што вклучува распоредување на задачите на една поширока мрежа од луѓе, позната и како толпа (‘crowd’) (во деловниот јазик ќе го пронајдете и како „надворешни набавки од мноштво добавувачи“). Следните неколку реченици се извадени од едно интервју со Сајмон Роџерс, во кое тој зборува за начинот на кој „Даталогот на Гардијан“ (Guardian Datablog) го искористи „краудсорсингот“ во известувањето за скандалот со трошоците на членовите на Парламентот, употребата на наркотици и документите за Сара Пејлин:

Понекогаш може да се случи да добиете купишта документи, папки, статистички податоци или извештаи што сам човек не може ни да се обиде да ги прегледа. Исто така, може да добиете материјали што се тешко читливи или се во непогоден формат за да можете со нив нешто да направите. „Краудсорсингот“ е корисен токму во такви ситуации.

Она што „Гардијан“ сигурно го има е огромен број читатели. Ако имаме интересен проект за кој ни треба помош од страна, можеме да ги прашаме да ни помогнат. Токму тоа го сторивме со проектот „Трошоци на членовите на Парламентот“ („MPs’ Expenses“). Требаше да обработиме 450,000 документи во многу кратко време. Има ли подобар начин од повикот до читателите да ни помогнат во извршувањето на задачата?

Проектот „Трошоци на членовите на Парламентот“ генерираше голем број дојави. Добивме многу повеќе стории отколку чисти податоци. Од гледна точка на сообраќајот што го остваривме, проектот претставуваше извонреден успех. Тој навистина им се допадна на луѓето.

Во моментов соработуваме со „Миксмаг“ (MixMag, <http://www.mixmag.net/>, водечки магазин посветен на електронската „денс“-музика и клупската сцена во светот) на прашањето на користењето дрога и соработката досега е фено-

Page 43 of Stephen Pound's Incidental Expenses Provision 2007/08

This document has 90 pages (see all)
← Previous | Next →

C1

Page 2 of 2

Authorisation and declaration

Signature

3/12/07

Date

MP

Substantially, exclusively and necessarily

Data protection

The House of Commons Administration will process the information you provide on this form for the purpose of administering and accounting for the Member's Estimate, making payments and keeping records in accordance with the rules agreed by the House of Commons and the Independent Commission. The information will also be disclosed to the National Audit Office for such purposes. The information may also be used within the House of Commons Administration or by its agents for the purpose of business analysis or research. For the purposes of the Freedom of Information Act 2000 the House of Commons Administration is a Public Authority and therefore the information it holds will fall within the scope of that Act. Under the Data Protection Act 1998 you have the right to see and receive a copy of any personal data that the House of Commons Administration holds about you. If you have questions about the contents of this notice or how your information is handled or about your rights under the Data Protection Act 1998, please call our Data Protection Officer on 020 7219 2032, who acts on behalf of the Data Controller (the Clerk of the House).

Send your completed form to

Validation Team, Operations Directorate, Department of Finance & Administration, House of Commons, London SW1A 0AA

Office use only

Validation

Claims received

Signature check

Funds check

Allowable expenditure

Member Free ID & Costs

Expenditure Cat 5 & subtotals added to form

Receipts/documentation present

Processing

Input

Initials Date

Input subtotals per Cat 5

Comments

What kind of page is this?

Claim
An expense form

Proof
Receipt, invoice or purchase order

Blank
Nothing to see here

Other
Something we haven't thought of

Is this page interesting? Should we investigate further?

Not interesting
e.g. a coversheet or stationery

Interesting
It's significant expenses data

Interesting but known
e.g. it's a duckhouse

Investigate this!
I would like to know more

Go to next unreviewed page

Enter individual line items here

You need to indicate if this page contains **claims** or **proofs** before you can add line items.

Слика 4-12. Редактиран примерок на неиланираниите трошоци на Стивен Паунд (Stephen Pound) incidental expenses („Гардијан“)

менална. Во моментов изгледа дека тоа ќе биде поголема приказна и од истражувањето на криминалот во Велика Британија, од гледна точка на бројот на луѓето што постојано ѝ се навраќаат, и тоа е одлично.

Заедничко за двата проекти е тоа што се однесуваат на прашања за кои луѓето навистина се заинтересирани и подготвени се да им го посветат своето време и внимание. Во поголемиот број случаи, употребата на „краудсорсинг“ во нашата работа во голема мера зависи од оние што опсесивно се занимаваат со некое прашање. Во случајот на трошоците на членовите на Парламентот, во почетокот имавме огромни количини сообраќај за потоа интересот да спласне. А сепак, сè уште има луѓе што опсесивно продолжуваат да ги прегледуваат сите материјали, барајќи аномалии и приказни. Еден од нив обработи 30.000 страници материјал. Таквите луѓе имаат големи познавања.

„Краудсорсингот“ го користевме и во случајот со „документите на Сара Пејлин“. Тој повторно ни беше од огромна помош во барањето приказни во суровите информации.

Од аспект на генерирањето приказни, „краудсорсингот“ се покажа навистина корисен. На луѓето им се допадна таквиот пристап, а се покажа многу добар и за „Гардијан“. Од друга страна, не го користиме во толкава мера во генерирањето податоци.

Некои од поуспешните проекти во кои употребувавме „краудсорсинг“ всушност повеќе наликуваат на старите добри анкети. Резултатите секогаш се добри кога ќе ги прашате луѓето да ви раскажат за нивните искуства, нивните животи, што направиле во животот, затоа што веројатноста дека ќе измислуваат е многу помала. Ќе ви го кажат она што го чувствуваат. Кога барате од луѓето практично да ја завршат работата за вас, мора да пронајдете некаква рамка која ќе обезбеди веродостојност на податоците што ќе ги произведат.

Во поглед на веродостојноста на податоците, сметам дека пристапот на „Олд веџер“ (Old Weather, „Климата низ историјата“) е навистина добар. Тие обезбедија на секоја одредница да работат десет луѓе, и тоа е добар начин да се обезбеди точност на податоците. Додека работевме на трошоците на членовите на Парламентот, се обидовме да го минимизираме ризикот некој од парламентарците да појде на интернет и да ги поправи своите податоци. Од друга страна, не можете постојано да стоите на стража за да го спречите тоа. Она што можете е да следите неколку специфични URL-адреси или дали таквите обиди доаѓаат од зоната „Лондон Југозапад 1“ (SW1). Значи, не е толку едноставно. Податоците што ги објавувавме не беа секогаш веродостојни. Дури и кога се работеше за одлични стории, процесот не произведуваше такви сурови податоци што би можеле да ги користиме без сомневање во нивната веродостојност.

Ако треба да понудам некаков совет за амбициозните дата-новинари што сакаат да користат „краудсорсинг“ за собирање податоци, би ги советувајќи тој пристап да го применат на теми за кои граѓаните навистина се заинтересирани и за кои ќе останат заинтересирани дури и откако таа тема ќе исчезне од насловните страници на весниците. Исто така, сигурно ќе биде полесно да обезбедите вистински ангажман од луѓето ако тоа што го правите го претставите како игра. Кога по втор пат ја обработувавме темата за трошоците на членовите на парламентот, тоа многу повеќе личеше на игра, на натпревар во кој сите имаа своја индивидуална задача што требаше да ја извршат. Тој пристап на давање специфични задачи навистина помогна. Навистина е поинаку отколку на луѓето едноставно да им дадете купишта информации кои треба да се прегледаат и да им кажете „ова треба да се прегледа“, што целата работа може да ја направи тешка, напорна, без чувство на лична исполнетост. Затоа сметам дека навистина е важно сето тоа истовремено да биде и забавно.

— Маријане Бушар, „Блог за гаша-новинарство“, интервју со Сајмон Роџерс од „Гардијан“

Како „Даблогот“ го користеше „краудсорсингот“ во известувањето за продажбата на билети за Олимпијадата

Мислам дека од сите „краудсорсинг“-проекти, најдобар одзив имаше една статија за лотаријата за распределување билети за Олимписките игри. Илјадници луѓе во Велика Британија се обидоа да обезбедат билети за Олимпијадата во 2012 година и беа гневни кога не ги добија. Некои луѓе нарачаа билети во вредност од неколку стотици фунти за потоа да им биде кажано дека нема да добијат ништо. Она што никој не го знаеше е дали се работи само за неколку луѓе што протестираат многу гласно додека повеќето всушност се задоволни. Токму тоа сакавме да го дознаеме.

Решивме, во отсуство на какви било добри податоци на таа тема, дека најдобро ќе биде да ги прашаме луѓето што мислат. И сметавме дека ќе мора да му пристапиме на проблемот како да не се работи за премногу значајно прашање, затоа што немавме урамнотежен примерок на располагање.

Подготвивме еден формулар со помош на „Google“ во кој поставивме мошне прецизни прашања. Се работеше за прилично обемен формулар со прашања за тоа колкава била вкупната вредност на билетите што ги порачале, колку им било одбиено од платежните картички, за кои натпреварувања биле заинтересирани, такви прашања.

Го поставивме како мала илустрација со „линк“ на насловната страница од веб-страницата и тој набрзина се прошири наоколу. Сметам дека тоа е една од клучните работи; не можете да размислувате „што сакам да знам за потребите на мојата сторија?“, туку „што сакаат луѓето да ми кажат во овој момент?“ А „краудсорсингот“ може да успее само ако успеете да допрете до она за коешто луѓето сакаат да зборуваат. Количеството одговори што ги добивме за овој проект, еден од нашите први обиди на полето на „краудсорсингот“, беше огромно. Само во првиот ден по објавувањето на прашалникот добивме вкупно 7.000 одговори, илјада од нив само во првиот час.

Во тој момент сфативме дека на презентацијата на наодите и резултатите ќе треба да ѝ пристапиме мошне сериозно. Во почетокот немавме поим колку успешен ќе биде обидот. Затоа додадовме и неколку предупредувања и оградувања: свесни дека читателите на „Гардијан“ генерално заработуваат повеќе од просекот, дека луѓето што добиле помалку од она што го очекувале ќе бидат поподготвени да зборуваат со нас итн.

Не знаевме каква вредност ќе имаат резултатите. На крај излезе дека имаме одлични 7.000 записи на кои можеме да ја засноваме нашата статија и откривме дека околу една половина од луѓето што се пријавиле за билети не добиле ништо. Објавивме сè што имавме и, поради големиот број луѓе што се одзгаа на нашиот повик претходниот ден, имаше голем интерес во јавноста за резултатите.

How many Olympic tickets did you get? Here's our readers' results

We asked how you had fared in the London 2012 ticket ballot.
Here's our analysis of the information you gave us

Tweet 21

Share 84

reddit this

Comments (144)



The London 2012 Olympic stadium: will you be there? Will anyone you know?
Photograph: Handout/Getty Images

How much money did London 2012 ticket buyers have to put on the line to stand even a 50:50 chance of getting at least one ticket? If the results submitted by Guardian readers are to be believed, at least £1,000.

Earlier this week, we asked readers of the Guardian London 2012 blog to let us know how their ticket-purchasing attempts had fared. By the end of the day – when this analysis was carried out – we'd had more than 5,000 responses.



Posted by
James Ball
Friday 3 June 2011 11.04
BST
guardian.co.uk
Article history



A larger | smaller

Sport
Olympic tickets - Olympic Games 2012

More from London 2012
Olympics blog on

Sport
Olympic tickets - Olympic Games 2012

More blogposts

Слика 4-13. Колку билети за Олимпискиите игри добивме?: резултати од анкетата со граѓанине („Гардијан“)

Неколку недели подоцна, кога беше објавен и официјалниот сумиран извештај за продажбата на билетите, се шокиравме кога сфативме колку блиски биле бројките што ние ги понудивме. Ги бевме промашиле точните бројки за влакно. Тоа делумно може да се должи на среќата, но дел сигурно се должи на тоа што добивме толку добар одзив на нашиот повик.

Ако побарате од читателите да направат нешто слично во корисничките коментари, ќе се соочите со ограничените можности за тоа што можете да сторите со резултатите. Значи, пред да почнете, треба добро да размислите: „Која е најдобрата алатка за да го дознаам она што сакам?“ Можеби коментарите се најдоброто решение? Можеби ќе биде подобро да подготвиме специјална апликација? Ако решите да изработите апликација, тогаш треба да размислите: „Дали ќе вреди да се чека толку време? Дали ќе ги оправда ресурсите што треба да ги потрошиме за изработка на апликацијата?“

Во нашиот пример, ние се сетивме на „Google Forms“ (сервис на „Google“ за изработка на формулари). Секогаш кога некој ќе го пополни формуларот, ре-

зултатите се појавуваат како нов ред во една сметководствена табела. Тоа значеше дека можев да ја отворам сметководствената табела и да ги прегледам сите резултати како што пристигнуваат и се ажурираат.

Иако можев да пробам да ја завршам сета работа во Google, ја снимив табелата во Microsoft Excel и извршив неколку операции, ги подредив бројките од најмала кон најголема, ги пронајдов внесовите каде што луѓето ги напишале сумите што ги потрошиле (наместо поединечните цени), ги поправив сите тие работи. Решив да задржам сè што ќе можам. Така, наместо да ги разгледувам само валидните одговори, се обидов да го поправам она што го имав на располагање. Некои од луѓето што го пополнија формуларот користеа странски валути, па ги пресметав износите во фунти, а сето тоа беше донекаде напорно.

Сепак, целата анализа беше завршена за неколку часа и, се разбира, ги отфрлив оние внесови што очигледно беа смешни, кога некој се обидел да се пошегува со нас. На пример, многумина решиле да посочат дека не потрошиле ништо за да набават билети. Тоа е вртење на работата на зафрканција, ама, што да се прави. Такви беа помалку од сто од вкупно повеќе од седум илјади внесови.

Потоа имаше неколку десетици одговори со невозможно високи износи во обид да ги извртат резултатите. На пример, дека потрошиле десет милиони фунти. На крај, пред себе имав група податоци на која можев да ги применим нормалните принципи за работа со податоци што ги користиме секојдневно. Подготвив таканаречена „приспособлива табела“ (pivot table). Извадив неколку просеци и слични пресметки. Не ни помислувавме, на почетокот, дека проектот ќе добие толкав замав, па единствена помош ми беше уредникот на спортскиот блог. Седнавме и сфативме дека проектот може да биде и многу забавен. Сета работа, од почеток до крај, ја завршивме за 24 часа. Добивме таква идеја, до ручек успеавме да склопиме нешто и да го поставиме на насловната страница на веб-страницата, се покажа дека е популарно, го задржавме на насловната до крајот на денот, а податоците ги поставивме на интернет следното утро.

Решивме да работиме во „Google Docs“ затоа што ни обезбедуваше целосна контрола над резултатите. Немаше потреба да користиме други аналитички алатки. Префрлувањето на табели од Google Docs во друг софтвер за работа со бази на податоци или во сметководствени табели е лесно. Ако почнете да користите специјализиран софтвер за анкети, производителите на таквиот софтвер често ве условуваат да морате да ги користите само нивните алатки. Да работевме со некакви особено чувствителни информации, можеби ќе размислевме дали да ги користиме „Google“ или да подготвиме нешто наше, „домашно“. Но, општо земено, поставувањето формулар од „Google Form“ на страницата на „Гардијан“ е мошне лесно, а нема шанса корисниците да откријат што сме користеле. Значи, ептен ни се најде.

Ако треба да давам совети за дата-новинарите што сакаат да користат „краудсорсинг“, тогаш нека го користат кога имаат мошне специфични барања.

Нека постават прашања што имаат повеќе од еден одговор. Нека се обидат да обезбедат барем основни демографски податоци од лицата со кои зборуваат за да можат да видат дали примерокот со кој работат ќе претежнува на некоја страна. Ака бараат износи и слични нешта, како во нашиот случај, нека даваат прецизни упатства за пополнување, дека се бараат цифри, дека треба да користат определена валута и слично. Многумина нема да ги почитуваат упатствата, но што повеќе ги водите за рака низ процесот, толку подобри ќе бидат наодите. Апсолутно секогаш нека додадат можност за оставање коментари. Иако многумина ќе ги пополнат и другите полиња, она што навистина сакаат да ви го кажат е нивното мислење за сторијата. Особено ако се работи за приказна или скандал што се однесува на измама на потрошувачите.

— Маријане Бушар, „Блог за даџа-новинарство“, интервју со Џејмс Бол од „Гардијан“

Користење и размена на податоци: Што велат прописите, што е напишано со мали букви и каква е реалноста

Во следните неколку параграфи ви нудиме еден краток преглед на законските прописи што ја регулираат областа на податоците и базите на податоци и како можете да ги „отворите“ своите податоци за јавноста со користење достапни јавни лиценци и правни алатки. Сепак, не дозволувајте ништо од она што е напишано подолу да ви го намали ентузијазмот да се занимавате со новинарство водено од податоци. Законските ограничувања за користење на податоците обично не се пречка, па лесно можете да обезбедите дека нема да се испречат на патот на други лица што сакаат да ги користат податоците што вие сте ги објавиле.

Со ризик да повториме нешто што е очигледно, денес е полесно да се обезбедат потребните податоци од кога било порано. Порано, пред објавувањето податоци на интернет да стане раширена практика, дури и кога знаевте кои податоци ви се потребни, моравте да ги молите оние што ги поседуваа да ви ги направат достапни, процедура што вклучуваше многу бирократија, писмена кореспонденција или посета на институцијата што ги има податоците. Денес, вашиот компјутер ќе побара од нивниот компјутер да му испрати еден примерок од податоците. Иако постои концептуална сличност, факт е дека сега, во овој момент, имате примерок од базата на податоци, а тие (создавачот или издавачот) не направиле ништо на тоа поле и веројатно воопшто немаат поим дека сте снимиле примерок од саканите податоци.

Што со прашањето на спуштање податоци со помош на компјутерска програма (се нарекува и „стругање“) и со „Условите за користење“ (Terms of Service

– ToS)? Да се вратиме на претходниот параграф: вашиот прегледувач е токму таква компјутерска програма.

Дали во Условите на користење може да се дозволи пристап само со определен вид компјутерска програма? Никој не ве спречува да ги читате таквите документи и да побарате совет од адвокат, секако, ако имате време и пари за фрлање. Сепак, не бидете „дрнкач“: Ако програмата што ја користите „удри“ по некоја интернет-страница, може да се соочите со блокада за пристап на вашата мрежа до таа страница – и да сте го заслужиле тоа. Денес постои обемна правна пракса што се однесува на пристапот и „стругањето“ податоци од интернет. Ако планирате да се занимавате со „стругање“, примери што ќе ви послужат како основа во работата може да пронајдете на веб-страниците како што е „ScrapingWiki“.

Кога еднаш ќе ги добиете податоците што ве интересираат, можете да ги прегледувате, распоредувате, визуелизирате, корелирате или да вршите друг вид анализа во вашиот примерок од податоците. Можете да објавите ваша анализа во која ќе ги цитирате и наведувате оние податоци што сте ги одбрале. Има нешто во паролата „фактите се слободни“ (во иста смисла како и слободата на говорот), но таа парола главно ја користат оние што постојано размислуваат за правните аспекти поврзани со базите на податоци или, во една поширока (и извртена) смисла, за управувањето со податоците.

Што ако, како амбициозен дата-новинар, планирате да ја објавите анализата, вклучувајќи и определени факти или податоци, но истовремено да ги објавите и базите или групите податоци што сте ги користеле во вашата анализа, можеби и како прилог на анализата? А можеби само ги чувате и се грижите за податоците, без нивна анализа (тоа е добро: на светов му се потребни повеќе луѓе што се грижат за чувањето податоци). Ако користите податоци собрани од некој друг субјект, тогаш може да настане проблем. (Дури и ако сами сте ги собрале податоците во базата, сепак прочитајте го следниот параграф како мотивација за практиките за споделување што следат потоа.)

Ако сте запознаени со ограничувањата за креативната работа што со себе ги носат авторските права – дали можеби носителот на авторските права не дал дозвола за користење на делото (дали делото е дел од јавниот домен или дали на начинот на кој вие го користите делото можат да се применат исклучоци од заштитата што ја нудат авторските права, како што е концептот за „фер употреба“), а вие сепак го користите – дистрибуирате, изведувате итн. – делото, носителот на авторските права може да ве натера да престанете со употребата. Иако фактите не подлежат на авторски права, збирките факти можат да бидат ограничени на сличен начин, иако во релевантното законодавство постојат повеќе варијации споредено со правилата за примена на авторските права кај креативните дела. Да скратиме, една база на податоци, како креативно дело, може да биде предмет на заштита според законите за авторски права. Во многу правни системи,

самиот труд вложен во собирањето податоци во една база на податоци, колку тој труд и да не бил креативен, значи дека базата на податоци подлежи на заштита на авторските права. Тоа особено е случај во САД, каде минималното количество креативност потребно за примена на авторските права е поставено многу ниско (разгледајте го судскиот процес на Феист против Рурал (Feist vs. Rural) за авторските права над еден телефонски именик, како класичен случај од САД). Во некои правосудства постојат и „права поврзани со сопственоста на базите на податоци“ што поставуваат ограничувања за користење на базите на податоци, и тие се одвоени од пошироката тема на авторските права (иако има многу преклопување од гледна точка на прашањата што ги покриваат, особено во случаите кога речиси и да не постојат критериуми за креативноста што би се применувале). Најпознати меѓу нив се *sui generis* правата над базите на податоци на Европската Унија. Повторно, а особено ако работите во Европа, можеби ќе сакате да побарате дозвола пред да објавите база на податоци што му припаѓа на некој друг.

Се разбира, таквите ограничувања не помагаат многу во воспоставувањето амбиент погоден за развојот на новинарството водено од податоци (не чинат ни за општеството генерално – социолозите и други научници ја предупредија ЕУ за тоа уште пред *sui generis* правилата да стапат во сила, а подоцнежните истражувања покажуваат дека имале право). Добра работа е што, како издавач на таква база на податоци, можете да ги остраниите таквите ограничувања од вашата база на податоци (под претпоставка дека таа не содржи елементи за кои немате овластување да одобрите натамошно ширење), обично со тоа што сите дозволи ќе ги дадете однапред. Една можност е да ја објавите базата на податоци со јавна лиценца или да ја прогласите за дел од јавното добро – слично на програмерите што го објавуваат својот код под слободна лиценца со отворен изворен код (бидејќи новинарството водено од податоци често содржи и програмерски код, а не само податоци, се разбира дека треба и кодот да биде јавно достапен за да може, по потреба, и други да го репродуцираат процесот на собирање податоци, како и анализата). Постојат повеќе причини зошто треба да ги направите податоците јавно достапни. На пример, вашата публика може да ги искористи да изработи нови визуелизации или апликации до кои можете да поставите хиперврска – нешто што „Гардијан“ го направи со неговиот „пул“ (pool, здружување на група луѓе или организации што работат во иста област или на иста тема, со цел да се здружат нивните ресурси и да се поттикне размената на информации, заб.прев.) за визуелизација на податоци на „Фликр“ (Flickr, <http://www.flickr.com>). Вашите податоци можат да се комбинираат со други податоци што ќе ви дадат вам и на вашите читатели подобар увид во некоја тема. Тоа што другите ќе го направат со вашите податоци може да ви даде идеја за нови стории или нови проекти водени од податоци. А секако ќе добиете и јавно признание.



Слика 4-14. Налейници за промоција на „Отворениите податоци“ (Фондација „Отворено знаење“)

Кога ќе сфатиме дека објавувањето на делата под јавна лиценца навистина е потребно, се јавува прашањето, која лиценца да ја употребиме? Одговорот на тоа прашање често ќе го пронајдете во проектот или во заедницата чиј труд го надградувате, или се надевате дека вашиот труд ќе придонесе кон нивните цели – користете ја лиценцата што тие ја користат. Ако сакате да „копате“ подлабоко, почнете од групата слободни и отворени лиценци – тоа значи дека сите имаат дозвола за сите видови користење (можно е да се бара именување на изворот или споделување на податоците под исти услови). Тоа што во областа на софтверот се „Дефиницијата за слободен софтвер“ (Free Software Definition) и „Дефиницијата за отворен изворен код“ (Open Source Definition), за сето друго знаење, вклучувајќи ги и базите на податоци, е „Дефиницијата за отворено знаење“ (Open Knowledge Definition). Таа определува што го прави делото отворено и што им дозволуваат отворените лиценци на корисниците.

За да видите кои се квалификуваните лиценци што важат во овој момент, посетете ја интернет-страницата на Фондацијата „Отворено знаење“ посветена на „Дефиницијата за отворено знаење“. Во основа постојат три класи отворени лиценци:

Донации во јавното добро

Се работи за максимално пермисивни лиценци; не се поставуваат никакви услови за употреба на делото.

Пермисивни лиценци, лиценци каде што се бара само правилно именување на изворот или авторот

Давањето признание на заслугите на авторот е единствен значаен услов што го поставуваат таквите лиценци.

„Копилефти“ (Copyleft), реципрочни или лиценци за споделување под исти услови

Овие лиценци бараат модифицираните дела, доколку се објавуваат, да бидат понудени за натамошна употреба под истата лиценца. (Помот „копилефт“ е антоним на англискиот термин „copyright“, односно означува дела кај кои авторските права не се заштитени и се дел од јавното добро, заб.прев.)

Ако користите податоци што некој друг ги објавил под отворена лиценца, горниот параграф ќе послужи како краток водич како да ги исполните условите наведени во отворената лиценца. Најверојатно ќе најдете на некоја од лиценците издадени од „Криејтив комонс“ (Creative Commons), „Отворена дата комонс“ (Open Data Commons) или од различни влади, и тие вообичаено содржат резиме што овозможува веднаш да видите кои услови треба да бидат исполнети за користење на податоците. Вообичаено е лиценцата да е посочена на веб-страницата од која ги спуштате (или ги „стругате“) податоците или на видно место во самата група податоци, во зависност од форматот. И вие би требало да ја наведете лиценцата на видно место кога ги отворате своите податоци за јавен пристап.

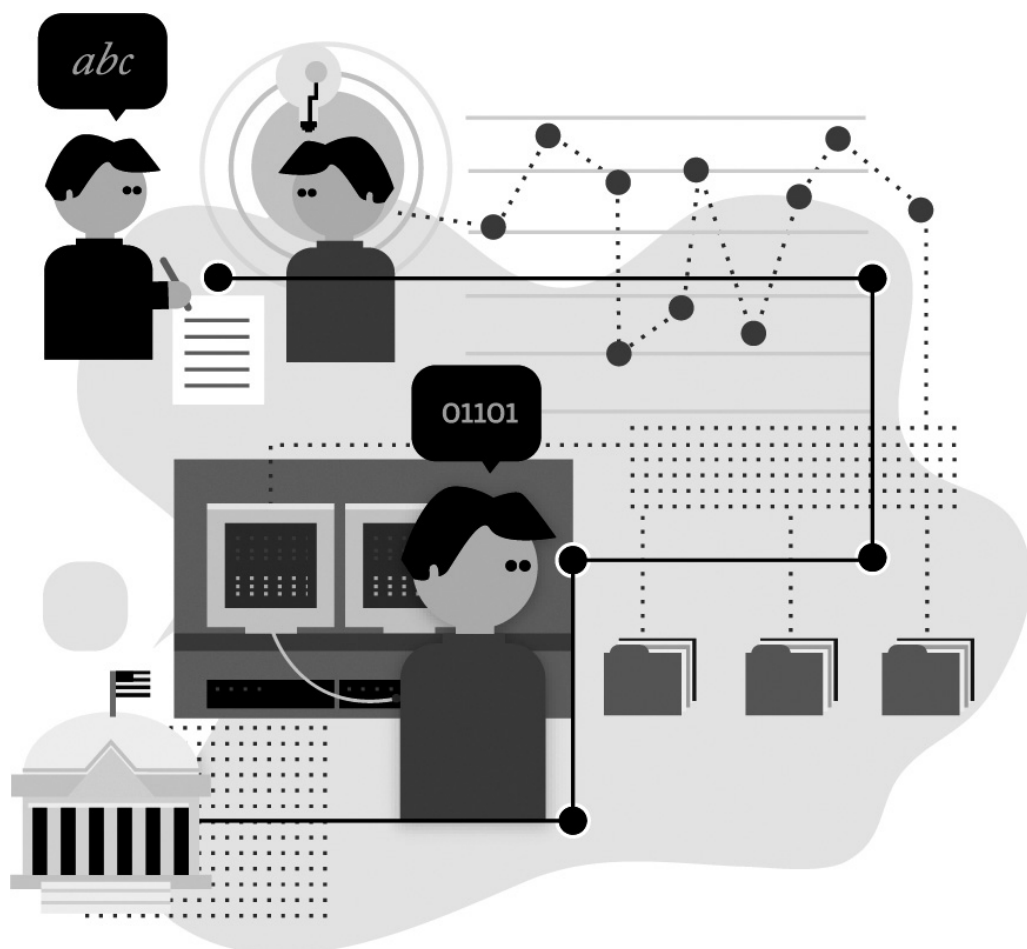
Да се вратиме на почетокот – што ако податоците што ви се потребни сè уште не се достапни на интернет, или се заштитени со некаква форма на контрола на пристапот? Размислете, кога барате пристап за себе, истовремено да побарате податоците да бидат отворени за пристап на сите што сакаат да ги користат. Не ќе е лошо да посочите и неколку позитивни примери за тоа што би можело да се случи со нивните податоци доколку се решат да ги отворат.

Споделувањето со цел свет, од друга страна, може да ги внесе во игра и правилата од областа на приватноста или некои други прашања. Вистина е дека тоа што отворените податоци елиминираат многу од техничките бариери или бариери поврзани со авторските права, не значи дека не треба да се придржуваме до другите закони и правила што би можеле да се применат. Но, тоа отсекогаш било така и на новинарите им стојат на располагање огромни ресурси, а понекогаш и заштитни инструменти, доколку здравиот разум ви вели дека би требало да истражите некое од тие прашања.

Ние ви посакуваме среќа во тој потфат! Иако, веројатно, среќата ќе ви биде многу попотребна во други сегменти на вашиот проект отколку во справувањето со (и инаку малите) законски ризици.

– Мајк Линксвеер, „Криејтив комонс“ (Mike Linksvayer, Creative Commons)

Разбирање на податоците



Кога ќе ги добиете податоците, што со нив? Што треба да барате? Кои алатки да ги користите? Ова поглавје го отвораме со неколку идеи за подобрување на вашата писменост на полето на работа со податоци, совети за начините како да постапувате со броевите и статистиките, како и неколку работи што треба да ги имате предвид кога работите со збркани, несовршени или, што често се случува, со групи податоци за кои недостига целосната документација. Потоа ќе видиме како од податоците да добиеме приказни, кои се омилените алатки на дата-новинарите и како се користи визуелизацијата на податоците за обезбедување подобар увид во темата што ја обработувате.

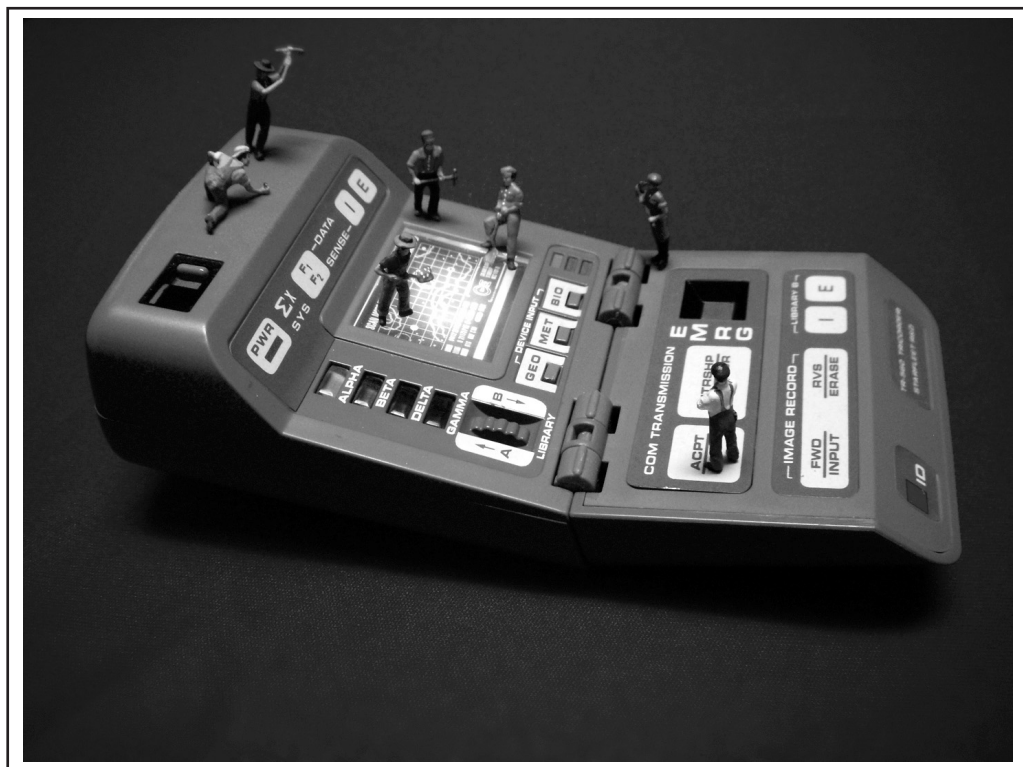
Подобреете ги познавањата на полето на работа со податоци во три лесни чекори

Поимот писменост се однесува на „способноста на читање заради стекнување знаења, кохерентно пишување и критичко размислување за прочитаните печатени материјали“. Следствено, писменоста во областа на користењето податоци е способност за употреба на податоците за стекнување нови знаења, производство на кохерентни податоци и критичко размислување за податоците. Писменоста во користењето податоци ја вклучува статистичката писменост, но и разбирањето на начинот на кој се работи со големи групи податоци, начинот на нивното производство, поврзувањето различни групи податоци и нивното толкување.

„Универзитетот за вести“ на институтот „Поинтер“ (Poynter’s News University) нуди курсеви по математика за новинари на кои новинарите можат да добијат помош во совладувањето концепти какви што се промените на процентите и просечните вредности. Она што е интересно, истите концепти во исто време се учат во непосредна близина на канцелариите на институтот „Поинтер“ (Poynter), во основните училишта на сојузната држава Флорида, каде што се дел од наставната програма за петто одделение (деца на возраст од 10-11 години).

Фактот што новинарите имаат проблем со математички операции што се учат уште во основно училиште, совршено покажува колку неписмени станале редакциите во областа на работа со податоци. А тоа навистина е голем проблем. Како некој дата-новинар да искористи еден куп бројки за климатските промени, ако не знае што е интервал на веродостојност? Како некој дата-новинар да напише приказна за тоа како се распределува националниот доход, ако не знае што е средна, а што просечна вредност?

Од друга страна, новинарот не мора да е дипломиран статистичар за да ги користи податоците на поефикасен начин. Неколку едноставни трикови можат да помогнат од бројките да произлезе многу подобра приказна. Како што вели професорот Герд Гигеренцер (Gerd Gigerenzer) од Институтот „Макс



Слика 5-1. Копање по податоциите (фото: Џ. Д. Хенкок - JDHancock, <http://www.flickr.com/photos/jdhancock/3386035827/>)

Планк“ (Max Planck Institute), ако не ги користиме правилно, тогаш подобрите алатки не мора да значат и подобар новинарски производ.

Дури и ако немате поим од математика или статистика, лесно можете да станете искусен дата-новинар со поставување три едноставни прашања.

1. Како се собрани податоците?

Неверојатен раст на БНД

Најлесниот начин да се пофалите со спектакуларни податоци е да ги измислите. Иако не изгледа така, податоците за кои често се зборува, на пример, податоците за бруто националниот доход (БНД), може да бидат и лажни. Крег Мареј (Craig Murray), поранешен амбасадор на Велика Британија, во својата

книга „Убиство во Самарканд“ (Murder in Samarkand), пишува дека стапките на економски раст во Узбекистан постојано се предмет на напнати преговори помеѓу Владата на Узбекистан и меѓународните институции. Со други зборови, таа стапка нема никаква врска со локалната економија.

БНД се користи како носечки показател затоа што на владите им е потребен за следење на главниот извор на приходи со кој располагаат – данокот на додадена вредност (ДДВ). Ако некоја влада не се финансира од собраниот ДДВ, или ако не го објавува буџетот во јавност, нема причина да ги собира податоците за движењето на БНД и тогаш за неа е подобро да ги фабрикува тие податоци.

Стапката на криминал постојано расте

„Стапката на криминал во Шпанија порасна 3 отсто“, пишува дневниот весник „Ел паис“ (El Pais). РТЛ (RTL) тврди дека Брисел станал жртва на растечката стапка на криминал поради нелегалните доселеници и зависниците од дрога. Таквото известување, засновано на статистиките што ги собира полицијата, е мошне често, но не ни кажува ништо за насилниот криминал.

Можеме да бидеме сосема сигурни дека никој не ги „штелува“ податоците во Европската Унија. Полицајците, од друга страна, сакаат добра мотивација. На пример, ако успехот се мери преку стапката на решени случаи, полицајците се мотивирани да пријават што е можно повеќе инциденти кај кои не е потребна дополнителна истрага. Пушењето „тревка“ е токму такво кривично дело. Тоа објаснува зошто стапката на криминал поврзан со трговијата и употребата на наркотици во Франција порасна четири пати во последните 15 години, иако потрошувачката на наркотици остана стабилна.

Што можете вие да направите

Кога се сомневате во веродостојноста на некој број, секогаш проверете го два пати, исто како што би направиле со изјавата на некој политичар. Во узбекистанскиот пример што веќе го спомнавме, доволен би бил телефонски разговор со некој што живее таму подолго време („Дали сметате дека државата денес е три пати побогата од 1995 година, како што тврдат официјалните податоци?“).

Што се однесува до податоците што ги нуди полицијата, социолозите често спроведуваат студии за виктимизацијата во кои луѓето ги прашуваат дали се чувствуваат загрозувани од криминалот. Таквите студии покажуваат многу помалку експлозивни податоци од она што го кажува полицијата. Можеби таму се крие причината зошто таквите вести ретко се појавуваат на насловните страници.

Постојат и други можности за прецизна процена на веродостојноста на податоците, како што е „Бенфордовиот закон“ [Benford's Law, именуван по физичарот Френк Бенфорд (Frank Benford), заб.прев], но ниту една од нив не е замена за критичкото мислење.

2. Што ново може да се научи?

Луѓето што работат ноќе се соочуваат со двапати повисок ризик од заболување од мултиплекс склероза

Сите нормални Германци и Германки ќе престанат да работат во ноќна смена ако прочитаат таков наслов. Но, статијата под тој наслов не открива целосно во што се состои ризикот.

Да земеме како пример група од 1.000 Германци и Германки. Веројатноста е дека еден или една од нив ќе добие мултиплекс склероза во животот. Ако сите 1.000 работеа во ноќна смена, бројот на болни од мултиплекс склероза ќе порасне на 2. Дополнителниот ризик од заболување од мултиплекс склероза кај оние што работат во ноќна смена е 1 во 1.000, не е 100%. Оваа информација секако е покорисна кога размислуваме дали да прифатиме некоја работа.

Во просек, секој петнаесетти Европеец е целосно неписмен

Тоа навистина е застрашувачки наслов. Исто така, насловот е целосно вистинит. Од 500 милиони Европејци, околу 36 милиони не знаат да читаат. Од друга страна, 36 милиони Европејци се деца под седум години (според податоците на Евростат, достапни на <http://bit.ly/eurostat-numeracy>).

Кога вадите просеци, секогаш размислувајте на линија „просек од што?“. Дали референтната популација е хомогена? Нерамномерните обрасци на распределба објаснуваат, на пример, зошто повеќето луѓе се подобри возачи од просекот. Мнозинството луѓе во животот учествуваат во само една, или воопшто не учествуваат во сообраќајни незгоди. Неколкумина непромислени возачи предизвикуваат многу повеќе сообраќајни незгоди и го креваат нивниот просечен број далеку над она што мнозинството луѓе го искусуваат. Истото правило важи и во случајот на распределбата на приходите: повеќето луѓе заработуваат помалку од просечната плата.

Што можете вие да направите?

Секогаш имајте ги предвид стапката на распределба и основната стапка. Проверете кои се средната и просечната вредност, како и најчестата вредност, што ќе ви помогне полесно да ги разберете податоците. Свеста за редот на големини со кои работите, како во случајот со мултиплекс склерозата, помага во дефинирањето на контекстот. Најпосле, читателите полесно ќе ве разберат ако известувате со користење изрази што се одраз на природната зачестеност (еден од 100) наместо во проценти (1%).

3. Колку веродостојни се информациите?

Проблеми поврзани со големината на примерокот

Едно истражување на јавното мислење за кое известува весникот „Диаро де Навара“ (Diario de Navarra) од Сарагоса тврди дека „80 отсто од граѓаните се незадоволни од правосудниот систем“. Може ли ставот на 800 испитаници да се екстраполира на 46 милиони Шпанци? Сигурно работата е „напумпана“.

Кога истражувањето што го спроведувате се однесува на голема популација (барем од неколку илјади), најчесто ви требаат барем 1.000 испитаници за да добиете резултати со можна статистичка грешка под 3 отсто. Тоа значи дека ако го повторите испитувањето на сосема друг примерок, во 19 од 20 случаи одговорите што ќе ги добиете ќе влезат во интервалот од плус/минус 3 процентни поени од вредноста што би ја добиле ако решите да ги прашате сите лица во една популација.

Пиењето чај го намалува ризикот од срцев удар

Статиите за придобивките од пиењето чај се многу чести. Кратката статија објавена во „Ди велт“ (Die Welt) што тврди дека чајот ја намалува опасноста од инфаркт на миокардот, го потврдува тоа правило. Иако ефектите на пиењето чај се предмет на сериозен научен интерес, многу истражувања не ги земаат предвид факторите што со себе ги носи стилот на живот, како што се исхраната, професијата или занимавањето со спорт и рекреација.

Во повеќето држави чајот се смета за пијалак за повисоките општествени класи, кои се грижат за сопственото здравје. Резултатите на истражувачите што во своите студии за ефектите на чајот не ги земаат предвид животните навики на испитаниците, можат да се сведат на тврдењето „богатите луѓе се поздрави и најверојатно пијат чај“.

Што можете вие да направите?

Пресметките на корелациите и маргините на грешка кај студиите за ефектите на чајот сигурно се точни, барем во најголемиот број случаи. Но ако истражувачите ги запостават корелациите (на пример, врската помеѓу пиењето чај и занимавањето со спорт), резултатите се од мала вредност.

Нема смисла новинарите да ги оспоруваат нумеричките резултати од некоја студија, на пример, големината на примерокот на кој е вршено истражувањето, освен во случаите кога постои основано сомневање поврзано со големината на примерокот. Од друга страна, лесно се проверува дали истражувачите пропуштиле во своето истражување да вклучат некои релевантни информации.

– Николас Кајзер-Брил, *Journalism++*

Совети за користење бројки во вестите

- Најдобриот совет за работење со податоци што можеме да ви го дадеме е да се препуштите и да уживате. Податоците можат да изгледаат одбивно. Ако им дозволите да ве заплашат, нема да стигнете никаде. Ако им пристапите како на нешто со што можете да си играте и да го истражувате, ќе бидете изненадени од леснотијата со која ви ги предаваат своите тајни и приказни. Значи, ракувајте со нив како и со сите други доказни материјали, без страв или стравопочит. Гледајте на нив како на вежба за вашата способност за имажинација. Бидете креативни и размислувајте за алтернативните приказни што се совпаѓаат со податоците и поблиску ги објаснуваат, а потоа проверете ги низ споредба со повеќе докази. „Дали ова може да се објасни со некоја друга приказна?“ е корисно прашање во размислувањето дали еден очигледно преголем или погрешен број, што би требало да послужи како доказ за нешто, можеби не е таков каков што изгледа.
- Не мешајте ја скепсата околу податоците со цинизмот. Скепсата е добра; цинизмот, од друга страна, кренал раце и се откажал. Ако верувате во вредноста на дата-новинарството (веројатно верувате, инаку не би ја читале книгава), тогаш мора да верувате дека податоците нудат нешто многу подобро од лагите, карикираниите наслови или „убиствените“ факти во „шашливите“ наслови. Податоците, ако ги користиме внимателно, често нудат длабоки сознанија. Треба да бидеме на штрек, а не цинични или наивни.
- Ако ви кажам дека употребата на алкохол пораснала за време на рецесијата, вие можеби ќе одговорите дека тоа е затоа што сите се депресивни. Ако ви кажам дека употребата на алкохол се намалила, можеби ќе одговорите дека тоа е затоа што сите се „шворц“. Со други зборови, она што податоците го велат не прави разлика за интерпретацијата што сакате да ја дадете, имено, дека работите се „за никаде“ како и да ги гледаме. Ако расте, не чини; ако паѓа, пак не чини. Поентата е дека ако верувате во податоците, дозволете им да кажат што имаат пред на нив да ги налепите сопствените расположенија, уверувања или очекувања. Има толку многу податоци што често ќе можете да ги потврдите вашите уверувања со просто разгледување наоколу. Со други зборови, и според моето скромно мислење, дата-новинарството не може да внесе додадена вредност ако не му пријдете отворено и без предубедувања. Дата-новинарството ќе биде онолку објективно колку што вие ќе му дозволите да биде. Тоа што е засновано на бројки не е доволно само по себе.
- Несигурноста е во ред. Ние ги поврзуваме броевите со авторитет и сигурност. Многу често, одговорот се сведува на тоа дека нема одговор, или дека одговорот што го имаме можеби е најдобриот можен, иако можеби е далеку од тоа да биде точен. Сметам дека за некои работи треба да зборуваме отворено. Ако вам ви звучи како добар начин да се „убие“ приказната, јас велам дека тоа е одличен начин да се постават нови прашања. Понекогаш постојат повеќе легитимни начини на гледање на податоците. Броевите, сами за себе, не мора да бидат ни вистинити ни лажни.

- Истражувањето само по себе е приказна. Приказната за тоа како сте се обиделе да откриете нешто може и самата да биде пример за одлично новинарство. Преминувањето од еден доказ на друг – а тоа особено се однесува на доказите што произлегуваат од податоци и каде еден единствен број ретко е доволен за такво нешто. Разновидните извори обезбедуваат нови агли на гледање, нови идеи, подобро разбирање. Се прашувам дали сме толку запнати да бидеме авторитативни и да им ги дадеме одговорите на луѓето што по пат забораваме да го прикажеме и процесот на душќање наоколу.
- Вистинските прашања постојат од поодамна: дали тоа навистина е голем број? Од каде произлегува? Дали сте сигурни дека значи тоа што мислите дека значи? Општо земено, тие прашања се само добра насока како да размислувате за податоците, за нештата по рабовите што претекле при гледањето во една бројка, компликациите што ги носи реалниот свет, широкиот дијапазон можни споредби во различни времиња, групи или географски области; накратко, да размислите за контекстот.

– Мајкл Бласџленг, слободен новинар

Основни чекори во работата со податоци

Кога почнувате некој проект што се заснова на податоци, треба да разберете барем три клучни концепти:

- Барањата за пристап до податоци треба да почнат со список на прашања на кои ви е потребен одговор.
- Податоците често се заплеткани и нечисти и им треба прочистување.
- Податоците можат да имаат својства што не се документирани.

2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 113 Payment for goods and services, 113.03 Office supplies, utensils and household goods, Base component, Non-personnel recurrent, 135100,
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 113 Payment for goods and services, 113.06 Assets and periodically, Base component, Non-personnel recurrent, 20000, 44000, 24372.18
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 113 Payment for goods and services, 113.09 Nutrition/food, Base component, Non-personnel recurrent, 0, 4300, 4455.3
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 113 Payment for goods and services, 113.10 Drugs and consumables, Base component, Non-personnel recurrent, 0, 1000, 1000
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 113 Payment for goods and services, 113.11 Telecommunication and mail services, Base component, Non-personnel recurrent, 384500, 384500, 35355
0, 87	
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 113 Payment for goods and services, 113.13 Transportation services, Special means, Non-personnel recurrent, 200000, 330400, 0, 35
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 113 Payment for goods and services, 113.18 Current repair of equipment and inventory, Base component, Non-personnel recurrent, 55000, 129300,
2412	
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 113 Payment for goods and services, 113.20 Use of government and local insignia/symbols, state distinction signs, Base component, Non-per
sonnel recurrent, 0, 31000, 0	
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 113 Payment for goods and services, 113.22 Printing services, Base component, Non-personnel recurrent, 111000, 200000, 240115.8
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 113 Payment for goods and services, 113.23 Representation costs, Base component, Non-personnel recurrent, 324000, 407000, 336312.74
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 113 Payment for goods and services, 113.30 IT and PC works, Base component, Non-personnel recurrent, 200700, 120700, 200105.81
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 114 Duty travels, 114.02 Travels abroad, Base component, Non-personnel recurrent, 196100, 187000, 67783.21
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 115 Employer's share of the mandatory health insurance premium, 115.01 Employer's share of the mandatory health insurance premium paid by
le in the country, Base component, Personnel, 88800, 88800, 8773	
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 115 Transfers to population, 115.01 Lifetime pensions and indemnities, Base component, Non-personnel recurrent, 221300, 221300, 221256
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 115 Transfers to population, 115.02 Other transfers made to population, Base component, Non-personnel recurrent, 184000, 184000, 183964.8
2005, Central, 102	Republic of Moldova's Presidency Office (apparatus), 0102 Republic of Moldova's Presidency Office (apparatus), 0103 Central apparatus (Office) of ministries and other administrative author
1115, 01 General	purpose state services, 01.02 Executive authorities, 242 Purchase of fixed assets, 242.00 Purchase of fixed assets, Base component, Capital, 112000, 144700, 113494.5
2005, Central, 103	Court of Accounts, 0103 Court of Moldova, 0103 Central apparatus (Office) of ministries and other administrative authorities, 01 General purpose state services, 01.03 Financial, budget/accou
ntion and control services, 111 Remuneration of work, 111.00 Remuneration of work, Special means, Personnel, 0, 94500, 64563.15	
2005, Central, 103	Court of Accounts, 0103 Court of Moldova, 0103 Central apparatus (Office) of ministries and other administrative authorities, 01 General purpose state services, 01.03 Financial, budget/accou
ntion and control services, 111 Remuneration of work, 111.00 Remuneration of work, Base component, Personnel, 630100, 741500, 723103.57	
2005, Central, 103	Court of Accounts, 0103 Court of Moldova, 0103 Central apparatus (Office) of ministries and other administrative authorities, 01 General purpose state services, 01.03 Financial, budget/accou
ntion and control services, 112 Mandatory state social insurance premium, 112.00 Mandatory state social insurance premium, Special means, Personnel, 0, 25200, 1742.17	
2005, Central, 103	Court of Accounts, 0103 Court of Moldova, 0103 Central apparatus (Office) of ministries and other administrative authorities, 01 General purpose state services, 01.03 Financial, budget/accou

Слика 5-2. Збркани њодајтоци

Знајте кои прашања сакате да ги одговорите

Во многу аспекти, работата со податоци наликува на интервјуирањето жив извор. И на податоците им поставувате прашања и чекате да ви одговорат. Но, исто како што изворот може да ги одговори оние прашања за кои има информации, и една група податоци може да одговори само на оние прашања за кои ги има вистинските записи и потребните променливи. Тоа значи дека треба внимателно да размислите кои прашања сакате да ги одговорите дури и пред да ги соберете податоците. Во основа, работете наназад. Најпрво направете список на тврдењата во вашата приказна што сакате да ги поткрепите со податоци. Потоа одлучете кои променливи и кои записи ќе треба да ги обезбедите и анализирајте за да излезете со тие ставови.

Да погледнеме еден пример што се однесува на извештаите за стапката на криминал на локално ниво. Да речеме дека сакате да пишувате статија за обрасците на криминалот во вашиот град, и дека сакате да се осврнете на временскиот период во денот и на деновите во неделата кога постои најголема веројатност за настанување различни видови криминални дејства, како и на тоа кои делови од градот се средишта за различни категории на криминал.

Ќе сфатите дека барањето за податоци ќе треба да ги опфати и датумот и точното време на сите кривични дела, видот на кривичното дело (убиство, кражба, провална кражба итн.), како и адресата на која се случило кривичното дело. Значи „Датум“, „Точно време“, „Категорија на кривично дело“ и „Адреса“ се минимумот од променливи што ви се потребни за да одговорите на тие прашања.

Сепак, треба да бидете свесни дека таквата група податоци со четири променливи не може да одговори на повеќе потенцијално интересни прашања, како што се расата или родот на жртвите, вкупната вредност на украдените предмети или кои полицајци извршиле најголем број апсења. Исто така, може да се случи да добиете податоци само за ограничен период, на пример, за последните три години, што значи дека нема да можете да зборувате дали и како криминалните обрасци се менувале во подолг временски период. Таквите прашања може да се надвор од планираните цели на вашата сторија, и тоа е во ред. Она што не е во ред е да почнете со анализа на податоците и одеднаш да решите дека сакате да го знаете процентот на кривични дела што биле решени со апсење во различни делови на градот.

Една од лекциите што можеме да ги научиме овде е дека е добро да ги побараме сите променливи и записи што постојат во базата на податоци, наместо само извадокот што ќе одговори на прашањата кои ги поставува приказната што ја имаме на ум. (Всушност, добивањето на сите податоци може да се покаже и поевтино отколку да се обидуваме да обезбедите само една подгрупа, особено ако треба на агенцијата да ѝ платите за програмирањето што е потребно за да се подготви извадокот.) Секогаш можете сами да го подготвите извадокот што ви е потребен, а пристапот до целата група податоци ќе ви овозмо-

жи да одговорите на прашањата што може да произлезат од известувањето и дури да донесат нови идеи за идни стории. Поради законите за доверливост на податоците или други прописи, некои променливи, како што се идентитетот на жртвите или имињата на доверливите информатори, нема да можат да бидат објавени. Но, дури и нецелосна база на податоци е подобра од никаква база на податоци, сè додека сте свесни на кои прашања редактираната база може, а на кои не може да одговори.

Чистење на збркани податоци

Еден од најголемите проблеми во работата со бази на податоци е тоа што често податоците што ги користите за анализа се собирали за бирократски, а не за новинарски потреби. проблемот лежи во тоа што тие два пристапи почитуваат различни стандарди за точност.

На пример, една од основните функции на која било база на податоци на кривичниот правосуден систем е да обезбеди обвинетиот Џонс на време да биде доведен од затвор на распит пред судијата Смит. За таа цел не е важно дали датумот на раѓање на Џонс е погрешен, дали името на улицата на која живее е погрешно напишано или дури дали му го погрешиле татковото име. Главно е дека системот сè уште може да го искористи таквиот несовершен запис за да обезбеди дека Џонс ќе се појави во судницата на Смит во зацртаното време.

Таквите грешки, од друга страна, можат да имаат негативно влијание врз обидите на новинарот да ги открие обрасците што се повторуваат во базата на податоци. Поради тоа, првата голема работа што треба да се заврши откако ќе добиете нова група податоци е да проверите колку е чиста и потоа да ја исчистите. Еден брз начин да побарате каде постојат збркани податоци е да подготвите табели за зачестеност за оние категориски променливи кај кои би се очекувало да имаат релативно мал број различни вредности. [Кога користите „Ексел“, на пример, за тоа можете да ги искористите опциите „Филтер“ (Filter) или „Приспособлива табела“ (Pivot tables) на секоја од категориските променливи].

Да ја разгледаме променливата „Род“, како лесен и едноставен пример. Ќе откриете дека полето „Род“ вклучува повеќе различни вредности, на пример: машки, женски, М, Ж, 1, 0, МАШКИ, ЖЕНСКИ итн., вклучувајќи и погрешно напишани како што е „Женки“. За вистинска родова анализа мора да ги стандардизирате вредностите – да решите да ги користите вредностите „М“ и „Ж“, на пример – и потоа да ги промените сите варијации во согласност со прифатениот стандард. Други често користени бази на податоци што имаат такви проблеми се податоците за финансирањето кампањи во САД, каде што во полето „Професија“ може да стои „Адвокат“, „Правник“, „Адв.“, „Правен застапник“, „Бранител“ и секакви други варијации или погрешно напишани звања; повторно, се работи за тоа да ги стандардизирате професионалните звања и да ги сведете на еден краток список од можности.

Чистењето на податоците станува уште попроблематично кога се работи за имиња. Дали во случаите „Џозеф Т. Смит“, „Џозеф Смит“, „Џ. Т. Смит“, „Џоз. Смит“ и „Џо Смит“ се работи за истото лице? Решението може да бара да се прегледаат и други променливи, како што е домашната адреса или датумот на раѓање, или подлабоко истражување на други записи и архиви. Алатките како што е „Гугл рифајн“ (Google Refine) можат да ја забрзаат и да ја олеснат задачата на чистење и стандардизација.

Непрочистени податоци

Благодарение на претежно строгото законодавство за јавните архиви и документи, добивањето податоци во САД не е толку голем проблем колку што може да биде во многу други земји. Сепак, откако ќе ги добиеме бараните податоци, сè уште се соочуваме со фактот дека работиме со податоци што се собирали за бирократски, а не за аналитички потреби. Податоците често се „валкани“ и нудат нестандардизирани вредности. Се случило повеќе пати да добијам податоци што не соодветствуваат на пропишаниот изглед и на поимникот приложен со датотеката. Некои агенции и тела и натаму инсистираат да ги доставуваат податоците во „незгодни“ формати како што е .pdf, па треба да се подложат на конверзија. Поради таквите проблеми, навистина цениме кога, од време на време, ќе добиеме група податоци што не ни прави никакви проблеми.

- *Стив Дојџ, Школа за новинарство „Волтер Кронкајт“, Државен универзитет на Аризона*

Податоците можат да имаат својства што не се документирани

Таканаречениот речник или поимник на податоците е „каменот од Розета“ што помага во разбирањето на соодветната база на податоци. Вообичаено, таквите поимници (може да се работи за текст, „PDF“ или дури датотека со сметководствени табели) покажуваат на кој начин е форматирана датотеката (делимитиран текст, текст со фиксна маргина, Excel, dBase итн.), редоследот на променливите и нивните имиња, како и видот податоци (datatype) во секоја променлива (линија текст, интеграл, децимали итн.). Тие информации ќе ви помогнат во правилното внесување на датотеката во аналитичката софтверска апликација што сакате да ја користите (Excel, Access, SPSS, Fusion Tables, некоја од многуте верзии на SQL-апликации итн.).

Друг клучен елемент на секој речник на податоци е објаснувањето на сите шифри што се користат за определени променливи. На пример, „Родот“ може да се шифрира така што вредноста „1“ ќе означува маж, а вредноста „0“ ќе означува жена. Кривичните дела можат да бидат шифрирани според статутарните броеви доделени за секое одделно кривично дело во вашата јурисдикција. Здравствените картони што се чуваат во болниците можат да користат

кој било од стотиците петоцифрени шифри за означување на дијагнозата на болеста од која е лечен пациентот. Без речник на податоци, анализата на таквите податоци е тешка, безмалку невозможна.

Сепак, проблеми можат да се јават дури и ако го имате на располагање речникот на податоци. Со таков случај, пред неколку години, се соочија новинарите на „Мајами хералд“ (Miami Herald) од Флорида во анализата на разликите во казните за лицата уапсени за возење под дејство на алкохол што ги изрекувале различни судии. Новинарите од судскиот систем ги добија податоците за пресудите и ги анализираа бројките наведени во три различни променливи вредности, според казните наведени во речникот на податоци: времетраењето на затворските казни, времетраењето на притворот и износот на досудените глоби. Бројките укажуваа на значајни разлики меѓу индивидуални судии и на новинарите им обезбедија докази за пишување сторија за тоа како некои судии се построги, а други се поблаги во казнувањето на престапниците.

Сепак, кај сите судии, во 1 до 2 отсто од случаите немаше прикажано ниту затворска казна, ниту притвор, ниту глоба. Значи, графиконот на кој беа прикажани обрасците на пресудите на индивидуалните судии вклучуваше и сосема мал број случаи во кои „Не е изречена казна“, речиси како отпосле да се сетиле на таа можност. По објавувањето на приказната и на графиконот, судиите бурно реагираа, велејќи дека „Хералд“ ги обвинува за кршење на законот на сојузната држава што пропишува задолжителна казна за сите осудени за возење под дејство на алкохол.

Новинарите повторно појдоа во судската архива за да прашаат на што може да се должи таквата грешка. Им беше одговорено дека во тие случаи се работело за сиромашни граѓани за кои тоа претставувало прв престап. Во нормални услови би им била изречена глоба, но немале пари да ја платат. Поради тоа, судиите изрекувале казна „општествено корисна работа“, на пример, чистење на сметот покрај јавните патишта. На крај се покажа дека законот што пропишува казнување на сите престапници бил усвоен откако структурата на базата на податоци веќе била подготвена. Сите службеници во судската администрација знаеле дека внесот „нула“ во некоја од променливите „затвор-притвор-глоба“ означува „општествено корисна работа“. Тоа сепак не беше наведено во речникот на податоци, па „Хералд“ мораше да објави исправка и да се извини.

Лекцијата што треба да ја научиме во овој случај е секогаш да побараме од институцијата што ни ги дава податоците дали постојат недокументирани елементи во податоците, без оглед дали се работи за новоусвоени шифри што не се вклучени во речникот на податоци, промени во изгледот на датотеката или што и да е друго. Исто така, секогаш проверувајте ги резултатите од вашата анализа и поставете си го прашањето: „Дали ова има смисла?“ Графиконот на новинарите на „Мајами хералд“ требаше да биде подготвен во определен рок, а тие беа толку фокусирани на просечните казни изречени од секој индивидуален судија што целосно ги занемарија малкуте случаи во кои, навидум,

не беше изречена никаква казна. Требаше да си го постават прашањето дали е логично буквално сите судии да го кршат законот на Флорида, дури и кога се работи за толку безначајно отстапување од законските норми.

- Стив Дојџ, Школа за новинарство „Волиер Кронкајџ“, Државен универзитет на Аризона

Измешани, скриени и отсутни податоци

Помнам една интересна ситуација што настана кога се обидовме да пристапиме до унгарските податоци за субвенциите за фармерите на ЕУ. Внатре имаше сè – но во форма на огромна „PDF“-датотека, измешано со податоците за субвенциите за земјоделството на национално ниво. Нашите програмери мораа да работат со часови за да можеме да ги искористиме податоците.

Друга интересна ситуација настана додека работевме со податоците за субвенциите за рибарство на ЕУ што националните тела за платен промет во сите 27 земји членки мораат задолжително да ги објават. Следи еден извадок од извештајот што го напишавме на таа тема (достапен на адресата <http://bit.ly/alfter-eu27>): „Во Велика Британија, на пример, форматот во кој се прикажани податоците варира од лесни за користење HTML-пребарувачи до прегледи на добитниците на субвенции објавени како „PDF“-прегледи или списоци на добитниците на субвенции, во различни формати, скриени на дното на соопштенијата за медиумите. А тоа се случува во само една земја членка на ЕУ. Во меѓувреме, Германија и Бугарија објавуваат празни списоци. Имињата на колоните и редовите се на место, но внатре нема никакви податоци“.

– Бриџит Алфјер, *Journalismfund.eu*

Векна леб од 32 фунти

Една сторија на „Велс он сандеј“ (Wales on Sunday) за сумите што владата ги троши на производите без глутен што се продаваат само со лекарски рецепт, во насловот тврдеше дека владата плаќа 32 фунти за векна леб. Всушност, таа сума се однесуваше на 11 векни од кои секоја чинеше по 2,82 фунти.

Тие бројки, добиени од писмениот одговор што новинарите го добија од Собранието на Велс и од објавените статистички податоци на Националниот систем за здравствена заштита на Велс, се однесуваа на трошоците по предмет од препишан лекарски рецепт. Сепак, добиените податоци, во речникот на податоци, не нудеа дополнително објаснување или дефиниција за тоа што е „предмет на лекарски рецепт“ или на кој начин би можеле да се дефинираат во засебна колона со наведени податоци за количините.

Претпоставката беше дека се однесува на единечен предмет – на пример, една векна леб – наместо, како што беше случај, на пакет од неколку векни.

Никој, ниту луѓето што одговорија на прашањето доставено во писмена форма, ниту канцеларијата за односи со медиумите, не постави прашање за количеството до понеделникот по објавувањето на сторијата.

Значи, не смееме да претпоставуваме дека дополнителните белешки што доаѓаат со владините податоци ќе објаснат какви информации се претставени или дека луѓето задолжени за работа со податоците во институциите ќе сфатат дека податоците не се јасни, дури и кога ќе ги соочите со вашата погрешна претпоставка.

Општо земено, весниците сакаат нешто од коешто ќе излезе добар наслов. Ако не постои ништо што сосема очигледно го оспорува таквото толкување на податоците, полесно е да се оди со она што ќе произведе добар наслов, а не да се проверува повторно, со ризик сторијата да пропадне, особено ако треба да се почитуваат строги рокови.

The screenshot shows the WalesOnline website interface. At the top, the logo 'WalesOnline.co.uk' is displayed with the tagline 'News from Wales and beyond, plus politics, business, sport, life and style... it's all here'. The weather is shown as 11°C - 11 kph wind. Below the logo is a navigation bar with links: Home, News, Rugby, Sports, Football, Lifestyle, Business, and Classifieds. Underneath this is another row of links: Health Check Wales, CardiffOnline, and Vouchers. A 'Hot Topics' section lists names like Nikkita Grender, Rebecca Aylward, Mike Phillips, Fracking, Gavin Henson, and Imogen. The main article headline is 'Prescriptions for gluten-free bread costing Welsh taxpayers £32' by Claire Miller, dated Jul 17 2011. The article text states: 'The Welsh NHS is forking out £32 a time for prescriptions for gluten-free bread. The average prescription for the specialist food cost £32.27, and was provided to people with the serious condition coeliac disease.' An image of a loaf of bread is shown. To the right of the image is a 'Related Tags' section with links: bread, gluten free, hayfever, nhs, over the counter medicines, painkillers, pasta, prescriptions, and wales. At the bottom of the article text, it says: 'Gluten-free loaves are available in shops for just £2.25, but are just one of scores of prescription items for which the Welsh NHS pays vast sums.'

Слика 5-3. Лекарските рецепти за леб без глутен за кои секој даночен обврзник од Велс плаќа по 32 фунти (WalesOnline)

Новинарите, сепак, се обврзани да ги проверат „неверојатните и смешни“ тврдења, дури и ако тоа значи дека сторијата ќе пропадне на списокот на важност на вестите.

– Клер Милер, „Велс онлајн“ (Claire Miller, WalesOnline)

Од податоци до сторија

За да ги привлечете читателите, треба да знаете како во насловот да ги „погодите“ со некаква бројка што ќе ги натера да се застанат и да се замислат. Мора да можат да ја прочитаат сторијата без да мора да знаат дека произлегла од некоја група податоци. Напишете возбудлива приказна и секогаш имајте на ум за кого пишувате, кој ќе го чита напишаното.

Еден пример за таквиот пристап имаме во проектот на „Бирото за истражувачко новинарство“ (Bureau of Investigative Journalism) што ги користеше податоците на „Системот за фискална транспарентност“ на Комисијата на ЕУ (EU Commission's Financial Transparency System). На податоците им пристапиме со однапред смислени прецизни прашања.

Во податоците баравме клучни зборови како што се „коктел“, „голф“ и „слободни денови“. Тоа ни овозможи да пресметаме колку Комисијата потрошила на тие активности и повлече многу прашања и воведе во игра многу приказни што требаше да се проверат.

Од друга страна, клучните зборови нема секогаш да ве одведат до тоа што го барате – понекогаш мора да подзастанете и да размислите што всушност барате. Во рамките на проектот сакавме да дознаеме и колку пари комесарите трошат на патувања со приватни авиони, но податоците не ја содржеа фразата „приватен млазен авион“, па моравме да ги побараме патничките агенции што ги користеле за такви услуги. Откако дознавме дека таквите услуги за Комисијата ги обезбедува компанијата „Абелаг“ (Abelag), можевме да ги пребаруваме податоците за да видиме колку пари се потрошени на услугите што ги нуди „Абелаг“.

Таквиот пристап ни овозможи да имаме јасно дефинирана цел во пребарувањето на податоците – да пронајдеме бројка што ќе ни даде добар наслов, а за нијансите одлучувавме подоцна.

Друг пристап што може да се примени е да се појде од некоја „црна листа“ и во неа да се бараат исклучоците. Еден лесен начин за извлекување приказни од податоците е да знаете што не би требало да пронајдете во нив! Добар пример за тоа е проектот на „Фајненшал тајмс“ и „Бирото за истражувачко новинарство“ што ги истражуваше структурните фондови на ЕУ.

Во пребарувањето на податоците се водевме од правилата на Европската комисија за тоа на кои компании и асоцијации ќе им биде забрането да добијат средства од структурните фондови. Еден пример беа средствата доделени на производителите на тутун.

Кога проверивме дали во податоците се јавуваат имиња на тутунски компании и тутунопроизводители, откривме дека компанијата „Бритиш-американ тобако“ (British American Tobacco) добила 1,5 милиони евра за една фабрика во Германија.

Тоа што тие средства ги кршеа правилата на Европската комисија за распределување на структурните фондови ни помогна набрзина да пронајдеме добра приказна во податоците.

Никогаш не можете да знаете однапред што сè ќе најдете во некои податоци, па слободно навалете. Тоа бара смел пристап кој, генерално, функционира најдобро кога се обидувате да ги идентификувате очигледните карактеристики што ќе излезат по филтрирањето (најголемите, екстремните вредности, најчестите вредности итн.).

– *Каелин Бар, „Сиѿивајр“ (Caelainn Barr, Citywire)*

Приказните во податоците

Дата-новинарството понекогаш остава впечаток како да се работи за начинот на прикажување на податоците – визуелизации што придонесуваат за брзо и интензивно разбирање на некој аспект на броевите, или интерактивни бази на податоци што можат да се пребаруваат и што на индивидуалните лица им овозможуваат да проверат каде во таа слика се наоѓа нивната улица или болницата во која се лекуваат. Тоа, секако, си има своја вредност, но како и сите други форми на новинарство, и во дата-новинарството приоритет се приказните. Значи, какви сè приказни можете да пронајдете во податоците? Врз основа на моето искуство во Би-Би-Си, подготвив еден список, една „типологија“ на различни видови приказни.

Сметам дека следната листа ќе ви помогне не само во анализата на податоците, туку и во фазите што претходат, во собирањето на податоците (без оглед дали користите јавно достапни податоци или ги подредувате податоците добиени преку барања за слободен пристап до информации).

Мерење

Наједноставната приказна во која нешто пребројувате или собираете: „Локалните совети минатата година потрошиле, вкупно на национално ниво, икс милијарди фунти на спојувалки за хартија“. Често е тешко да се каже дали тоа е многу или малку. За тоа ви треба и контекстот, што можете да го обезбедите по пат на:

Пройорционала сјоредба

„Минатата година, локалните совети потрошиле две третини од буџетот за канцелариски материјали на спојувалки за хартија“.

Внајрешна сјоредба

„Локалните совети трошат повеќе на спојувалки за хартија отколку на дистрибуција на оброци за старите лица“.

Надворешна споредба

„Парите што локалните совети лани ги потрошиле на спојувалки се два пати поголеми од износот резервиран за меѓународна помош во националниот буџет“.

Постојат и други начини за контекстуализирано или компаративно прегледување на податоците:

Промени во определен временски период

„Парите што локалните совети ги трошат на спојувалки се зголемиле три пати во последните четири години“.

„Лиѓашки табели“

Тие често се поделени по географска област или по институција, при што мора да се осигурате дека основата за споредба е фер и правична (на пример, со земање предвид на големината на локалното население). „Советот на Борсетшајр троши повеќе на спојувалки по вработен од сите други локални самоуправи, со стапка што е четири пати повисока од националниот просек“.

Или може да ги групирате податоците по тема:

Анализа по категоријата

„Локалните совети во кои мнозинство има Пурпурната партија трошат 50 отсто повеќе на спојувалки од советите контролирани од Жолтата партија“.

Или можете да вадите нумерички врски и односи помеѓу факторите:

Асоцијација

„Советите во кои седат политичари што добиле донации од компании што се занимаваат со продажба на канцелариски материјал трошат повеќе на спојувалки, а трошоците во просек растат за 100 фунти за секоја донирана фунта“.

Се разбира, секогаш имајте на ум дека постои разлика помеѓу корелацијата и каузалноста. Ако ги истражувате трошоците на набавка на спојувалки, дали се обидувате да ги обезбедите и следниве бројки?

- Вкупните набавки за да го обезбедите контекстот?
- Географски/историски/друг преглед за да обезбедите споредбени податоци?
- Дополнителни податоци за да се осигурате дека споредбите се фер и правични, на пример, бројот на населението?
- Други интересни податоци што можат да се анализираат и резултатите да се споредат или да се поврзат со трошоците?

— *Марџин Розенбаум, Би-Би-Си*

Новинарите и нивните омилен алатки

„Псссссссс“. Тоа е звукот на декомпресијата на вашите податоци од херметички затворениот сад во кој се чуваат. Што понатаму? Што точно барате? Кои алатки ги користите за да „нурнете“ во податоците? Прашавме неколку дата-новинари да ни кажат нешто повеќе за начините на кои работат со податоците. Следат нивните одговори:

Ние, во „Даталогот на Гардијан“, навистина сакаме да оствариме интеракција со читателите, а со тоа што се трудиме брзо да им овозможиме да го реплицираат нашето дата-новинарство, им овозможуваме да го надградат она што сме го сработиле и понекогаш да забележат нешто што ние сме пропуштиле. Тоа значи дека алатките се онолку добри колку што нивната употреба е интуитивна. Се обидуваме да избереме алатки што можат да ги користат сите, без потреба претходно да совладаат некаков програмски јазик, да посетуваат специјална обука или да плаќаат некаков надоместок за користењето.

Токму поради тие причини, во моментот во голема мера ги користиме производите на „Гугл“. Сите податоци што ги обработуваме и објавуваме се достапни во форма на сметководствена табела на „Гугл“ (Google Spreadsheet), што значи дека сите оние што имаат корисничка сметка на „Гугл“ можат да ги спуштат податоците, да ги внесат во сопствената корисничка сметка и да ги искористат за изработка на сопствени графички, да ги „спастрат“ податоците и да изработат приспособливи табели, или да ги внесат податоците во друга алатка што сакаат да ја користат.

За мапирање на податоците ги користиме табелите на „Гугл фјужн“ (Google Fusion tables). Кога изработуваме термички мапи во „Фјужн“, ги споделуваме нашите „КМЛ“ (Keyhole Markup Language – KML) геодетски датотеки за да можат читателите да ги „спуштат“ и самите да изработат свои термички мапи – можеби преку додавање дополнителни слоеви податоци на оригиналната мапа на „Даталогот“. Друга одлична можност што ја нудат алатките на „Гугл“ е што работат на повеќето платформи (персонален компјутер, мобилен телефон или таблет) што ги користат нашите читатели.

Освен „Google Spreadsheets“ и „Google Fusion“, во секојдневната работа користиме уште две алатки. Првата е „Табло“ (Tableau), софтвер за визуелизација на повеќедимензионални групи податоци; а втората е „Мениајс“ (ManyEyes, <http://www-958.ibm.com/software/analytics/manyeyes/>), за брза анализа на податоците. Ниедна од спомнатите алатки не е совршена, па затоа постојано бараме подобри алатки за визуелизација што ќе бидат привлечни за нашите читатели.

-- Лиза Еванс, „Гардијан“

Дали некогаш ќе станам програмер на компјутерски код (кодер)? Веројатноста тоа да се случи е мала! Во секој случај, не мислам дека сите новинари мора да знаат да пишуваат компјутерски код. Она што сметам дека ќе им се исплати е да бидат свесни, во најопшта можна смисла, за сите можности што постојат и да знаат како да разговараат со кодерите.

Ако штотуку почнувате, нема потреба од брзање. Треба да ги убедите колегите и уредниците дека работата со податоци ќе ви донесе стории што инаку не би ги добиле и дека тоа се исплати. Кога ќе ја согледаат вредноста на таквиот приод, тогаш можете да преминете на посложени стории и проекти.

Ви советувам да научите да работите во „Ексел“ (Excel) и да почнете со неколку едноставни приказни. Почнете со помали проекти и полека преминете на анализа на бази на податоци и мапирање. „Ексел“ овозможува толку многу работи – се работи за извонредно моќна алатка, а повеќето луѓе не користат ни делче од сите можности што ги нуди. Ако можете, посетувајте курс за користење на „Ексел“ за новинари, на пример, обуката што ја нуди „Центарот за истражувачко новинарство“.

Во поглед на толкувањето на податоците, пријдете му крајно сериозно. Морате да бидете совесни и темелни во работата. Внимавајте и на најмалите детали и сомневајте се во сопствените наоди. Водете белешки за начинот на кој ги обработувате податоците и чувајте еден примерок од оригиналните податоци. Лесно се прават грешки. Јас, на пример, ги повторувам моите анализи по два или три пати, и секогаш почнувам буквално од почеток. Уште подобро е ако вашиот уредник или некој друг ги анализира податоците засебно и потоа ги споредите резултатите.

– Синџија О’Мурку, „Фајненшил џајмс“

Способноста за пишување и активирање комплексни софтверски решенија, со брзина со која новинарот може да напише статија, е нова работа. Тоа порано беше многу подолг процес. Нештата се променија благодарение на развојот на две рамки за брз развој што функционираат со слободен/отворен изворен код, „Џанго“ (Django, <https://www.djangoproject.com/>) и „Руби он реилс“ (Ruby on Rails, <http://rubyonrails.org/>), некаде кон средината на првата деценија од 21 век.

„Џанго“, изработен во програмскиот јазик „Питон“ (Python), е дело на Адријан Холовати (Adrian Holovaty) и редакцискиот тим на „Лоренс журнал-ворлд“ (Lawrence Journal-World) од Лоренс, Канзас. „Руби он реилс“ беше развиен во Чикаго, од страна на Дејвид Хајнемејер Хенсон (David Heinemeier Hansson) и „37сигналс“ (37Signals), компанија за развој на веб-апликации.

Иако двете рамки нудат различен приод кон „образецот МВЦ“ (Model-View-controller (MVC) pattern, образец за примена на кориснички интерфејси), и двете се одлични алатки што овозможуваат брз развој дури и на најкомплексни веб-апликации. Нивното користење елиминира некои од

најосновните постапки при изработувањето апликации. На пример, креирањето и носењето предмети од базата на податоци, врзувањето на „URL“-адреси со соодветниот компјутерски код се вградени во рамките, па девелоперите не мора да губат време на програмирање на толку основни работи.

Иако до сега никој не направил формално истражување на тимовите што изработуваат новински апликации во САД, општоприфатено е дека најголемиот број тимови користат една од тие две рамки во изработката на новински апликации што зад себе имаат бази на податоци. Ние, во „ПроПублика“, го користиме „Руби он реилс“.

Равојот на сервиси за брзо обезбедување „исечоци“ од веб-сервери (rapid web server „slice“ provisioning) каков што е „Амазон веб сервисис“ (Amazon Web Services, <http://aws.amazon.com/>), исто така, отстрани неколку проблеми што го забавуваа активирањето веб-апликации.

Настрана од сето тоа, алатките што ги користиме за работа со податоци се прилично стандардни. Користиме „Гугл рифајн“ (Google Refine) и „Мајкрософт Ексел“ (Microsoft Excel) за прочистување на податоците; „Ес-Пи-Ес-Ес“ (SPSS) и „Ер“ (R) за статистичка обработка; „АркГИС“ (ArcGIS, <http://www.arcgis.com/features/>) и „КјуГИС“ (QGIS) за „ГИС“ (Geographic Information Services – сервиси за географски информации) операции; „Гит“ (Git) за работа со изворниот код; „Текстмејт“ (TextMate), „Вим“ (Vim) и „Сублајм текст“ (Sublime Text) за пишување код; и микс од „МајЕс-Кју-Ел“ (MySQL), „ПостгреЕс-Кју-Ел“ (PostgreSQL) и „Ес-Кју-Ел сервер“ (SQL Server) за базите на податоци. Изградивме сопствена рамка во „Џаваскрипт“ (JavaScript) што ја нарекуваме „Глас“ (Glass) и ја користиме за брза изработка на апликации со „тешка“ насловна страница во „JavaScript“.

– Скоџ Клајн, ПроПублика

Понекогаш наједноставната алатка е најдобра алатка – лесно се случува да ја потцениме моќта и способноста на сметководствените табели. А токму користењето сметководствени табели во времето кога сè се чуваше во оперативниот систем „ДОС“ (DOS), ми овозможи да разберам една комплексна формула додека работев на договорот за партнерство на сопствениците на клубот за американски фудбал „Тексас ренџерс“ (The Texas Rangers), во периодот кога и Џорџ В. Буш (George W. Bush) беше еден од главните сопственици. Една сметководствена табела ми помогна да ги означам координатните точки и грешките во пресметките. Можам да пишувам „скрипти“ (кратки компјутерски програми за автоматско извршување повторливи постапки) за чистење и многу други работи. За еден дата-новинар, тоа е дел од основните алатки.

Инаку, моите омилен алатки се уште помоќни – „SPSS“ за статистичка анализа и програми за мапирање што ми овозможуваат да ги видам географските обрасци.

– Черил Филијс, „Сиеџл џајмс“

Јас сум голем љубител на „Питон“ (Python). Тоа е извонреден програмски јазик со отворен изворен код, лесен за читање и пишување (на пр., секој ред не мора да завршува со „точка-запирка“). Освен тоа „Python“ нуди огромна база на корисници и додатоци (наречени пакети), буквално за сè што ви треба.

Според мене, дата-новинарите многу ретко имаат потреба да ја користат рамката „Џанго“ (Django). „Џанго“ е рамка за веб-апликации изработена во „Python“, т.е. алатка за изработка на големи веб-апликации што се потпираат на бази на податоци. Дефинитивно е „претерана“ кога се работи за мали интерактивни инфографики.

Користам и „QGis“ (<http://www.qgis.org>), алат со отворен изворен код што нуди широк спектар „GIS“-функции кои повремено им се потребни на дата-новинарите што работат со географски податоци. Ако сакате да префрлате геопросторни податоци од еден формат во друг, тогаш „QGis“ е вистинската алатка за вас. Може да работи со скоро сите формати за чување геоподатоци што постојат (Shapefile, KML, GeoJSON, итн.). Ако сакате да исцртате нови региони, „QGis“ ќе ви помогне и во тоа. Освен тоа, околу „QGis“ е собрана огромна заедница, така што на интернет постојат тони и тони ресурси, прирачници и упатства.

Примарната цел на софтверската апликација на проектот „R“ (<http://www.r-project.org/>) е да послужи како алатка за научни визуелизации. Тешко дека постои метод на визуелизација или рачна работа со податоци што веќе не е интегриран во „R“. „R“ е цел универзум сам за себе, „Мека“ за анализа на визуелни податоци. Еден недостаток е што мора да научите (уште еден) цел програмски јазик. Но, кога ќе ја поминете првата етапа на кривата на учење, не постои помоќна алатка. Добро обучените дата-новинари можат да го користат „R“ за анализа на големи групи податоци што ги надминуваат можностите на „Excel“ (на пример, табела со милион редови податоци).

Она што навистина е добро кај „R“ е што можете да водите прецизен „протокол“ од тоа што го правите со податоците низ целиот процес – од читањето „CSV“ датотека до генерирањето графикони. Ако има промена на податоците, графиконот се обновува со само едно кликување со глушецот. Ако некој има прашања за интегритетот и валидноста на графиконот, можете да го прикажете точниот извор, што значи дека секој може сам или сама повторно да го изработи графиконот (или, можеби, да открие ако сте направиле некоја грешка).

„Нампај“ (NumPy) и „МатПлотЛиб“ (Matplotlib) се еден поинаков начин да сторите сè што е спомнато погоре со користење „Питон“. Меѓутоа, нивното користење бара повисоко ниво на познавање на работата со „Питон“. „NumPy“ (<http://www.numpy.org/>) и „Matplotlib“ (<http://matplotlib.org/>) всушност се два примери за готови софтверски пакети подготвени во „Питон“. Се користат за анализа и визуелизација на податоци, со тоа што двата пакети можат да произведуваат само статични визуелизации. Не можат да се користат за изработка на интерактивни графикони со појаснувања за користените алатки и други напредни решенија.

Иако лично не користам „Мепбокс“ (MapBox) (<https://www.mapbox.com/>), слушам дека се работи за одлична алатка ако сакате да подготвите посоставени мапи врз основа на „OpenStreetMap“. Алатката ви дозволува, на пример, да ги приспособувате стилските на прикажување на мапата (бои, ознаки итн.). „MapBox“ има и придружен елемент што се вика „Leaflet“ (Леток). Кај „Leaflet“ во основа се работи за повисоко ниво на „JavaScript“-библиотека за мапирање што овозможува лесен премин помеѓу различни сервиси што нудат мапи („OSM“, „MapBox“, „Google Maps“, „Bing“ итн.).

„Рафаел-Џеј-Ес“ (RaphaelJS, <http://raphaeljs.com/>) е библиотека со визуелизации со понизок квалитет што ви овозможува да работите со основни елементи (крugови, линии, текст) и од нив да изработите анимација, да додадете интерактивни елементи итн. Не содржи, на пример, готови хоризонтални столбести дијаграми, што значи дека сами ќе мора да ги нацртате потребните правоаголници.

Од друга страна, кај „Raphael“ е добро тоа што сè што ќе создадете ќе функционира и во „Интернет експлорер“ (Internet Explorer, IE), што не е случај со многу други (наистина извонредни) библиотеки со визуелизации како што е „Дез“ (d3). Жално е што многу корисници сè уште користат „Интернет експлорер“, а јас не познавам ниту една редакција што ќе се осмели да игнорира 30 отсто од своите корисници.

Освен „RaphaelJS“, постои можност да се направи резервна верзија за „Интернет експлорер“, изработена во „Флеш“ (Flash). Тоа, во основа, е она што „Њујорк тајмс“ го прави со своите датотеки. Секако, тоа значи дека секоја апликација мора да ја развивате два пати.

Јас сè уште не сум најубеден кој е „најдобриот“ процес за пренос на визуелизации за „Интернет експлорер“ и за модерните прегледувачи. Често откривам дека апликациите изработени во „RaphaelJS“ работат мошне бавно во „IE“, околу десет пати побавно отколку во „Flash“, со користење посовремени прегледувачи. Значи, резервните верзии во „Flash“ можат да се покажат како подобра опција ако сакате на сите корисници да им понудите анимирани визуелизации со висок квалитет.

— Греџор Ајш, Фондација „Одговорно знаење“

Мојата омилена алатка за сите намени е „Excel“. „Excel“ може да се справи со мнозинството проблеми што можат да се појават во известувањето со помош на компјутер (ИПК), а ја има и таа предност што лесно се учи и им е достапен на мнозинството новинари. За соединување табели вообичаено користам „Access“, за потоа повторно да ја префрлам соединетата табела во „Excel“ за натамошна обработка. За географска анализа го користам софтверот „ArcMap“ на „ЕСРИ“ (ESRI; <http://www.esri.com>). Тоа е моќна алатка што ја користат агенциите и институциите што собираат геокодирани податоци. „TextWrangler“ е одлична алатка за прегледување текстуални податоци со многу чудни планови и граници, а дозволува и софистицирано пребарување и замена на вообичаени изрази. Кога имам потреба од статистички техники како што е линеар-

ната регресија, го користам софтверот „SPSS“ со мени со опции што лесно се користат. За сериозна работа, како што се групите на податоци со милиони записи кај кои е потребно сериозно ниво на филтрирање и програмирани трансформации на варијабилите, го користам софтверот „SAS“.

-- Стив Доџ, Школа за новинарство „Волиер Кронкајв“

Омилени алатки за хакирање, „стругање“ и играње со податоците ни се „Python“ и „Django“. За градење „луди“ веб-мапи ги користиме „ПостГИС“ (PostGIS), „QGIS“ и алатките на „MapBox“. Во моментот имаме мртва трка помеѓу „R“ и „NumPy+ Matplotlib“ за тоа кој ќе биде нашиот омилен алат за анализа на истражувачки податоци, иако во последно време најмногу се служиме со „CSVKit“, алатка што сами ја произведовме: Безмалку сè што правиме се чува во „облак“ на интернет (web-cloud).

— Брајан Бојер, „Чикаго трибун“

Во „Ла насион“ користиме:

- „Excel“ за чистење, организирање и анализа на податоците;
- „Google Spreadsheets“ за објавување и поврзување со услугите како што се „Google Fusion Tables“ и платформата за отворени податоци „Junar Open Data Platform“;
- „Junar“ (<http://www.junar.com/>) за споделување и размена на нашите податоци и нивно вградување во нашите статии и блогерски постови;
- „Tableau Public“ (<http://www.tableausoftware.com/public/>) за интерактивните визуелизации на податоци;
- „Qlikview“ (<http://www.qlikview.com/>), мошне брза алатка од арсеналот на деловното разузнавање за анализа и филтрирање големи групи податоци;
- „NitroPDF“ (<http://www.nitropdf.com/>) за конверзија на „PDF“-датотеки во текст и „Excel“-датотеки; и
- „Google Fusion Tables“ за визуелизации прикажани на мапа.

— Анхелика Пералта Рамос, „Ла насион“, (Арџентина)

Како „грасрутс“-заедница без предрасуди во полза на некоја технологија, ние во „Транспаренсија хакер“ (Transparência Hacker) користиме многу различни алатки и програмски јазици. Секој член на заедницата си има свој избор на софтвер што најмногу му се допаѓа и таквата разновидност истовремено е и наша силна страна и наша слабост. Неколкумина од нас, всушност, работиме на „Линукс дистрибуција на Транспаренсија хакер“ (Transparência Hacker Linux Distribution) што ќе можеме да ја подигнеме на кој било компјутер и да почнеме со хакирање податоци. Комплетот на алатки содржи неколку интересни алатки и библиотеки за работа со пода-

тоци како што се „Рифајн“ (Refine), „ЕрСтудио“ (Rstudio) и „Опен офис калк“ (Open Office Calc) („Calc“ е запоставен од многу добри познавачи на работата со компјутери, но навистина е корисен за работа на мали и брзи проекти). Исто така, многу го користиме „Скрејпервики“ (Scrapewiki) за брза подготовка на прототипови и зачувување на резултатите на интернет.

Што се однесува до визуелизацијата и графиките, ни се допаѓаат повеќе алатки. „Python“ и „NumPy“ се мошне моќни алатки. Малку луѓе во заедницата го пробале „R“, но мислам дека, во крајна линија, некако излегува дека мнозинството од нашите проекти ги работиме со користење на библиотеки за исцртување графикони на „Javascript“ – „d3“, „Флот“ (Flot, <http://www.flotcharts.org/>) и „RaphaelJS“. На крајот, многу експериментираме со мапирање и „Тајлмил“ (Tilemill, <https://www.mapbox.com/tilemill/>) се покажа навистина како интересна алатка.

– Педро Маркун, „Трансјаренција хакер“

Користење визуелизаци за подобар увид во податоците

Визуелизацијата е клучна во анализата на податоците. Таа е првата линија на нападот и ја открива сложената структура на податоците што не може да биде апсорбирана на ниту еден друг начин. Откриваме незамисливи ефекти и поставуваме прашања за ефектите што сме ги очекувале.

– Вилијам С. Кливленг (William S. Cleveland), од книгата „Визуелизирање на податоци“, издание на „Хобарт Прес“ (Visualizing Data, Hobart Press)

Податоците, затоа што се состојат од „битови“ и „бајти“ зачувани во датотека на тврдиот диск („hard drive“) на компјутерот, сами по себе се невидливи. За да можеме да ги видиме и сфатиме податоците, мораме да ги визуелизираме. Во ова поглавје, поимот визуелизација ќе го користам во едно пошироко значење што ќе ги вклучува и чисто текстуалните прикази на податоци. На пример, и внесувањето некоја група податоци во софтвер за работа со сметководствени табели може да се смета за форма на визуелизација на податоците. Невидливите податоци наеднаш се трансформираат во видлива „слика“ на нашиот екран. Значи, прашањето не е дали новинарите треба или не треба да ги визуелизираат податоците, туку кој вид визуелизација ќе се покаже како најкорисен во определена ситуација.

Со други зборови, прашањето е во кои случаи има смисла да одиме подалеку од визуелизацијата преку табела? Најкраткиот одговор би бил скоро секогаш. Табелите сами за себе не се доволни за да добиеме добар преглед на некоја група податоци. Табелите сами за себе не овозможуваат веднаш да ги откриеме постоечките

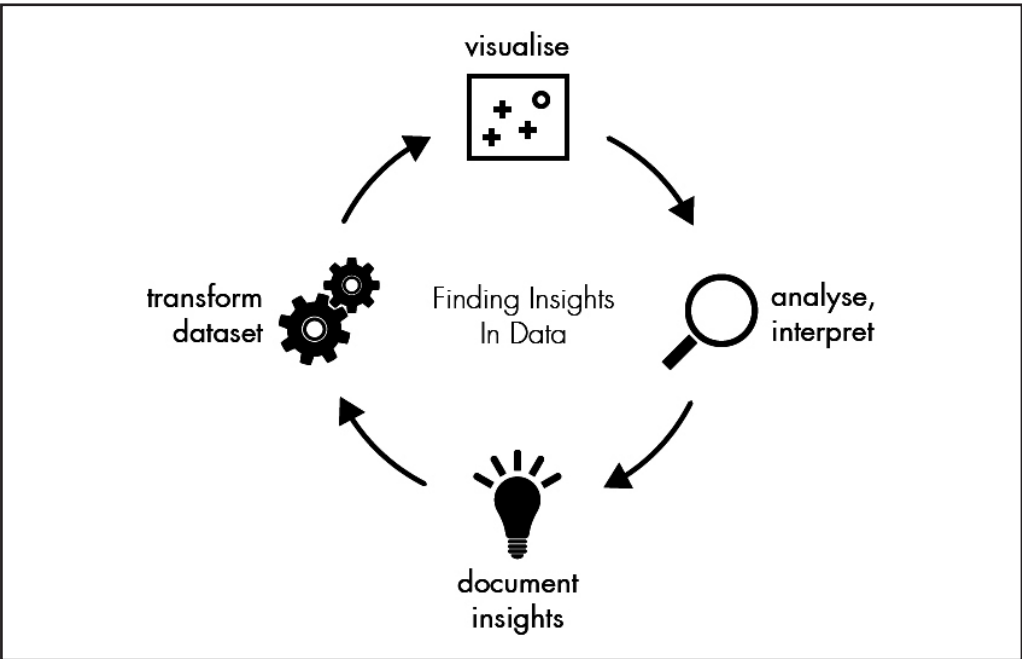
обрасци во податоците. Најчест пример за тоа се географските обрасци што можат да се набљудуваат само ако ги визуелизираме податоците на мапа. Но, како што ќе видиме подолу во ова поглавје, постојат и други видови обрасци.

Користење на визуелизацијата за нови сознанија

Очекувањето дека алатките и техниките за визуелизација ќе ослободат бран од готови приказни од податоците е нереално. Нема правила, нема некаков „протокол“ што би ни гарантирал дека ќе добиеме приказна. Сметам дека има повеќе смисла да бараме „сознанија“ кои добар новинар може вешто да ги вткае во сториите.

Секоја нова визуелизација веројатно ќе ни даде некакви нови сознанија за податоците со кои работиме. Некои од сознанијата ќе бидат познати од порано (иако можеби не се докажани), а други ќе бидат целосно нови или дури ќе претставуваат изненадување. Новите сознанија можат да го означат почетокот на една сторија, но некои од нив можат да бидат резултат од грешки во податоците што најлесно би биле откриени преку визуелизација на податоците.

Мислам дека процесот за кој се зборува на Слика 5-4 (и понатаму во ова поглавје) може да биде од голема помош во зголемувањето на успехот на процесот на откривање сознанија во податоците.

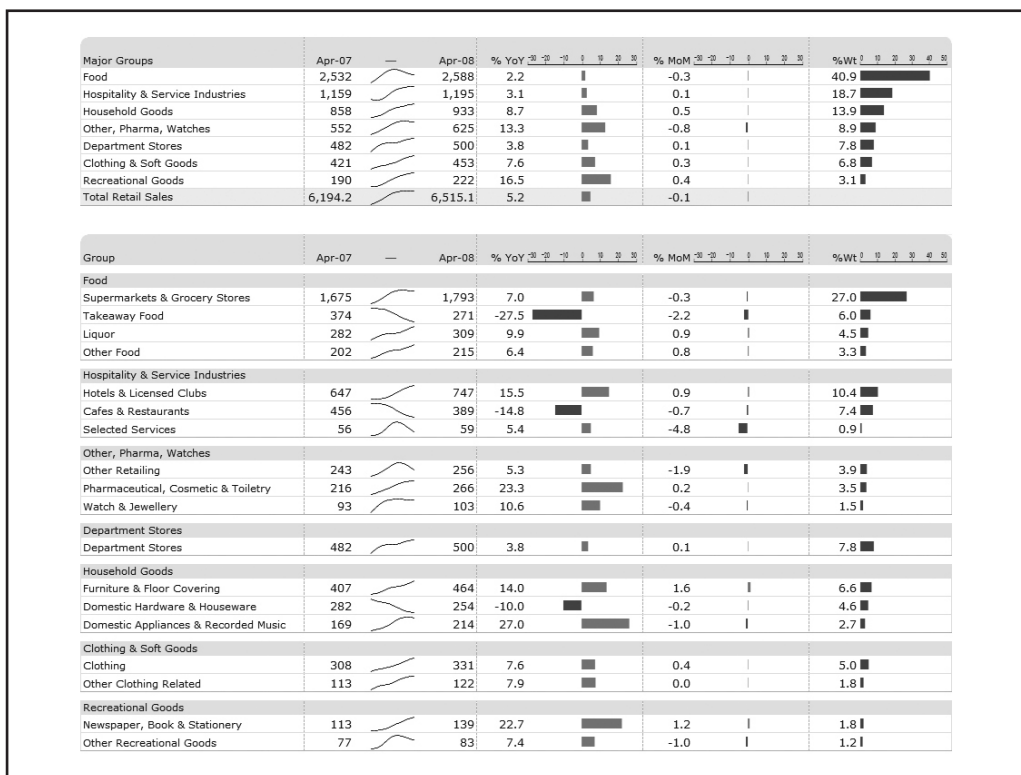


Слика 5-4. Увид во податоците: визуелизација (Грегор Ајш)

Научете како да ги визуелизирате податоците

Визуелизацијата нуди единствен поглед на податоците. Податоците можат да се визуелизираат на многу различни начини.

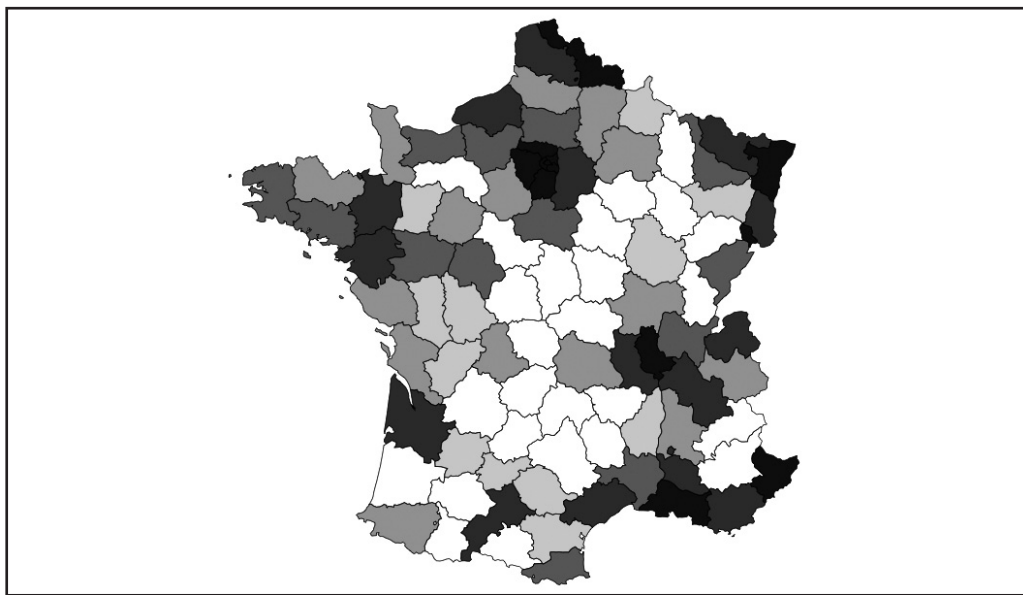
Табелите се добри за визуелизација кога работите со релативно мал број статистички точки. На нив се покажани ознаките и износите на сосема структуриран и организиран начин и тие го откриваат својот целосен потенцијал ако ги комбинираме со способноста за сортирање и филтрирање на податоците. Дополнително, Едварт Тафти (Edward Tufte) сугерираше дека внатре во колоните на табелите треба да се додадат мали графикони – на пример, една хоризонтална линија за секој ред или мал линиски графикон [познат и како „светкава линија“ (sparkline)]. Сепак, како што веќе спомнавме, табелите очигледно си носат и определени ограничувања. Тие, секако, се одличен начин за прикажување на еднодимензионалните координати како што се „најдобрите 10“ (во што било), но не се толку успешни во споредбата на повеќе димензии во исто време (на пример, популациските движења по земји во различни точки во времето).



Слика 5-5. Едварт Тафти советиува: „сџарклајнс“ (Греџор Ајш)

Графиконите, генерално, дозволуваат мапирање на димензиите во податоците според визуелните својства на геометриските фигури. Многу е напишано за ефективностa на индивидуалните визуелни својства, а скратената верзија би гласела: бојата е тешка, позицијата е најзначајна. Кај распределените дијаграми, двете димензии се мапираат според позицијата на оските „x“ и „y“. Можете да прикажете дури и трета димензија за да ја определите бојата или големината на прикажаните симболи. Линиските дијаграми се особено погодни за прикажување еволуција во времето, додека столбестите дијаграми се совршени за споредба на категорични податоци. Елементите на дијаграмот можете да ги наредите еден врз друг. Ако сакате да направите споредба меѓу мал број групи во вашите податоци, едно одлично решение е да прикажете повеќе единечни случаи во еден дијаграм. Во сите дијаграми можете да користите различни методи на степенување за проучување различни аспекти на вашите податоци (на пример, линеарна или логаритамска скала).

Всушност, повеќето податоци со кои работиме на еден или на друг начин се однесуваат на вистински, опипливи луѓе. Моќта на мапите лежи во способноста повторно да ги поврзат податоците со материјалниот свет околу нас. Замислете си дека имате група податоци во која се наведени географските локации на кои се извршени кривични дела. Разбирливо е дека сакате да видите каде се извршени кривичните дела. Мапите можат да ги откријат и географските врски помеѓу податоците (на пример, движењето на трендот од север кон југ, или од урбаните кон руралните средини).



Слика 5-6. Пример за „хоройлей“ (Греѓор Ајш)

Кога зборуваме за меѓусебните односи, дијаграмите се четвртиот најзначаен вид визуелизација. Дијаграмите ги прикажуваат меѓусебните врски (работи) на вашите статистички точки (јазли). Позицијата на јазлите се пресметува со помалку или повеќе сложени алгоритми за определување на изгледот на дијаграмот, а тоа ни овозможува веднаш да ги видиме структурите што постојат внатре во мрежата. Кај визуелизацијата преку дијаграми, трикот е да се најде вистинскиот и најсоодветен начин за моделирање на мрежата. Сите групи податоци немаат однапред дефинирани односи, а дури и ако ги имаат, тоа можеби нема да биде најинтересното нешто поврзано со нив. Понекогаш му останува на новинарот да ги дефинира рабовите помеѓу јазлите. Еден совршен пример за гореопишаното е Дијаграмот на социјална мрежа на Сенатот на САД (<http://slate.me/senate-social>), каде рабовите ги поврзуваат сенаторите што гласале на ист начин во повеќе од 65 отсто од сите гласања.

Анализирајте го и толкувајте го тоа што го гледате

Откако сте ги визуелизирале податоците, следниот чекор е да научите нешто од сликата што ја создадовте. Можете да се прашате:

- Што гледам на оваа слика? Дали тоа е она што го очекував?
- Има ли некои интересни обрасци?
- Што значи ова во контекст на податоците што ги имам?

Понекогаш ќе добиете визуелизација која, и покрај тоа што изгледа пре-красно, имате впечаток дека не ви кажува ништо интересно за податоците. Но, скоро секогаш можете нешто да научите од која било визуелизација, независно за колку банално или тривијално нешто се работи.

Документирајте ги вашите сознанија и чекори

Ако на процесот му пристапите како да се работи за патување низ податоците, документацијата ќе биде вашиот патопис. Таа ќе ви покаже каде сте патувале, што сте виделе таму и како сте ги донеле одлуките за тоа кои ќе бидат вашите следни чекори. Водењето на документацијата можете да го започнете и пред воопшто да фрлите поглед на податоците.

Во најголемиот број случаи, кога почнуваме со работа на група податоци што дотогаш не сме ја виделе, веќе имаме некакви очекувања и претпоставки поврзани со тие податоци. Обично постои некаква причина поради која сме заинтересирани за таа група податоци. Не е лошо работата на документирањето да ја започнете со запишување на почетните очекувања и претпоставки. Тоа помага во идентификувањето на предрасудите и ја намалува опасноста од погрешно толкување на податоците со тоа што ќе го откриеме само она што изворно сме сакале да го пронајдеме.

Навистина сметам дека документацијата е најзначајниот чекор во целиот процес. Исто така, тоа е чекорот за кој постои најголема веројатност да биде прескокнат и заборавен. Како што ќе видиме од примерот наведен подолу, опишаниот процес вклучува многу внесување координати и рачна обработка на податоци. Разгледувањето на комплетот од 15 дијаграми и графикони што сте го подготвиле може да се покаже како збунувачко искуство, особено ако од изработката поминало подолго време. Всушност, тие дијаграми и графикони имаат вредност (за вас или за кое било друго лице на кое сакате да му ги претставите наодите) само ако ги прикажете во контекстот во кој биле создадени. Оттаму, би требало да пронајдете време и да водите белешки за следниве нешта:

- Зошто го изработив овој дијаграм?
- Што сторив со податоците за да можам да го создадам дијаграмот?
- Што ми кажува тој дијаграм?

Трансформација на податоците

Откако ги собравте сознанијата од последната визуелизација, нормално е да имате некаква идеја што е следното нешто што сакате да го видите. Можеби сте откриле некој интересен образец или шема во податоците и сега сакате подетално да ги проучите.

Можни се следните трансформации:

Зумирање

Разгледување определен детаљ од визуелизацијата

Агрегација

Комбинирање повеќе статистички точки во една единствена група

Филтрирање

(Привремено) отстранување на статистичките точки што не се главен предмет на интерес

Отстранување на крајните вредности

Отстранување на референтните точки што не се репрезентативни за 99 отсто од податоците во групата.

Да замислиме дека сте визуелизирале еден дијаграм, а резултатот е збрка од јазли меѓусебно поврзани со стотици рабови (што е чест резултат кај визуелизирањето на таканаречените густо поврзани мрежи). Вообичаената трансформација би била да исфрлиме некои од рабовите со филтрирање. Ако, на пример, рабовите го прикажуваат протокот на пари од државите донатори до државите примачи на донации, можеме да решиме да ги отстраниме сите текови за износи под определена граница.

Кои алатки да ги користиме

Не е лесно да се одговори на прашањето кои алатки се најдобри. Секоја од алатките за визуелизирање податоци си има свои добри страни. Визуелизацијата и рачното внесување на податоците треба да бидат лесни и евтини. Ако промената на параметрите на визуелизациите би ви одзела неколку часови, малку е веројатно дека ќе сакате да експериментирате. Тоа, од друга страна, не значи дека нема потреба да научите како да ја користите таа алатка. Но, кога еднаш ќе научите како се употребува, алатката би требало да биде навистина ефикасна.

Често изгледа разумно да избереме алатка што истовремено ги опфаќа и прашањето на рачно внесување и прашањето на визуелизација на податоците. Одвојувањето на тие две задачи на две различни алатки значи дека ќе морате често да ги импортирате и експортирате податоците. Следи краток список на алатки за визуелизација и рачно внесување податоци:

- Сметководствени табели како што се „Либреофис“ (LibreOffice), „Excel“ или „Google Docs“
- Рамки за статистичко програмирање како што се „R“ (<http://r-project.org>) или „Пандас“ (Pandas, <http://pandas.pydata.org>)
- Географски информатички системи [Geographic Information Systems (GIS)] како што се „Квантум ГИС“ (Quantum GIS), „АркГИС“ (ArcGIS) или „ГРАСС“ (GRASS)
- Библиотеки со визуелизации како што се „d3.js“ (<http://mbostock.github.com/d3>), „Прифјуз“ (Prefuse, <http://prefuse.org>) или „Флеар“ (Flare, <http://flare.prefuse.org>)
- Алатки за рачно внесување податоци како што се „Гугл рифајн“ (Google Refine) или „Датаренглер“ (Datawrangler)
- Софтверски решенија за визуелизација кај кои не е потребно програмирање, како што се „Мениајс“ (ManyEyes) или „Табло паблик“ (Tableau Public, <http://tableausoftware.com/products/public>)

Примерите за визуелизации во следната секција се изработени со користење „R“, софтвер што се смета за некаков вид швајцарски армиски нож на полето на визуелизацијата на (научни) податоци.

Пример: Разбирање на податоците за донациите на изборните кампањи во САД

Да ја разгледаме базата на податоци за финансирање на кампањите на Претседателските избори во САД (US Presidential Campaign Finance Database), што содржи податоци за околу 450.000 донации за претседателските канди-

дати. Соодветната „CSV“-датотека е со големина од 60 мегабајти и е премногу голема за да се обработува во софтверска апликација како што е „Excel“.

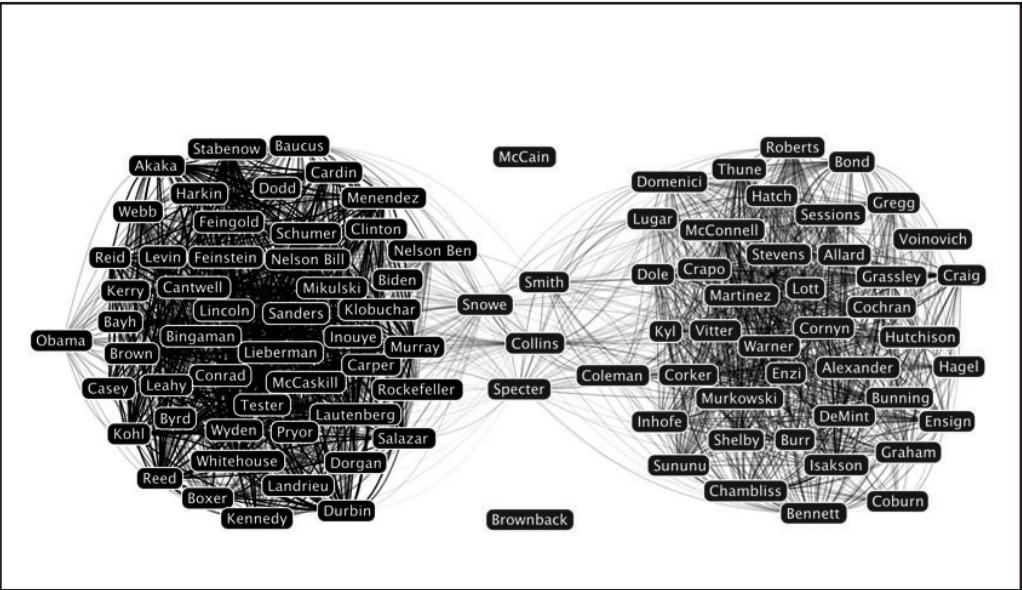
Како прв чекор, јасно ќе ги запишам моите почетни претпоставки за податоците во таа база на податоци:

- Обама ќе собере најмногу донации (тој е актуелен претседател и ужива најголема популарност меѓу кандидатите).
- Бројот на донации ќе се зголемува како што се приближува денот на изборите.
- Обама ќе добие многу повеќе помали донации од кандидатите на републиканците.

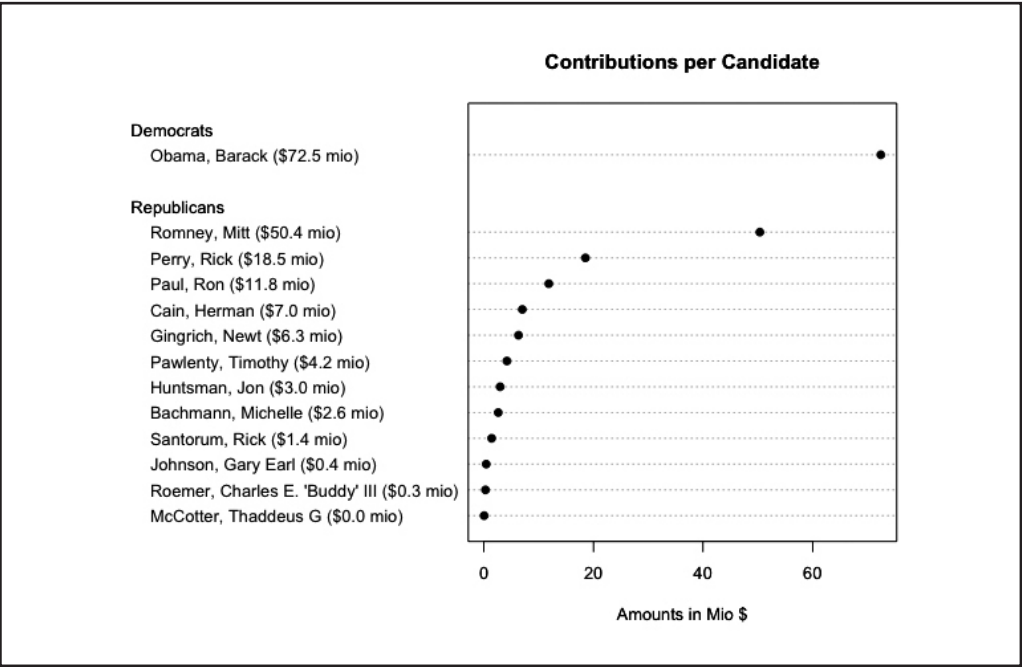
За да добиеме одговор на првото прашање, треба да ги трансформираме податоците. Наместо поединечни донации, потребни ни се збирните износи на донациите за секој од кандидатите. Со визуелизација на резултатите, подредени во табела од најголемиот кон најмалиот износ, можеме да ја потврдиме претпоставката дека Обама ќе собере најмногу пари:

Кандидат	Сума (во долари)
Обама, Барак (Obama, Barack)	72.453.620,39
Ромни, Мит (Romney, Mitt)	50.372.334,87
Пери, Рик (Perry, Rick)	18.529.490,47
Пол, Рон (Paul, Ron)	11.844.361,96
Кејн, Херман (Cain, Herman)	7.010.445,99
Гингрич, Њут (Gingrich, Newt)	6.311.193,03
Поленти, Тимоти (Pawlenty, Timothy)	4.202.769,03
Хантсмен, Џон (Huntsman, Jon)	2.955.726,98
Бахман, Мишел (Bachmann, Michelle)	2.607.916,06
Санторум, Рик (Santorum, Rick)	1.413.552,45
Џонсон, Гери Ерл (Johnson, Gary Earl)	413.276,89
Ромер, Чарлс Е. (Roemer, Charles E. Buddy III)	291.218,80
Мекотер, Тадеус Г. (McCotter, Thaddeus G)	37.030,00

Иако табелата ни ги покажува минималниот и максималниот износ и поредокот на кандидатите, таа не нуди многу информации за обрасците што се кријат зад рангирањето на кандидатите. На Слика 5-7 имаме еден друг поглед на податоците, „расподелен дијаграм“ на кој можеме да видиме сè што е прикажано на табелата, плус обрасците во полето. На пример, дијаграмот ни овозможува веднаш да ја споредиме разликата помеѓу Обама и Ромни и помеѓу Ромни и Пери, без да мора да ги одземаме износите едни од други. (Забелешка: дијаграмот беше изработен со користење на софтверот „R“. Линковите до изворниот код се наведени на крајот од ова поглавје).



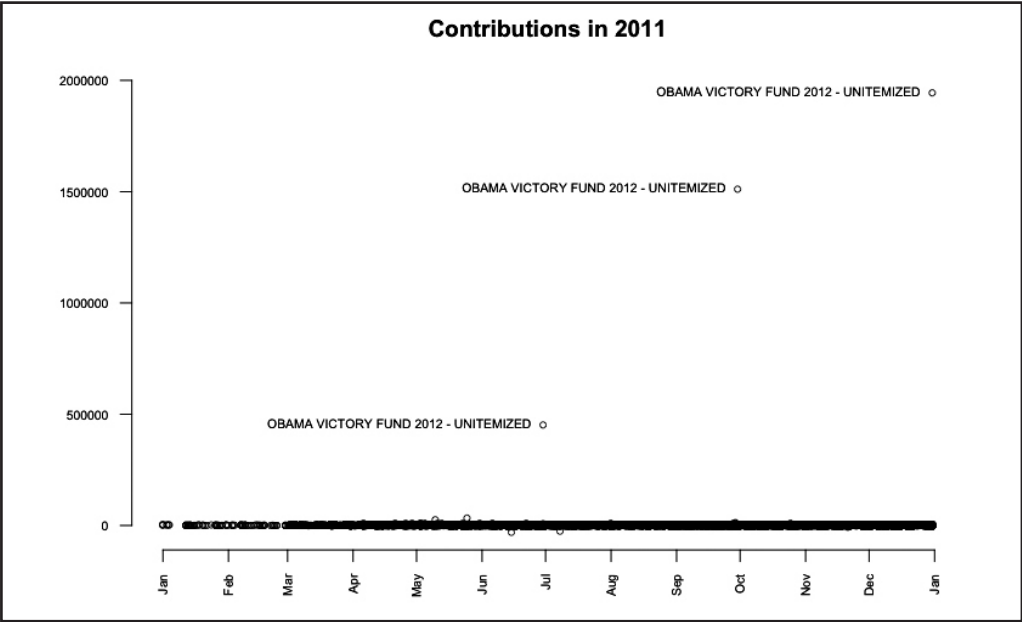
Слика 5-7. Визуелизациите како помош во откривањето на подјаточките образци (Греџор Ајш)



Слика 5-8. Три јасно видливи крајни вредности (Греџор Ајш)

Сега можеме да преминеме на пошироката слика што ја нудат податоците. Мојот прв чекор беше да ги визуелизирам сите износи на донациите, хронолошки подредени, во еден едноставен распределен дијаграм. Гледаме дека безмалку сите донации се сосема, сосема мали ако ги споредиме со трите навистина големи крајни вредности. Истрагата понатаму открива дека големите донации доаѓаат од „супер ПАК-от“ „Фондот за победа на Обама 2012“ (Obama Victory Fund 2012). Тие донации се направени на 29 јуни (450.000 долари), 29 септември (1,5 милиони долари) и 30 декември (1,9 милиони долари).

Иако донациите од Супер ПАК-овите сами за себе се, несомнено, најголемата приказна која се крие во податоците, може да биде интересно да погледнеме и малку подалеку. Поентата е во тоа што таквите големи донации го нарушуваат нашиот поглед на помалите донации што доаѓаат од индивидуални лица, па затоа ќе ги отстраниме од податоците. Таквата трансформација е позната како отстранување на крајните вредности. Кога ќе ја повториме визуелизацијата, ќе видиме дека најголемиот број донации се движат околу 10.000 долари или се помали од 5.000 долари.



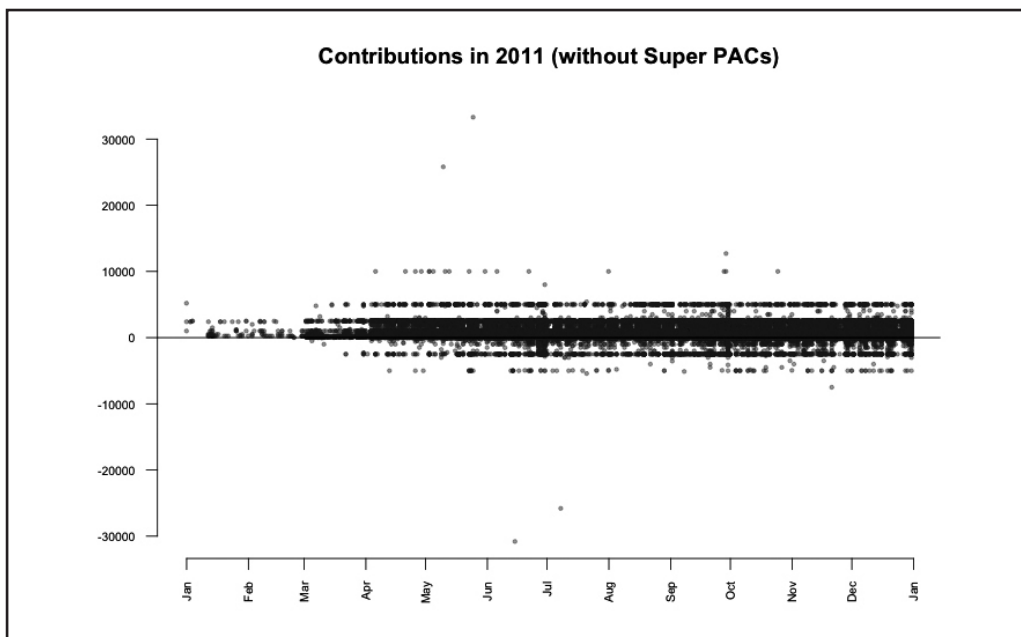
Слика 5-9. Оџсџранување на крајните вредности (Греџор Ајш)

Сојузниот закон за изборни кампањи (Federal Election Campaign Act – FECA) ги ограничува индивидуалните донации на најмногу од 2.500 долари за индивидуален кандидат. Како што ќе видиме во распределениот дијаграм, голем

број од донациите го надминуваат тоа ограничување. Две големи донации од мај 2012 година привлекуваат особено внимание. Изгледа дека си имаат свои пандани во негативните (рефундирани) суми во јуни и јули. Натамошното истражување ги открива следните трансакции:

- На 10 мај, Стивен Џејмс Дејвис (Stephen James Davis), од Сан Франциско, адвокат во фирмата „Банекер партнерс“ (Banneker Partners), донирал 25.800 долари за кампањата на Обама.
- На 25 мај, Синтија Марфи (Cynthia Murphy) од Литл Рок (Little Rock), вработена во фирмата за односи со јавноста „Марфи Груп“ (Murphy Group), донирала 33.000 долари за кампањата на Обама.
- На 15 јуни, на Синтија Марфи ѝ се рефундирани 30.800 долари, со што донацијата е намалена на 2.500 долари.
- На 8 јули, на Стивен Џејмс Дејвис му се рефундирани 25.800 долари, со што донацијата е намалена на 0 (нула) долари.

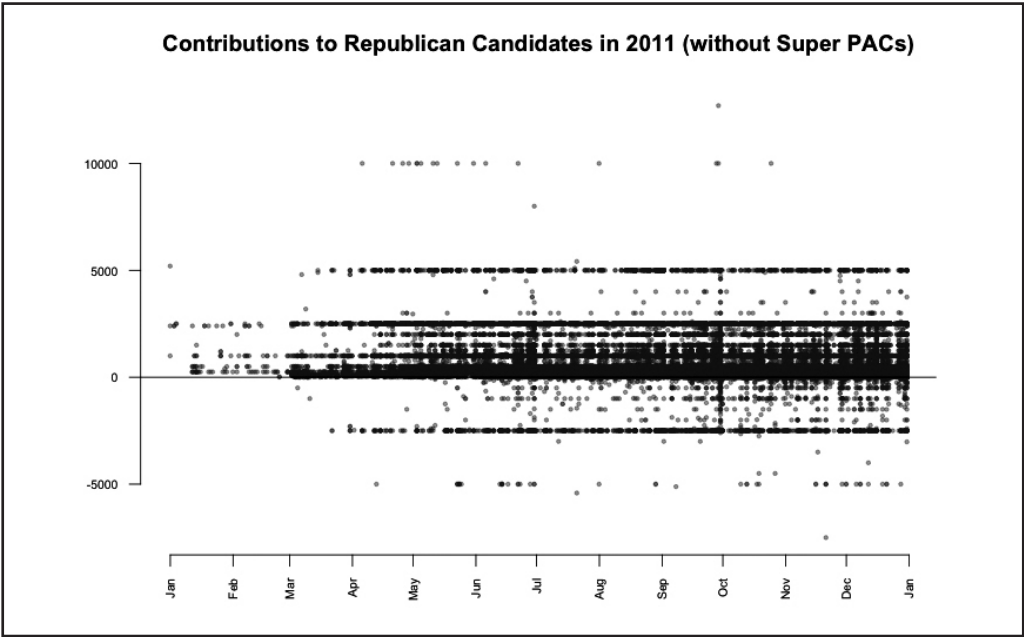
Што е интересно кај овие броеви? Сумата од 30.800 долари што ѝ била рефундирана на Синтија Марфи се совпаѓа со максималниот износ што индивидуални лица можат да го донираат на националните партиски комитети на годишно ниво. Таа можеби сакала едноставно да ги соедини двете донации во една трансакција, што било отфрлено. Износот од 25.800 долари што му биле рефундирани на Стивен Џејмс Дејвис може да се совпаѓа со разликата што ја добиваме ако од 30.800 одземеме 5.000 долари (ограничувањето за донации до некој друг политички комитет).



Слика 5-10. Осигурување на крајните вредности, виор џаџ (Греџор Ајш)

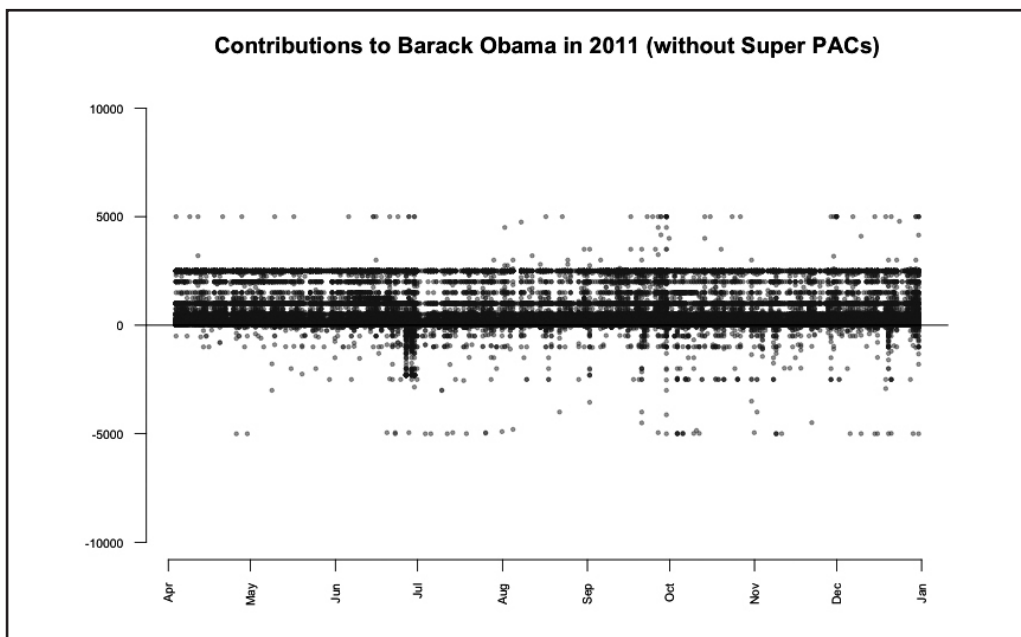
Едно друго интересно откритие во последната поставка на дијаграмот е образецот на донации за републиканските кандидати во износ од 5.000 долари и износи помали од 2.500 долари, што следи една хоризонтална линија. За подетално да ги разгледам, изработив визуелизација само на донациите за републиканците. Дијаграмот што го добив е одличен пример на обрасците и шемите што податоците ги следат, а кои би останале невидливи без соодветна визуелизација.

Можеме да видиме дека републиканските кандидати добиле голем број донации во износ од 5.000 долари. Всушност, ако ги провериме податоците, има 1.243 такви донации, што е само 0,3 отсто од вкупниот број донации. Сепак, затоа што правилно се распоредени низ времето, се појавува образец на хоризонтална линија. Еден интересен аспект на таа линија е фактот дека донациите од индивидуални лица се ограничени на максимум 2.500 долари. Следствено, сè над тој износ им беше вратено на донаторите, па како резултат ја имаме втората линија за износите под 2.500 долари. Ако ги споредиме со донациите за Барак Обама, ќе видиме дека таму не се појавува таков образец.

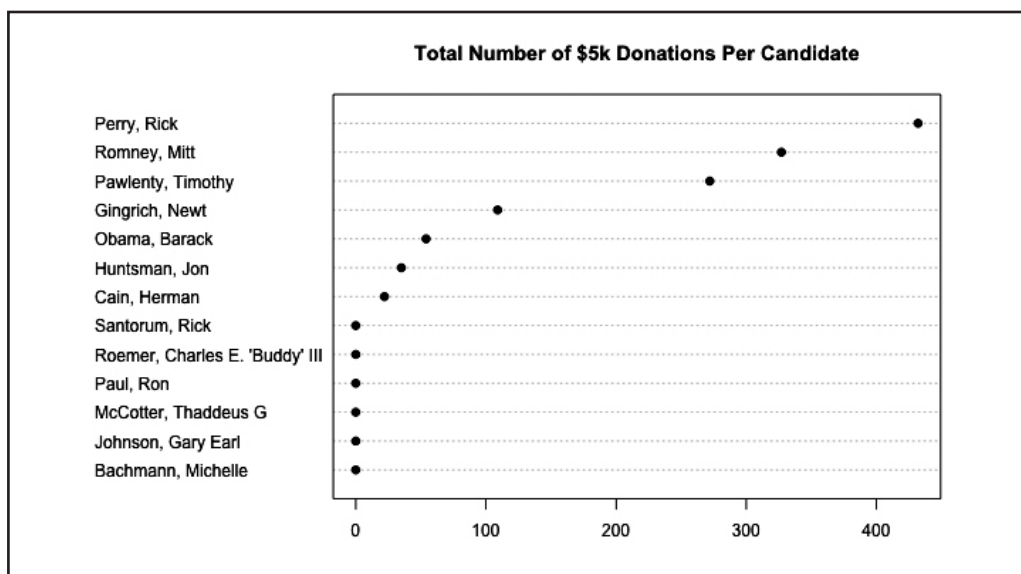


Слика 5-11. Оџсиџранување на крајниџе вредности, џреџ џаџ (Греџор Аџи)

Значи, можеби ќе биде интересно да видиме зошто илјадници републикански донатори не го забележале или почитувале ограничувањето за индивидуални лица. За уште подлабока анализа на тоа прашање, можеме да ги разгледаме вкупните бројки на донации од 5.000 долари за секој кандидат поединечно.



Слика 5-12. Донацииите по кандидату (Грегор Ајн)



Слика 5-13. Од каде доаѓаат парите што ги троши сенаторот?: донацииите по кандидату (Грегор Ајн)

Се разбира, таквиот поглед е донекаде искривен затоа што не ги зема предвид вкупните износи на донациите што ги добил секој од кандидатите. Следниот дијаграм прикажува колкав процент од вкупните донации за секој од кандидатите се донации во износ од 5.000 долари.

Што можеме да научиме од овој пример?

Често, таквата визуелна анализа на некоја нова група податоци наликува на возбудливо патување во непозната земја. Почнувате како странец, пред вас се само податоците и вашите очекувања, и со секој нов чекор, со секој нов дијаграм, стекнувате нови сознанија за темата на која работите. Тие сознанија ви служат како основа за одлуките за тоа кои ќе бидат вашите следни чекори и кои прашања заслужуваат дополнително истражување. Како што видовме во ова поглавје, процесот на визуелизација, анализа и трансформација на податоците може да се повторува во недоглед.

Преземете го изворниот код

Сите дијаграми прикажани во ова поглавје беа изработени со користење на извонредно моќниот софтверски пакет „R“. Тешко е да се пронајде метод на визуелизација или рачна обработка на податоци, а тој веќе да не е интегриран во „R“, софтвер наменет, пред сè, за визуелизација во научни цели. Оние што се интересираат за начините на визуелизација и обработка на податоците во „R“, можат да го преземат изворниот код за сите дијаграми на линковите што следат:

- распределен дијаграм: донации по кандидат (<https://gist.github.com/gka/1769733>)
- внесување координати: сите донации прикажани хронолошки (<https://gist.github.com/gka/1816161>)
- внесување координати: сите донации по овластени комитети (<https://gist.github.com/gka/1816169>)

На располагање ви стои и обемна литература на оваа тема.

— Греџор Ајш, Фондација „Ојен нолиџ“

Испорака на податоците



Откако ги разгледавте податоците и заклучивте дека определен интересен аспект заслужува да се пишува за него, на кој начин ќе ѝ го испорачате на јавноста? Го отвораме ова поглавје со кратки анегдоти во кои неколку водечки новинари во областа на дата-новинарството објаснуваат како тие им ги послужиле податоците на читателите – од инфографики, преку платформи за отворени податоци, до линкови што водат до веб-страници за спуштање датотеки. Потоа преминуваме на еден поширок преглед на процесот на изработка на новински апликации и на карактеристиките на визуелизациите на податоци. Најпосле, ќе разгледаме и што можете да направите за да обезбедите вашата публика да се ангажира во спроведувањето на проектот.

Претставување на податоците во јавноста

Постојат многу различни начини што можете да ги искористите за да ѝ ги прикажете податоците на јавноста – од објавување на суровите податоци проследени со стории, до изработка на прекрасни визуелизации и интерактивни веб-апликации. Прашавме неколку дата-новинари да ни кажат нешто повеќе за начините на кои тие ѝ ги презентираат податоците на јавноста.

Да се визуелизира или не?

Понекогаш податоците можат да ја раскажат приказната многу подобро од зборовите или фотографиите, и тоа е основната причина поради која, во последно време, во редакциите сè почесто се слушаат фрази како „новинска апликација“ или „визуелизација на податоци“. Друга работа што побудува интерес е богатата жетва од (често бесплатни) нови алатки и технологии наменети да им помогнат дури и на технички најмалку вештите новинари да ги трансформираат податоците во возбудлив пример на визуелно раскажување приказни.

Достапноста на алатките како „Гугл фјужн тејблс“ (Google Fusion Tables), „Мени ајс“ (Many Eyes), „Tableau“, „Дипити“ (Dipity) (<http://www.dipity.com/>) и други, значи дека денес е лесно да изработите мапи, дијаграми, графикони, па дури и комплетни апликации засновани на податоци, активности што порано беа резервирани за специјалистите во таа област. Денес влезните критериуми се сведени на најниско можно ниво, па прашањето што новинарите си го поставуваат не е дали можат да ги трансформираат податоците во визуелизација, туку дали воопшто треба тоа да го прават. Понекогаш е подобро да нема те никаква визуелизација, отколку да понудите лошо изведена визуелизација.

– Арон Пилхофер, Њујорк ѿајмс

Користење анимирани графички решенија

Со добро сценарио, синхронизирани анимации и јасни објаснувања, анимираната графика може да послужи во оживувањето сложени броеви или идеи и да ја води вашата публика низ приказната. Видео-предавањата на Ханс Рослинг (Hans Rosling) се добар пример како податоците можат да оживеат и да раскажат една приказна. Без оглед дали се согласувате со нивните методи, сметам дека „Индексот на фрлачи на чевли“ на „Економист“ (Economist's Shoe-throwers' Index, http://www.economist.com/blogs/dailychart/2011/02/daily_chart_arab_unrest_index) е добар пример како видеото може да се користи за раскажување приказна што се заснова на бројки. Графичките аспекти на индексот не би требало, или не би смеело да бидат прикажани како статична слика.

Едноставно, таму се случуваат толку многу нешта што не можат да бидат прикажани на тој начин. Но, откако ќе го изградите приказот, чекор по чекор, ќе стекнете

разбирање за тоа како и зошто нешто се нашло во индексот. Со користење анимирани графики и кратки филмови можете да го засилите тоа што вашата публика го слуша. Гласовната наратија, комбинирана со визуелни елементи што даваат добри објаснувања, е моќен и впечатлив начин на раскажување некоја приказна.

— Лулу Пини, независна дизајнерка на инфографики

Нека слушне цел свет

Нашата работа обично почнува со „Excel“. Тоа е лесен начин набрзина да утврдите дали податоците кријат нешто интересно. Ако чувствуваме дека тука има нешто, тогаш преминуваме во информативниот деск, а среќна околност е тоа што сме сместени веднаш до главниот информативен деск во „Гардијан“. Потоа размислуваме како да ги визуелизираме или да ги прикажеме на страниците на весникот. Го пишуваме текстот што ќе ја придружува визуелизацијата. Додека пишувам, обично на екранот имам отворено и скратена верзија на сметководствената табела. Често, додека пишувам, по малку и анализирам за да ги издвојам наинтересните поенти. Потоа го објавувам „постот“ и малку „твитувам“ за него, им пишувам на различни луѓе за да се осигурам дека до него ќе водат линкови од сите важни места.

Кај некои постови, барем половина од сообраќајот доаѓа преку „Твитер“ и „Фејсбук“. Горди сме на фактот што, во просек, посетителите на нашата веб-страница поминуваат 6 минути на една статија на „Датаблогот“, споредено со просечно една минута на останатите страници на веб-страницата на „Гардијан“. Шест минути е одлична бројка, а времето поминато на страницата е еден од клучните показатели за анализата на сообраќајот на нашата веб-страница.

Тоа ни помага и во процесот на убедување на колегите за вредноста на тоа што го работиме. Секако, заедно со големите приказни засновани на податоци што сме ги подготвиле и што ги знаат сите во редакцијата: „КОИНС“, „Викиликс“ и „Немиритите во Велика Британија“. Во случајот на податоците за трошоците од базата на податоци „КОИНС“, пет-шест специјализирани новинари од „Гардијан“ ги прегледаа податоците за да ни дадат свое мислење за податоците по нивното објавување од страна на Владата на Велика Британија. Еден друг тим од пет-шест новинари работеше со нас кога Владата ги објави податоците за трошоците од над 25.000 фунти, вклучувајќи и познати новинари како Поли Кертис (Polly Curtis). „Викиликс“, исто така, претставуваше голем зафат, со многу стории од војните во Ирак и Авганистан. И немиритите во Велика Британија постигнаа голем успех со над 550 илјади посети во период од само два дена.

Сепак, не се работи само за тоа да имате краткотраен хит: Работата е да станете веродостоен извор на корисни информации на кој може да му се верува. Се обидуваме да станеме место на кое ќе добиете добри и обмислени информации за темите што ги покриваме.

— Сајмон Роџерс, „Гардијан“

Објавување на податоците

Често, податоците ги вградуваме во нашата веб-страница како визуелизација и во формат што овозможува нивно лесно спуштање и снимање. Нашите читатели можат да ги прегледаат податоците на кои се засновани нашите стории преку интерактивната визуелизација или со користење на податоците на друг начин. Зошто е тоа важно? Тоа ја зајакнува транспарентноста на „Сиетл тајмс“. На читателите им ги нудиме истите податоци што сме ги користеле за носење цврсти заклучоци. А кој го користи? Се разбира, тука се оние што не критикуваат, како и оние што едноставно се интересираат за приказната и сите нејзини последици. Бидејќи нашите податоци се достапни за јавноста, можеме да ги искористиме советите од нашите критичари, а и од читателите воопшто, за она што можеби сме го пропуштиле или за тоа што уште би можеле да истражимо – мошне вредни совети кога се занимавате со сериозно новинарство.

– Черил Филијс, „Сиетл тајмс“

Давање пристап до податоците

Постојат неколку причини поради кои одлуката на потрошувачите на нашите вести да им дадеме пристап до податоците е вистински потег. Читателите можат да бидат сигурни дека не ги „злоупотребуваме“ податоците за да стигнеме до нефер заклучоци. Отворањето на нашите податоци ја следи традицијата на општествените науки за давање можност на истражувачите да ги реплицираат нашите постапки. Охрабрувањето на читателите да ги проучат податоците може да резултира со нивни совети што ќе не одведат до натамошни приказни. Најпосле, ангажираните читатели што се заинтересирани за податоците што ги користите и нудите, веројатно повторно ќе се враќаат на вашата веб-страница.

-- Стив Дојџ, Школа за новинарство „Волиер Кронкајџ“, Државен универзитет на Аризона

Основање платформа за отворени податоци

Во „Ла насион“, објавувањето отворени податоци е интегрален дел од активностите во областа на дата-новинарството. Аргентина нема закон за слобода на информациите, ниту портал за пристап до податоците на национално ниво, така што сме сосема сериозни во намерата да им ги понудиме на читателите сите податоци што ги користиме во известувањето.

Поради тоа, суровите структурирани податоци ги објавуваме преку нашата интегрирана платформа изработена на платформата „Јунар“ (Junar), како и преку „Google Spreadsheets“. Отворено овозможуваме и поттикнуваме користење на нашите податоци, а нудиме и кратки упатства, достапни во пишана форма и на видео (<http://bit.ly/lanacion-tutorials>).

Понатаму, определени податоци и визуелизации нудиме и преку нашиот „Даталог на Насион“ (Nacion Data Blog, <http://blogs.lanacion.com.ar/data/>). Тоа го правиме за да ја проповедаме вредноста на податоците и алатките за објавување податоци во Аргентина, да им покажеме на другите како ги собираме и користиме податоците и на другите да им покажеме како можат понатаму да ги користат податоците.

Од отворањето на платформата во февруари 2012 година, добивме многу сугестии и идеи за собирање податоци. Повеќето идеи доаѓаат од академската и истражувачката заедница, како и од универзитетски студенти што ни благодарат секогаш кога ќе одговориме со решение или со определена група податоци. Луѓето, исто така, ги користат нашите податоци на „Табло“ (Tableau), даваат коментари за нив, а неколку пати нашите предмети беа најмногу посетувани и коментирани на тој сервис. Во 2011, седум од стотите најпосетувани визуелизации беа наши.

– Анхелика Пералта Рамос, „Ла насион“ (Аргентина)

Давање човечки лик на податоците

Упадливо е отсуството на човечкиот фактор во дискусијата за „големите“ податоци што се шири во јавната свест. Иако многумина гледаат на податоците како на неврзани бројки во просторот околу нас, сепак се работи за мерки на сосема опипливи нешта што често се поврзани со луѓе. Податоците се прикачени за животите на вистински живи луѓе и, кога работиме со бројки, мораме да ги имаме предвид системите во реалниот свет од кои доаѓаат.

Да ги разгледаме, на пример, локациските податоци што, додека ова го пишуваме, се собираат преку стотици милиони телефони и други мобилни уреди. Лесно е за таквите податоци (броеви што ги бележат географската ширина и должина и точното време) да размислуваме како за „дигитални нус-производи“, а сепак се работи за дестилирани моменти од нашите лични приказни. Иако изгледаат сувопарни и неутрални кога ги набљудувате во форма на сметководствена табела, ако им дозволите на луѓето да ги репродуцираат на мапа и да го следат своето движење, тие ќе искушат повторно преживување на сеќавањата, што истовремено е моќно и, во крајна линија, мошне човечко чувство.

Во моментот, локациските податоци се користат од многу трети страни – од девелоперите на апликации, од големите брендови, од огласувачите. Иако втората страна (телекомуникациските компании и субјектите што управуваат со уредите) ги поседува и чува податоците, првата страна во тоа равенство – вие лично – немате никаков пристап или контрола врз таквите информации. Одделението за истражување и развој на „Њујорк тајмс“ (NYTimes R&D Group) почна со реализацијата на проект на еден прототип под името „Опенпатс“ (OpenPaths, „отворени патеки“, <https://openpaths.cc>), што на јавноста, односно на граѓаните им помага да ги проверат сопствените локациски податоци, но

и да го искушат концептот на сопственост над податоците. Најпосле, луѓето би требало да имаат контрола врз тие броеви, толку тесно врзани за нивните животи и искуства.

Новинарите имаат значајна улога во изразувањето на таквата вродена човечка природа на податоците. Тоа им дава моќ да го променат начинот на кој јавноста гледа на нештата – на податоците и на системите од кои излегуваат бројките.

– Џер Торп, резиденџен „дајта-уметник“, Одделение за истражување и развој на „Њујорк тајмс“ (Jer Thorp, Data Artist in Residence: New York Times R&D Group)

Отворени податоци, отворен изворен код, отворени вести

2012 може да биде и година на отворените вести. Таа идеја е во суштината на нашата уредувачка идеологија и е клучната порака на начинот на кој во моментов сме брендирани. Јасно е дека ни треба отворен процес што ќе одговара на развојот на новинарството придвижувано од податоци. Не смееме да оставиме тој процес да го туркаат напред само отворените податоци. Мора да имаме отворени алатки за спроведување на процесот. Се надеваме дека, до крајот на годината, ќе можеме, кон секоја објавена визуелизација, да ги додадеме и податоците на кои се заснова и компјутерскиот код што ја придвижува.

Многу од алатките што денес ги користиме за изработка на визуелизации се со затворен изворен код. Ни се нудат визуелизации со рестриктивни лиценци што го забрануваат користењето на изведените податоци. Постоечките библиотеки со отворен изворен код често решаваат индивидуални проблеми, но не успеваат да понудат поширока методологија. Сето тоа ја отежнува надградбата на трудот на други лица. Тоа му става крај наместо да го поттикне разговорот. За таа цел, во рамките на проектот „Мисо“ (Miso Project, @themisoproject на „Твитер“) развиваме еден комплет отворени алатки за интерактивно раскажување приказни.

За проектот преговараме со неколку медиумски организации. Разбирањето на целосниот потенцијал на софтверот со отворен изворен код бара ангажман на целата заедница. Ако успееме во намерата, тоа ќе внесе сосема нова динамика во односите со читателите. Нивниот придонес ќе оди многу подалеку од можноста да даваат коментари на нашиот труд – ќе можат да ја анализираат работата што сме ја завршиле, да решаваат проблеми и да ги користат податоците на нови, неочекувани начини.

– Алистер Денџ, „Гардијан“

Додадете линк за снимање на податоците

Во последниве неколку години сигурно сум обработил неколку гигабајти податоци во спроведувањето проекти и изработката на статии, од скенови на печатени табели од 1960-тите до дипломатските телеграми објавени од „Вики-

ликс“ со вкупна големина од 1,5 гигабајти. Отсекогаш било проблем како да ги убедиме уредниците дека треба систематски да ги објавуваме изворните податоци во отворен и лесно достапен формат. За да го премостам тој проблем, додавав линкови „Сними ги податоците“ (Download the Data) во статиите, но сејќи ги луѓето до архивите каде што се чуваат датотеките или до релевантните документи на „Google Docs“. Интересот за повторно користење беше на нивото што го гледаме кај сличните програми под покровителство на владата (т.е. мал, многу, многу мал). Сепак, малкуте примери на повторна употреба обезбедија нови сознанија и поттикнаа расправа, така што неколкуте минути дополнителна работа на секој проект не се залудни!

– *Николас Кајзер-Брил, Journalism++*

Познавајте ги своите можности

Познавајте го досегот на вашите можности. Огромна е разликата помеѓу хакирањето „колку да ви мине времето“ и инженерингот што е потребен за да се покрие определен опсег или да се обезбедат некакви перформанси. Осигурајте се дека партнерите со кои сте се здружиле ги поседуваат вештините потребни за спроведување на проектот. Никогаш не запоставувајте го дизајнот. Лесното користење, корисничките искуства и дизајнот на презентацијата можат да имаат огромно влијание врз успехот на вашиот проект.

– *Крис Ву, „Хекс/Хакерс“*

Како да изработите новинска апликација

Новинските апликации се прозорци за податоците што стојат зад приказната. Може да се работи за бази на податоци што овозможуваат пребарување, блескави и мазни визуелизации или нешто сосема друго. Без оглед на нивната форма, новинските апликации ги поттикнуваат читателите да работат со податоците во оној контекст што им значи: да ги погледаат движењата на криминалот во нивното соседство, колку е добра локалната болница или амбулантата, да проверат кој дал донација за кампањата на кандидатот на изборите за кој сакаат да гласаат.

Најдобрите апликации не се само од инфографики што користат висока технологија. Тие имаат долг рок на траење. Тоа што им помагаат на читателите во решавањето на нивните реални проблеми, или што даваат иновативни и корисни одговори, им помага да опстојуваат надвор од циклусот на вести како трајни ресурси. Кога новинарите на „ПроПублика“ сакаа да ја истражат безбедноста на американските клиници што нудат дијализа за бубрежно-болните, тие изработија апликација (<http://projects.propublica.org/dialysis/>) што им помагаше на корисниците да проверат колку е безбедна клиниката во градот

во кој живеат. Обезбедувањето таква значајна и релевантна услуга помага во воспоставувањето однос со корисниците кој далеку ги надминува можностите на наративната приказна да го стори истото.

Токму таму, во можноста да се создаде нешто што има трајна вредност, се кријат и предизвиците и потенцијалите на подготовката на најнапредните новински апликации. Без разлика дали сте девелопер или менаџер, дискусијата околу изработката на добра новинска апликација мора да го користи пристапот на развивање нови производи. Значи, останете фокусирани на корисниците и осигурете се дека ќе добиете најмногу што можете за вложените пари. Значи, пред да почнете со работата, добро би било да си поставите три прашања. Тие три прашања ќе ги разгледаме на следните неколку страници.

PRO PUBLICA Journalism in the public interest.

Receive our top stories daily
Subscribe via email **SIGN UP**
zip-code optional

Home Our Investigations Tools & Data MuckReads About Us

Search ProPublica

Dialysis
The High Costs and Hidden Perils of a Treatment Guaranteed to All

Dialysis Facility Tracker Updated Dec. 22, 2010

By Robin Fields, Al Shaw, and Jennifer LaPleur, ProPublica, Dec. 22, 2010

This site is for dialysis patients and others who want to learn about the quality of care at individual dialysis clinics. Among other things, you can learn how often patients treated at a facility have been hospitalized, report certain types of infections or are placed on the transplant list. The information is submitted by facilities and collected by contractors of the Centers for Medicare and Medicaid Services, the federal agency that oversees most dialysis care.

Related story: Dialysis Data, Once Confidential, Shines Light on Clinic Disparities »

Find a facility near you

Address, ZIP, or facility name within 10 mi. **SEARCH**

For example, 1606 E. 55th St. Chicago, IL 60615, 77054, or UCLA

Safeguard the public interest.
Support ProPublica's award-winning investigative journalism.
DONATE

Get Updates
Sign up here to be alerted when we update our dialysis facilities database.

Our Stories

Facilities in Your State

Name	Facilities
Alabama	121
Alaska	8
Arizona	103
Arkansas	64
California	484

Слика 6-1. Апликација за следење на установите што нудат дијализа (ПроПублика)

Која е мојата публика и какви се нејзините потреби?

Новинските апликации не треба да ѝ служат само на приказната, туку и на корисникот. Во зависност од тоа за каков проект се работи, корисникот може да е пациент што сака да знае колку е безбедна болницата во која прима дијализа, или некој што сака да знае колкава е опасноста од земјотрес во регионот каде што се наоѓа неговиот дом. Независно од тоа за кого се работи, изработката на новинска апликација, како и секој друг добар производ, треба да почне со разговор за луѓето што ќе ја употребуваат.

Една апликација може да им послужи на голем број корисници. На пример, проектот „Курбвајз“ (Curbwise) на „Омаха ворлд-хералд“ (Omaha World-Herald, дневен весник од Омаха, Небраска) им помага на сопствениците на куќи што сметаат дека плаќаат превисоки даноци; на локалните жители што се заинтересирани за цените на недвижностите во регионот; и на агентите за недвижности што се обидуваат да го следат пазарот на недвижности. Во секој од тие случаи, апликацијата одговара на определена потреба, поради што корисниците постојано ѝ се навраќаат.

Сопствениците на куќи, на пример, можат да побараат помош во собирањето информации за недвижностите во непосредната околина, за да ги аргументираат своите тврдења дека даноците што ги плаќаат се неоправдано високи. Собирањето такви информации е комплицирано и одзема многу време, проблем што „Курбвајз“ го решава преку подготовка на извештај, лесен за користење, што ги содржи сите информации што се потребни за оспорување на висината на данокот на имот пред локалните власти. „Курбвајз“ таквите извештаи ги продава за 20 долари, цена што луѓето се подготвени да ја платат затоа што им помагаат во решавањето реални проблеми.

Без оглед дали вашата апликација треба да нуди решенија за реални проблеми, како што прави „Курбвајз“, или да ја дополни сторијата со интересна визуелизација, секогаш имајте ги на ум луѓето што ќе ја користат. Фокусирајте се на дизајнот и вградете функции што ќе одговараат на нивните потреби.

Колку време да му посветам на проектот?

Девелоперите во редакциите се како вода во пустина: многу се барани, а ги има малку. Изработката на новински апликации носи потреба за балансирање помеѓу секојдневните потреби на редакцијата и долгорочната посветеност што е потребна за изработка на вистински извонредни производи.

Да замислиме дека вашиот уредник ви пристапил со следната идеја: градскиот совет следната недела треба да гласа за предлогот да се урнат неколку историски градби во вашиот град. Ќе предложи изработка на една едноставна апликација со мапа на која се прикажани зградите.

Како девелопер, имате неколку можности. Можете да искористите сè што знаете и да изработите убава мапа со користење наменски софтвер. Или можете да ги искористите постоечките алатки, „Google Fusion Tables“ или библиотеците за мапирање со отворен изворен код, и да ја завршите работата за неколку часа. Ако ја изберете првата можност, ќе добиете подобра апликација. Но, втората опција може да ви остави повеќе време да изградите нешто друго што ќе има поголем капацитет за трајно влијание.

Не мора да градите комплексни новински апликации што изгледаат прекрасно само затоа што приказната го дозволува тоа. Клучно е да ја одржувате

рамнотежата помеѓу приоритетите. Она што треба да го запомнете е дека секоја апликација е поврзана со определени трошоци: имено, потрошеното време сте можеле да го искористите за работа на друга, потенцијално повлијателна апликација.

Како да достигнам ново, повисоко ниво?

Конструирањето висококвалитетни апликации може да биде скапо и да ви одземе многу време. Токму поради тоа, секогаш се исплати да прашате што ќе добиете за возврат. Како еднократниот хит да прерасне во нешто навистина посебно?

Еден пристап е да осмислите долготраен проект што оди подалеку од вообичаениот циклус на вестите. Тоа, од друга страна, можете да го постигнете и со изработка на алатка (ако е со отворен изворен код, дотолку подобро!) што ќе ви заштеди време во некој подоцнежен стадиум од работата, или со вградување напредни аналитички постапки во вашата алатка, што ќе ви овозможи да дознаете повеќе за вашата публика.

Многу организации подготвуваат „пописни мапи“ (Census maps) за да ги илустрираат демографските промени во нивните градови. Но, кога тимот за новински апликации на „Чикаго трибјун“ ја изработи својата пописна мапа (<http://bit.ly/chicago-census>), ја подигнаа целата работа на повисоко ниво со тоа што развија алатки и техники за побрза изработка на такви мапи, и сите алатки и техники ги направија достапни за да можат да ги користат и други организации.

Во Центарот за истражувачко новинарство, каде што работам, споивме неколку бази на податоци со можност за пребарување со чувствителна рамка за следење настани, и тоа ни овозможи да дознаеме, меѓу другото, и колку корисниците ги ценат инспирацијата и можноста за истражување вградени во нашите новински апликации.

Иако звучи ситничаво, секогаш размислувајте за исплатливоста на инвестицијата. Решете некој генерички проблем; измислете нов начин за ангажирање на корисниците; отворете го изворниот код за делови од тоа што сте го изработиле; користете аналитички алатки за да дознаете повеќе за своите корисници; пронајдете ги случаите во кои апликацијата може да ви донесе приходи, слично на примерот на „Curbwise“.

Пакување

Новинските апликации постигнаа огромен напредок во краток временски период. Новинските апликации од првата генерација (News Apps 1.0) во голема мера личат на инфографиките од втората генерација (Infographisc 2.0) – интерактивни визуелизации помешани со бази на податоци што можат да се

пребаруваат, а нивната основна намена беше да го унапредат раскажувањето на приказните. Денес, многу такви апликации можат да бидат подготвени од новинари, што работат со кратки рокови, со користење алатки со отворен изворен код, така што на девелоперите им останува време да размислуваат за поголеми нешта.

Кај новинските апликации од втора генерација (News Apps 2.0) – а индустријата генерално се движи во таа насока – се работи за комбинирање на добрите страни на новинарството на полето на раскажувањето приказни и вршењето јавна функција, со дисциплината во развојот на производи и стручноста што ги поседува светот на технологијата. Несомнено, како резултат ќе имаме експлозија на иновативност во процесот податоците да ги направиме порелевантни, интересни и полезни за нашата публика, со надеж дека таа иновативност ќе му помогне и на новинарството да стане релевантно, интересно и полезно.

– Чејс Дејвис, Ценитар за истражувачко новинарство

Новинските апликации на „ПроПублика“

Новинската апликација е голема интерактивна база на податоци што раскажува некаква новинска историја. Гледајте на неа како на секое друго новинарско дело. Таа едноставно наместо зборови и слики користи софтвер.

Со тоа што на индивидуалните читатели им нуди податоци што имаат некакво значење за нив, новинската апликација може да им помогне да ја разберат приказната на начин што е крајно индивидуализиран. Таа може да им помогне на читателите да разберат како тие лично се поврзани со нешто што се случува во целата држава, да им помогне да го поврзат тоа што го знаат со тоа што не го знаат и на тој начин да придонесат за подобро разбирање на апстрактните концепти.

Ние вообичаено подготвуваме новински апликации кога имаме некакви податоци (или кога мислиме дека можеме да набавиме податоци) што, по својот опсег, се значајни на национално ниво, а сепак се доволно раслоени за да откријат значајни детали.

Новинската апликација треба да раскаже приказна и, како и на секоја добра вест, потребни ѝ се наслов, наднаслов, добар вовед и ударна вест. Понекогаш е тешко да ги препознаеме тие концепти во некоја интерактивна софтверска програма, но ако погледнеме повнимателно, тие секако се таму.

Исто така, новинската апликација треба да генерира нови студии и натамошно известување. Најдобрите апликации на „ПроПублика“ беа користени како основа за раскажување локални приказни.

Да ја земеме за пример апликацијата „Доллари за докторите“ (Dollars for Docs, <http://projects.propublica.org/docdollars/>). Таа беше прв обид да се следат милионските суми долари што фармацевтските компании им ги плаќаат на лекарите за консултантски услуги, за настапи како предавачи на нивни настани и конференции итн. Нашата апликација им овозможува на читателите да го проверат својот лекар и сите такви уплати на негово или нејзино име. Податоците беа користени и од новинари од други редакции. Повеќе од 125 локални медиумски организации, вклучувајќи ги и „Бостон глоуб“ (Boston Globe), „Чикаго трибјун“ и „Сент Луис пост-диспеч“ (St. Louis Post-Dispatch), подготвија истражувачки приказни за локалните доктори во тие градови, врз основа на податоците од „Доллари за докторите“.

Некои од тие приказни беа резултат на формално партнерство со тие медиуми, но повеќето беа спроведени независно – во некои случаи воопшто не знаевме дека се работи на приказната сè до нејзиното објавување. Таквите нешта се од клучно значење за нас како мала новинска организација која сепак работи на национално ниво. Ние самите не можеме да знаеме што се случува на локално ниво во 125 градови, но ако нашите податоци им помогнат на локалните новинари да раскажат приказни што ќе имаат вистинско влијание, нашата мисија е исполнета.

Една од моите омилен новински апликации е „Мапирање на Л. А.“ (Mapping L.A., <http://maps.latimes.com/neighborhoods/>) на „Лос Анжелес тајмс“ (Los Angeles Times), која започна како мапа на бројните маала на Лос Анжелес, подготвена со „краудсорсинг“, што до моментот на објавување на „Мапирање на Л. А.“ немаа општоприфатени граници. По почетниот „краудсорсинг“-проект, „Л. А. тајмс“ ги користеше маалата како рамки за пообемно дата-новинарство – стапката на криминал, квалитетот на училиштата во индивидуалните маала итн., нешто што претходно не можеа да го сторат. Значи „Мапирање на Л. А.“ не само што истовремено е широка и специфична, таа генерира нови приказни и ги раскажува приказните на индивидуалните лица.

Потребните ресурси за изработка на новинска апликација се разликуваат од еден случај до друг. „Њујорк тајмс“, на пример, има десетици вработени што работат на новински апликации и интерактивни графики. Од друга страна, совршената апликација за следење на политичките анкети на информативната веб-страница „Токинг поинтс мемо“ (Talking points memo, <http://talkingpointsmemo.com/>) беше произведена од само двајца вработени, при што ниту еден од нив немаше професионални квалификации од областа на компјутерските науки.

Како и повеќето редакциски програмери (кодери), и ние во изработката на апликации следиме една модифицирана верзија на методологијата „Еџајл“ (Agile methodology, <http://agilemethodology.org>). Работиме брзо и нацртите им ги покажуваме на останатите вработени во редакцијата. Уште поважно, соработуваме со новинарите и ги читаме нивните нацрти за статии, дури и

оние настанати на почетокот на процесот на пишување. Повеќе работиме како новинари отколку како традиционални програмери. Освен пишување компјутерски код, контактираме со изворите, собираме информации и ја подобруваме нашата стручност. Би било тешко да подготвиме добра новинска апликација од материјал што не го разбираме.

Постојат три причини поради кои редакциите треба да се заинтересирани за производство на новински апликации придвижувани од податоци. Тие се форма на добро новинарство, популарни се – новинските апликации се најпопуларните сегменти во понудата на „ПроПублика“, и ако ние не го правиме тоа, тогаш ќе го прави некој друг. Помислете на сите големи приказни што би ги пропуштите! Најважно од сè, редакциите би требало да знаат дека и тие можат да се занимаваат со новински апликации. Многу полесно е отколку што изгледа.

– Ској Клајн, ПроПублика

Визуелизацијата како мотор за дата-новинарството

Пред да се обидете да цртате графикон или мапа со вашите податоци, застанете за момент и размислете на колку начини статичните и интерактивните графички елементи имаат улога во новинарството што го произведувате.

Во фазата на известување, визуелизациите можат:

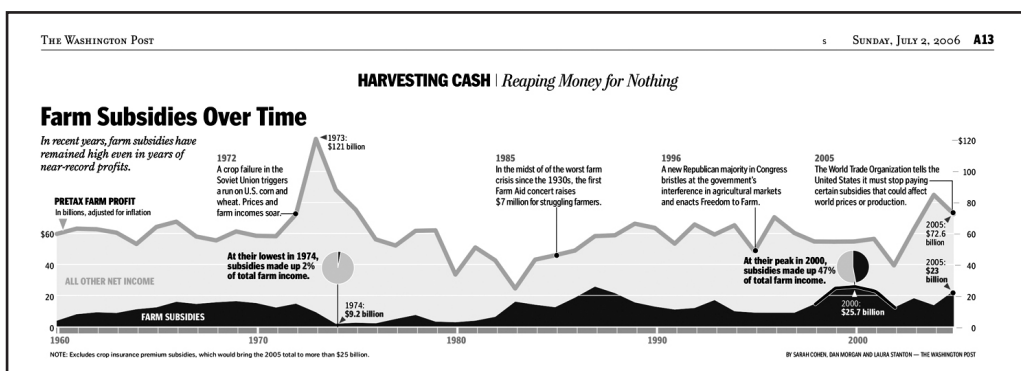
- да ви помогнат да ги идентификувате темите и прашањата за остатокот од известувачките активности
- да ги идентификуваат координатните точки: добрите приказни или можните грешки содржани во вашите податоци
- да ви помогнат да пронајдете типични примери
- да ви ги покажат „дупките“ во известувањето.

Визуелизациите играат повеќе улоги и во објавувањето. Тие можат:

- да ја илустрираат поентата на приказната на многу поубедлив начин
- да ги отстранат непотребните технички информации од текстот
- на вашите читатели да им обезбедат транспарентност на процесот на известување, особено ако се работи за интерактивни апликации што дозволуваат натамошно истражување.

Горенаведеното укажува дека треба да ја внесувате визуелизацијата во известувањето што порано и што почесто, без разлика дали ќе започнете да подготвувате електронски верзии од податоците или не. Не гледајте на визуелизацијата како на одвоен чекор, нешто за што ќе размислувате дури кога ќе ја завршите работата на сторијата. Дозволете ѝ таа да го води вашето известување.

Понекогаш добар почеток ќе биде и само да ги претставите забелешките што дотогаш сте ги направиле во визуелна форма. Разгледајте го графиконот на Слика 6-2, објавен во „Вашингтон пост“ (Washington Post) во 2006 година.



Слика 6-2. Хронолошки преглед на субвенциите за земјоделството („Вашингтон пост“)

Визуелизацијата, изградена во текот на неколку месеци, го прикажува оној дел од приходите од земјоделството што е поврзан со субвенциите и клучните настани во последните 45 години. Пронаоѓањето на податоците користени во подолг временски период, со слични дефиниции и значења, претставуваше огромен предизвик. Истражувањето на сите највисоки и најниски вредности не принуди постојано да размислуваме за контекстот на нашата приказна. Тоа, исто така, значеше дека една задача е во најголема мера завршена уште пред сториите да бидат напишани.

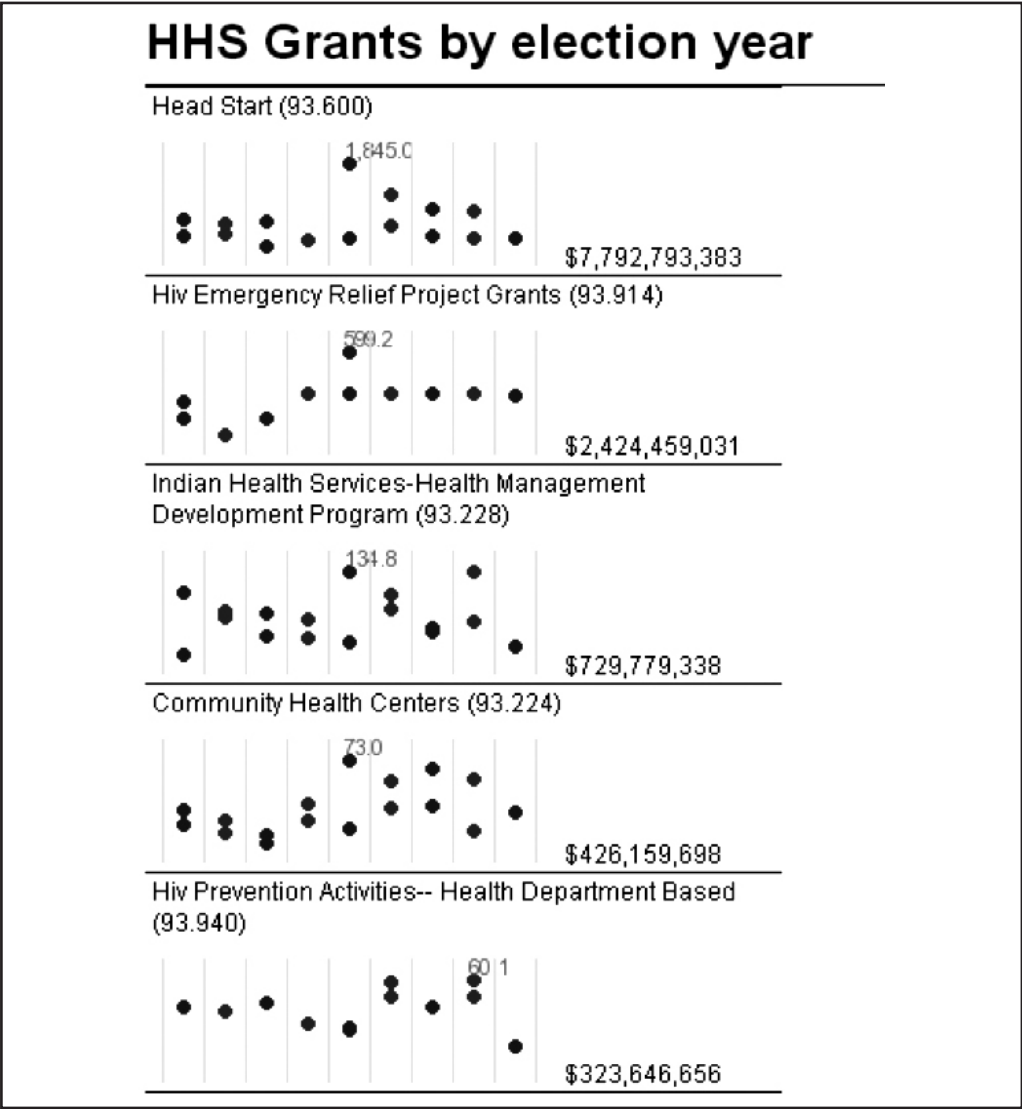
Следат неколку корисни совети за начините на кои визуелизацијата ќе ви помогне да почнете со обработка на податоците.

Совет 1: Користете мали пресметки за брза ориентација во некоја голема група податоци

Таа техника ја користев во „Вашингтон пост“ кога проверувавме една дојава дека администрацијата на Џорџ В. Буш доделува грантови со политичка мотивација, без вистинска основа. Повеќето програми за помош се избираат со помош на формула, други се финансирани со години, така што бевме љубопитни дали можеби постои некој образец во над 1.500-те дискрециски одлуки за доделување грантови.

За секоја од финансираните програми креирав посебен дијаграм во кој црвените точки ги означуваат годините во кои се одржувале претседателски избори, а зелените точки ги означуваа годините во кои имало избори за

Конгресот. Што беше проблемот? Шест месеци пред претседателските избори, кај неколку програми се забележуваше нагло зголемување – на сликата прикажани како црвени точки со највисоки бројки до нив – но се работеше за погрешна изборна година. Истиот образец постојано се прикажуваше и за време на Претседателските избори 2000 година, кога се натпреваруваа Ал Гор (Al Gore) и Џорџ В. Буш, но не и за време на изборите во 2004 година.



Слика 6-3. Граничовиите на Министерството за здравство и социјална политика на САД: „сјарклајнс“ помагаат во откривањето на приказната („Вашингтон пост“)

Сето тоа беше полесно да се види во серија дијаграми, а не во табели со бројки, а еден интерактивен формулар ни дозволуваше да ги провериме различни видови грантови, по региони и по агенции и тела. Мапите со мали пресметки можат да ги прикажат местото и времето како статична илустрација што е лесна за споредба, понекогаш дури и полесна од интерактивните прикази.

Во тој пример користевме кратка компјутерска програма напишана во програмскиот јазик „RНР“, но тоа е уште полесно да се направи со „Excel 2007“ или со линеарните дијаграми „спарклајнс“ (sparklines) вградени во Excel 2010. Стручњакот за визуелизации Едвард Тафти ги измисли тие „интензивни, едноставни графички решенија што наликуваат на зборови“ за пренесување информации со еден брз поглед на податоците. Денес тие „спарклајнс“ ги гледаме насекаде, од малите графикони со берзански извештаи до „гол-разликата“ на табелите во разни спортови.

Совет 2: Погледнете ги податоците свртени наопаку и од страна

Нема лош агол на гледање кога се обидувате да сфатите некоја приказна или група податоци; обидете се со сè што ќе ви падне на ум и ќе добиете нови перспективи. Ако известувате за криминалот, можете да разгледате еден комплет од дијаграми и графикони што ги покажува промените кај насилните кривични дела во текот на една година; друг комплет може да ги прикажува процентуалните промени; третиот може да нуди споредби со други градови; четвртиот може да ги прикажува промените низ историјата. Користете сурови бројки, проценти и индекси.

Споредете ги во различни размери. Следете го правилото дека оската „икс“ мора да ја претставува нултата позиција. Потоа прекршете го тоа правило и видете дали сте дознале нешто повеќе. Пробајте да ги обработите податоците со чудни обрасци на распределба со употреба на логаритамски операции, квадрирање или коренување.

Имајте ги предвид резултатите од истражувањата на визуелната перцепција. Експериментите на Вилијам Кливленд (William Cleveland) покажуваат дека окоото гледа промени во сликата ако просечниот отклон изнесува околу 45 степени. Сево погоре наведено сугерира дека треба да ги игнорирате оние што ви советуваат секогаш да почнете од нула. Наместо тоа, преминете веднаш на приказот што нуди најмногу информации. Други истражувања на полето на епидемиологијата сугерираат дека треба да пронајдете целно ниво што ќе служи како гранична вредност за вашиот дијаграм. Сите горе-наведени поенти ќе ви помогнат да ги погледнете податоците на поинаков начин. Ќе знаете дека сте завршиле кога графичките прикази ќе престанат да ви откриваат нови работи.

Совет 3: Не претпоставувајте ништо

Разгледувајќи ги податоците на повеќе начини и од повеќе агли, веројатно откријте внесови што не изгледаат како што треба. Можеби воопшто не знаете што значат, можеби некои координати изгледаат како да се печатни грешки, а можеби некои трендови изгледаат како да се движат наназад.

Ако сакате да објавите нешто врз основа на претходно извршени истражувања или објавени визуелизации, треба да одговорите на тие прашања и не смеете да правите никакви претпоставки. Или се работи за интересни приказни или се работи за грешки. Во секој случај, ги предизвикуваат општоприфатените ставови или недоразбирања.

Често се случува локалните власти да ви достават сметководствени табели полни со грешки, а, исто така, лесно се случува погрешно да го разберете жаргонот што владите го користат во работата со податоците.

Најпрво навратете се на она што вие сте го сработиле. Дали ја прочитавте документацијата и наведените предупредувања? Дали проблемот се јавува и во изворната верзија на податоците? Ако од вашата страна сè изгледа исправно, тогаш ќе треба да се фатите за телефонот. Ќе треба да го разрешите проблемот ако сакате да ги користите тие податоци, па подобро е да почнете веднаш.

Од друга страна, сите грешки не се подеднакво значајни. Во регистрите на финансирањето на изборните кампањи, нормално е да пронајдете стотина непостоечки поштенски броеви во база на податоци со 100.000 записи. Сè додека сите грешки не се однесуваат на ист град или на ист кандидат, повремени лоши записи на податоци немаат значење.

Прашањето што мора да си го поставите е дали, ако го искористам ова, читателите ќе добијат во основа точна информација за тоа што велат податоците?

Совет 4: Не бидете опседнати со прецизноста

Наспроти непоставувањето прашања ја имаме опседнатоста со прецизноста уште пред таа воопшто да има значење. Графичките решенија во кои ги нудите објаснувањата треба да се генерално точни и прецизни, но не грижете се премногу ако треба да заокружите некоја децимала, ако зборовите на процентите не изнесуваат точно 100 отсто, или ако во период од 20 години ви недостигаат податоците за една или две години. Сето тоа е дел од процесот на истражување. Сè уште ќе можете да ги идентификувате значајните трендови и ќе знаете што треба да соберете пред да дојде времето за објавување. Всушност, можете да размислите да ги отстраните ознаките за имињата и размерот, како на графиконите наведени погоре, за да добиете подобро чувство за податоците.

Совет 5: Подгответе хронологи на случаите и настаните

На почетокот на секоја комплексна сторија, почнете да правите хронолошки табели со клучните настани и случаи. За таа цел можете да користите „Excel“, „Word“ или некоја од специјалните алатки каква што е „Тајмфлоу“ (TimeFlow, <http://hint.fm/projects/timeflow/>), но во еден момент ќе сретнете податоци што ќе треба да ги разложите хронолошки, одејќи наназад низ времето. Со редовно читање на хронологијата ќе видите кои дупки во известувањето треба да ги пополните.

Совет 6: Состанете се со луѓето од графичкото одделение што порано и состанувајте се со нив често

Разговарајте за можните графички решенија со уметниците и дизајнерите што работат во вашата редакција. Тие ќе имаат сопствен начин на гледање на податоците, предлози како да ги прикажете на интерактивен начин и како да ги поврзете податоците и приказните. Во известувањето, тоа ќе ви помогне да знаете што треба да соберете што е можно порано, или да можете на време да го известите тимот дека некое графичко решение не е можно затоа што не можете да ги соберете потребните работи.

Совети за објавувањето

Понекогаш ќе ви требаат само неколку денови или часови, а понекогаш сторијата бара истражување што трае неколку месеци. Како што се приближува времето за објавување, два аспекти сè повеќе добиваат на значење.

Ја помнете годината што ви недостигаше на почетокот на истражувањето? Одеднаш излегува дека не можете да продолжите понатаму без тие податоци. Сите лоши податоци што ги игнориравте во известувањето? Ќе се вратат да ве прогонуваат. Поентата е дека не можете да пишувате врз основа на лоши податоци. Работата со графичките решенија е едноставна, или имате сè што ви треба или немате, нема средина.

Трудот вложен во собирање на податоциите мора да одговара на трудот вложен во изработката на интерактивниот графички приказ

Интерактивниот графички приказ не нуди можност да скриете некој недостаток. Ако навистина сакате вашите читатели да ги истражуваат податоците на секој начин што ќе го посакаат, тогаш сите податоци треба да бидат она што тврдат дека се. Корисниците секогаш можат да пронајдат некоја грешка и таа потоа да ве прогонува со месеци или години. Ако градите сопствена база на податоци, тоа значи дека од вас ќе се очекува да ги препрочитате, проверите и да ги внесете потребните исправки во целата база. Ако користите документи добиени од владата, ќе треба да одлучи-

те колку проверки ќе направите и што ќе правите кога ќе откриете некоја грешка. Грешки секако ќе има, тоа е неизбежно.

Дизајнирајте за два вида читателци

Графичкиот приказ – без оглед дали се работи за автономна интерактивна функција или за статична визуелизација што ја придружува статијата – треба да задоволи два различни вида читатели. Треба да биде лесно разбирлив на прв поглед, а сепак да е доволно комплексен за да им понуди нешто интересно на оние што сакаат да одат подлабоко. Ако се работи за интерактивен приказ, на читателите обезбедете им нешто повеќе од еден единствен број или име.

Претставете една идеја, постојат поединствени ја

Сигурни сте дека сакате луѓето да видат само една работа? Одлучете каков впечаток сакате да оставите и погрижете се да скриете сè друго. Во мнозинството случаи, тоа значи дека треба да отстраните некои информации дури и ако интернетот ви дозволува да понудите сè. Повеќето од деталите собрани во хронолошкиот преглед едноставно не се многу значајни, освен во случаите кога единствена грижа ви е транспарентноста на извештавањето. Во статичен графички приказ, тоа може да биде застрашувачко. Во интерактивен графички приказ, од друга страна, ќе биде здодевно.

– Сара Коен, Универзитетот „Дјук“ (Sarah Cohen, Duke University)

Користење визуелизации во раскажувањето приказни

Постојат неколку причини поради кои треба да се осврнеме на визуелизацијата на податоците. Не само што визуелизацијата може да биде прекрасна и да привлекува внимание – што е вредна валута во процесот на размена и во привлекувањето читатели – таа, исто така, има на располагање и една моќна когнитивна предност. Се работи за фактот дека цела една половина од човечкиот мозок е посветена на обработката на визуелни информации. Кога на корисниците им нудите графички прикажана информација, до нив допирате преку портата што води до човечкиот мозок што има најголема проточност. Убаво дизајнираната визуелизација ќе остави силен и длабок впечаток на оние што ја гледаат, тоа ќе го направи од прва и ќе се пробие низ шумот директно до најзначајната поента на една комплексна приказна.

За разлика од другите визуелни медиуми – како што се фотографијата и видеото – визуелизацијата на податоци е длабоко вкоренета во мерливите факти. Иако естетски привлечна, таа носи помал емоционален набој и повеќе се грижи да расветли отколку да „загрее“ нешто. Во денешнава ера на тесно фокусирани ме-

диуми, често приспособени за публика со точно определени ставови, визуелизацијата на податоците (и дата-новинарството воопшто) нуди возбудлива можност за раскажување приказни што, пред сè, е водена од факти, а не од фанатизам.

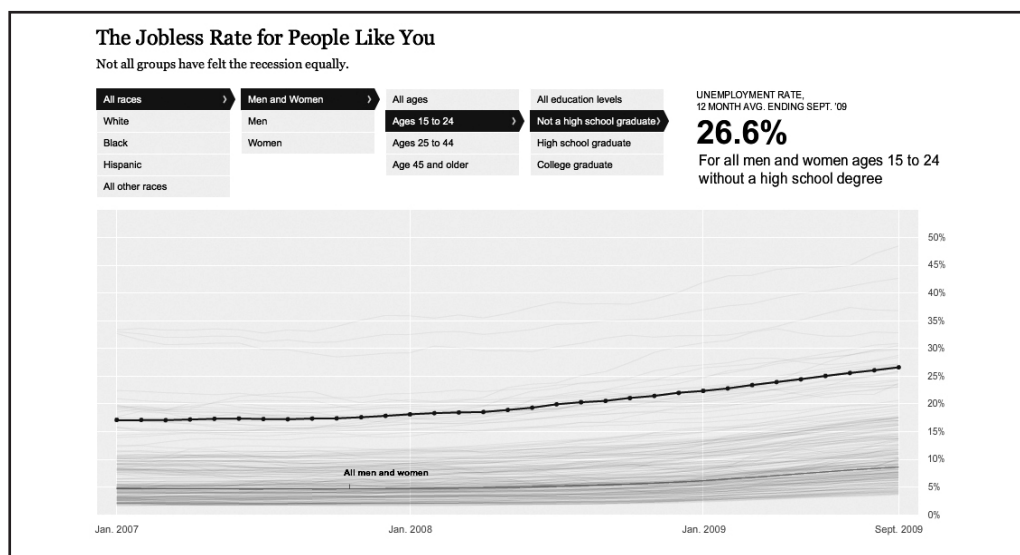
Понатаму, како и другите форми на наративно новинарство, визуелизацијата може да се покаже ефикасна и на полето на откривање ударни вести – брзо пренесување нови информации, на пример, местото на кое се случила сообраќајна незгода и бројот на жртвите – и за подолги студии („фичери“), каде може да влезе подлабоко во некоја тема и да понуди нова перспектива, да помогне да согледате нешто што веќе ви е познато од нов агол.

Нов поглед на познатото

Добра илустрација за способноста на визуелизацијата да ги тестира и предизвика конвенционалните знаења нуди едно интерактивно графичко решение објавено од „Њујорк тајмс“ кон крајот на 2009 година, една година по избувнувањето на светската економска криза (<http://nyti.ms/employment-lines>). Во време кога националната стапка на невработеност во САД се движеше околу 9 отсто, корисниците можеа да го филтрираат населението на САД низ различни образовни и демографски филтри и да видат дека постојат драматични разлики во стапката на невработеност. Се покажа дека стапката варира. Варијациите одеа од помалку од 4 отсто кај средовечните жени со завршено високо образование, до безмалку половина од младите црни мажи без завршено средно образование. Таквиот диспарат воопшто не беше нов, а тоа го потврдија линиските дијаграми на кои беа прикажани историските вредности за секоја од индивидуалните групи.

Дури и кога ќе престанете да гледате во неа, добрата визуелизација на податоци ви влегува в глава и остава траен ментален модел за некој факт, тренд или процес. Кој знае колку луѓе ја погледаа анимацијата (http://nctr.pmel.noaa.gov/indo_1204.html) објавена од истражувачите што ги проучуваат цунамијата во декември 2004 година, на која се прикажани брановите како се шират од епицентарот на земјотресот во Индонезија преку Индискиот Океан, заканувајќи се на милиони луѓе во крајбрежните области во Јужна Азија и Источна Африка?

Визуелизациите на податоци – и естетските асоцијации што ги носат со себе – можат да прераснат во културни знаци. Таква беше претставата за длабоките политички поделби во САД по изборите во 2000 и 2004 година, кога „црвените“ држави во кои победија републиканците ги обоија централните региони, додека „сините“ држави во кои доминираат демократите беа собрани во гроздови на североистокот и на западниот брег на САД. Притоа не игра улога фактот што водечките радиодифузни мрежи во САД, до 2000 година, често ги менуваа сината и црвената боја како ознаки за партиите, а некои од нив тоа го правеа на секои четири години. Оттаму и фактот што голем број Американци ја помнат изборната 1984 година како година на епската „сина“ победа на Роналд Реган, кога републиканскиот кандидат ја освои победата во 49 држави.



Слика 6-4. Слика на невработеност на луѓето како вас („Њујорк тајмс“)

На секое графичко решение што се придржува до востановените визуелни клишеа, доаѓа друго што нуди моќно сведоштво засновано на фактите. Така ја имаме мапата на „Њујорк тајмс“ од 2006 година (http://www.nytimes.com/imagepages/2005/10/02/national/nationalspecial/20051002diaspora_graphic.html) што користеше кругови со различна големина за да покаже каде се преселиле стотиците илјади раселени лица од Њу Орлеанс, расфрлени по целиот континент како резултат на личните врски и програмите за релокација. Дали „заталканите“ раселени лица некогаш ќе се вратат дома?

Сега, кога ја разгледаваме моќта на визуелизациите на податоците, ред е да се постави прашањето кога треба, а кога не треба да ги користиме? Но, најпрво да погледнеме неколку примери во кои визуелизацијата на податоците може да помогне на читателите да им раскажете некоја приказна.

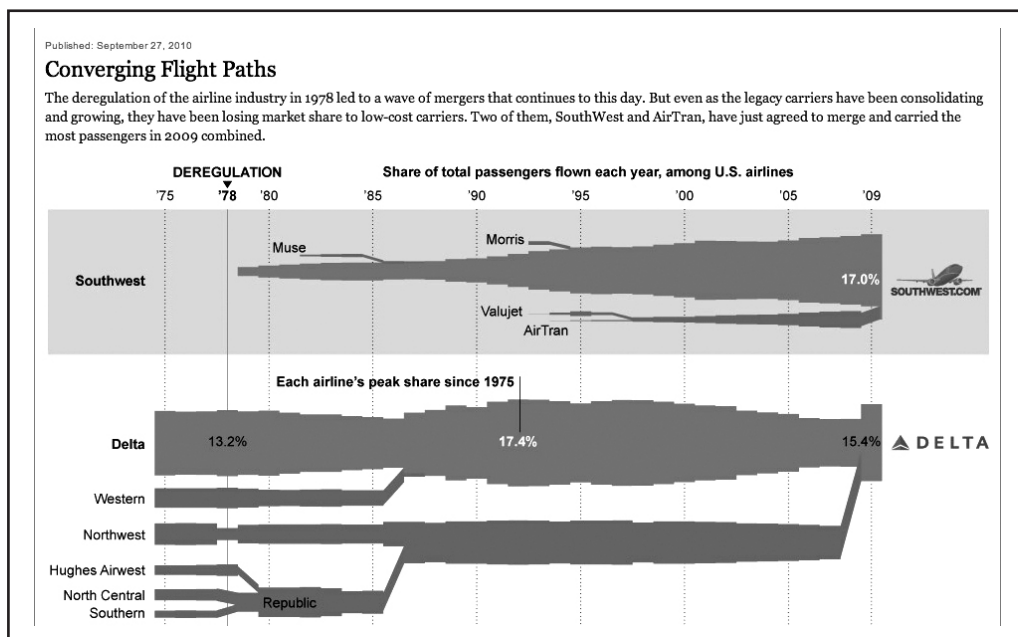
Прикажување на промените што настанале со текот на времето

Веројатно најчестата употреба на визуелизацијата на податоци е прикажувањето како вредностите се менуваат со текот на времето, најчесто на обичен линиски дијаграм. Порастот на населението на Кина од 1960 година наваму (<http://bit.ly/google-china-population>) или наглиот пораст на невработеноста по економскиот крах од 2008 година се добри примери за такви визуелизации. Но, промените со текот на времето можат успешно и впечатливо да бидат прикажани и со други графички форми. Португалскиот истражувач Педро М. Круз (Pedro M. Cruz) понуди драматичен приказ на падот на западноевропските им-

перии (<http://pmcruz.com/visual-experiments/visualizing-empires>) од почетокот на 19 век наваму со помош на анимирани графикони со кружници. Прикажани во однос на вкупното население, Британија, Франција, Шпанија и Португалија пукаат како меури од сапуница со прогласувањето независност на секој нов прекуморски посед. Си заминуваат Мексико, Бразил, Австралија, Индија и, само момент... си одат најголемиот број африкански колонии во првата половина на 1960-тите, безмалку целосно бришејќи ја Франција од графиконот.

Графиконот на „Волстрит журнал“ (<http://blogs.wsj.com/venturecapital/2009/08/25/how-long-does-it-take-to-build-a-technology-empire/>) покажува колку месеци им биле потребни на сто претприемачи да ја достигнат магичната граница на приход од 50 милиони долари. Изработена со помош на слободната алатка за подготовка на дијаграми и анализа на податоци „Табло паблик“ (Tableau Public), споредбата наликува на трагите на полетување на повеќе авиони – некои се побрзи, некои се побавни, некои се тешки, исцртани еден преку друг.

Кога сме кај авионите, друг интересен графикон со историски промени го исцртува пазарниот удел на водечките авиокомпани од САД во текот на неколкуте децении на консолидирање на авиотранспортната индустрија во САД. Откако администрацијата на претседателот Картер (Jimmy Carter) ја дерегулираше областа на патничката авијација, еден бран аквизиции, финансирани преку јавниот долг, создадоа неколку национални авиокомпани од помали регионални авиопревозници, процес прикажан на графиконот објавен од „Њујорк тајмс“.

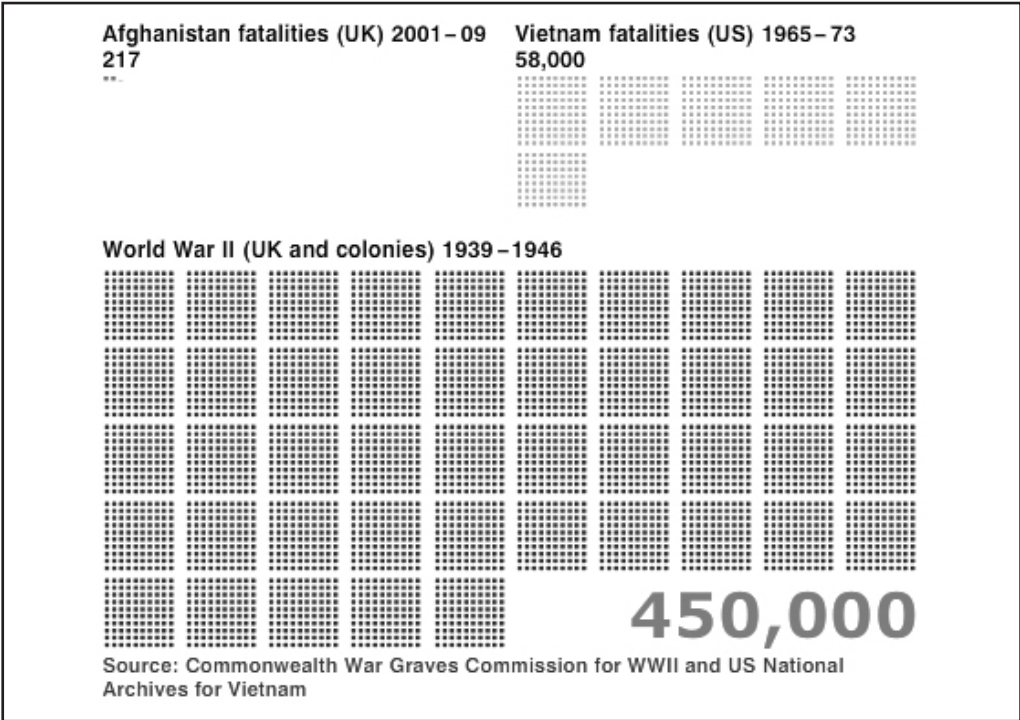


Слика 6-5. Преклојување на летиовише („Њујорк тајмс“)

Со оглед на фактот што за обичните читатели хоризонталната „икс“ оска на дијаграмот го претставува течењето на времето, може лесно да се помисли дека сите визуелизации ги прикажуваат промените во текот на времето.

Споредување вредности

Визуелизацијата, исто така, е мошне успешна во помагањето на читатели-те да споредат две или повеќе дискретни вредности, без разлика дали треба да се обезбеди контекстот на трагично загинатите војници во Ирак и Авганистан (преку споредба со илјадниците загинати во Виетнам или милионите жртви од Втората светска војна, што е приод што Би-Би-Си го примени во анимираниот приказ на слајдови што ја придружуваше нивната база на податоци на загинатите во конфликтите). Или кога „Нешнал џеографик“ (National Geographic), преку еден минималистички дијаграм, покажува колку е поголема веројатноста да починете од срцева болест (веројатноста е еден на пет) или срцев удар (1 на 24), отколку да загинете во авионска несреќа (1 на 5.051), од каснување од пчела (1 на 56.789), со приказ на релативната веројатност на смртта (секако, во сенка на огромниот лак што ја прикажува општата веројатност на умирање која изнесува 100 одсто, односно сите еден ден ќе умреме!).

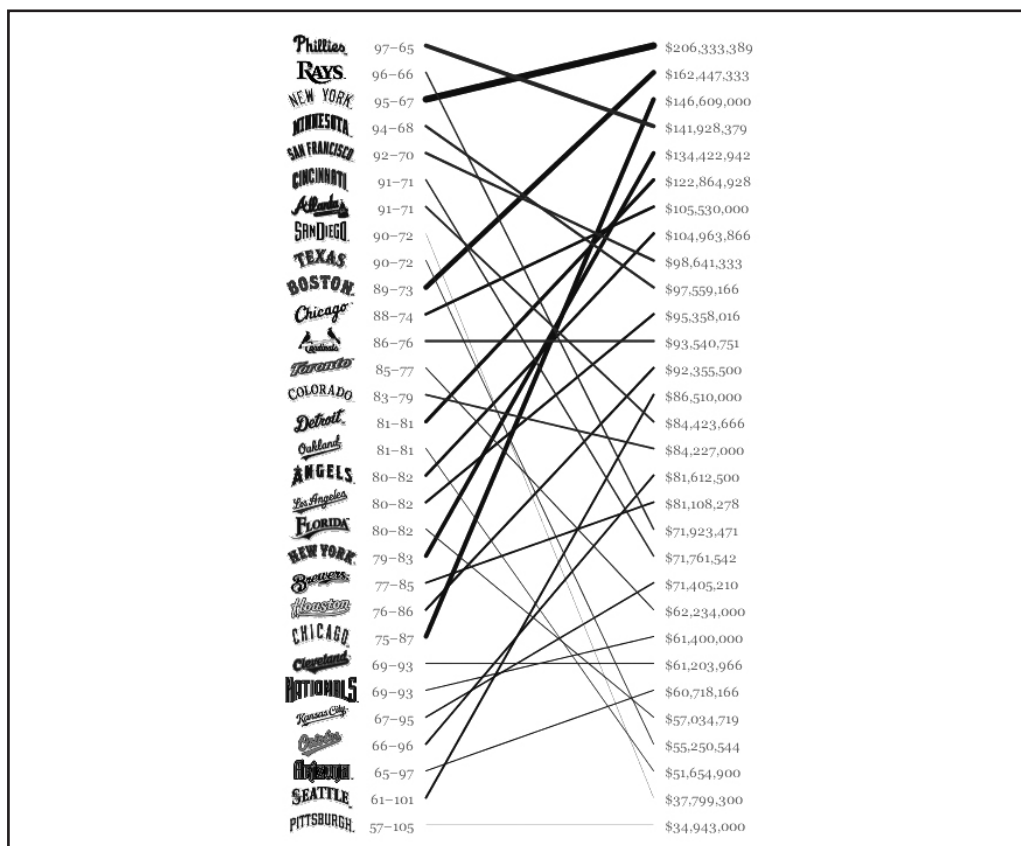


Слика 6-6. Споредба на војните по бројот на загинајќи (Би-Би-Си)

Во соработка со агенцијата „Берг дизајн“ (Berg Design), Би-Би-Си, исто така, ја разви и веб-страницата „Димензии“ (Dimensions, <http://howbigreally.com/>), каде што можете, на мапата на „Гугл“ на која е прикажана вашата локална заедница, да ги прикажете, во соодветен размер, важните светски настани – на пример, истекувањето нафта од нафтната платформа „Дипвотер хорајзон“ (Deerwater Horizon) и поплавите во Пакистан.

Прикажување на меѓусебните врски и тековите на податоците

Воведувањето на супер-брзите возови во Франција во 1981 година можеби не ја намали физички земјата, но затоа имаме една „згодна“ визуелна претстава која покажува колку помалку време е потребно за да се стигне до различни дестинации отколку со конвенционална железница. Мрежата од линии нацртана преку Франција, на сликата „пред“ е создадена од правилни квадрати. На сликата „потоа“ таа е „собрана“ во центарот во правец на Париз, покажувајќи не само колку се



Слика 6-7. Плаќањата споредена со усвојувањето (Бен Фрај)

„поблиски“ дестинациите, туку и дека најголемата заштеда на време се јавува во првиот дел на патувањето од Париз накај периферијата, пред возовите да стигнат до деловите од пругата што не биле подобрени и поради тоа морале да забават.

За пример на споредба помеѓу две засебни променливи вредности, погледнете го графиконот на Бен Фрај (Ben Fry) што нуди оценки на успехот на тимовите во американската безбол лига (Major League Baseball) споредено со вкупната вредност на договорите со нивните играчи. Во левата колона, тимовите се подредени според нивниот скор, додека од десната страна се прикажани вкупните плати исплатени на нивните играчи. Двете вредности се поврзани со црвена (успех под очекувањата) или со сина (успех над очекувањата) линија, што овозможува да видиме кои сопственици на тимови жалат што преплатиле некои играчи кои се покажале како промашувања. Понатаму, ако ги прегледаме хронолошките приказите, ќе добиеме жива анимација на „борбата за титулата“ за определена сезона од почеток до крај.

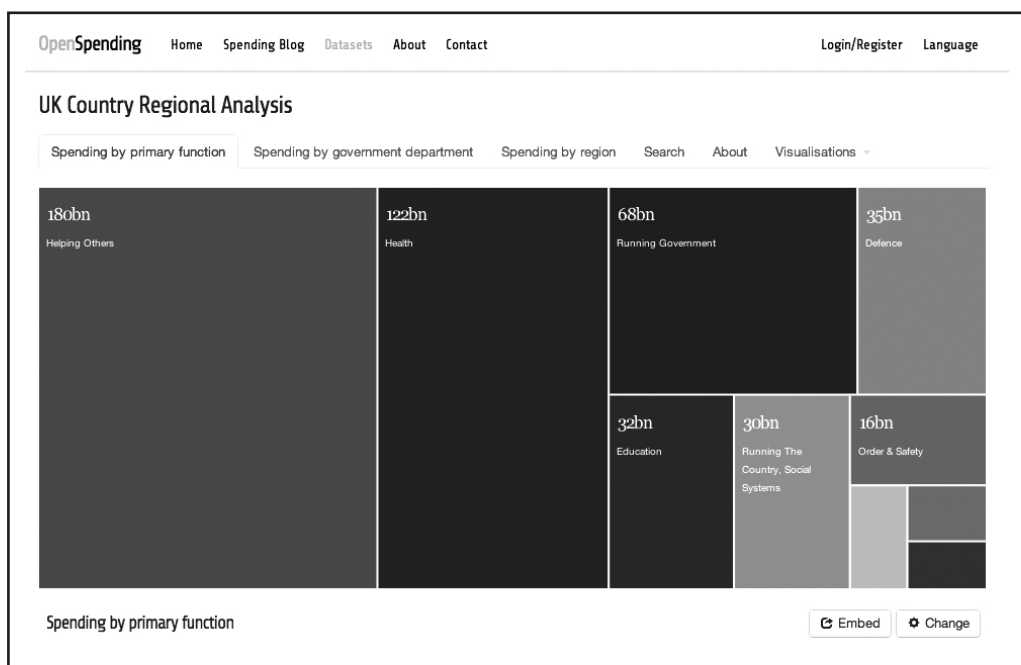
Дизајнирање со помош на податоци

Донекаде слични на графичкото прикажување на врските меѓу податоците, дијаграмите на тек нудат информации преку линиите што ги поврзуваат полињата, вообичаено со користење линии со различна ширина или боја. На пример, во ситуација на криза во Евроната во која неколку земји членки не можеа да ги сервисираат своите долгови, „Њујорк тајмс“ се обиде да ја размрси мрежата на заеми што ги поврзува земјите членки на ЕУ со нивните трговски партнери од другата страна на Атлантикот и од Азија. Во еден „стадиум“ на визуелизацијата, ширината на линијата го претставува износот на кредитите одобрени од една држава на друга, додека преминот од жолта до портокалова боја укажува колку „загрижувачки“ е статусот на заемот, односно колку е мала веројатноста тој да биде вратен!

На една друга, малку повесела тема, магазинот „Нешнал џеографик“ (National Geographic) подготви еден, навидум едноставен дијаграм, на кој се прикажани врските што три големи града во САД – Њујорк, Чикаго и Лос Ангелес – ги имаат со големи винарски региони, и како начинот на транспорт на виното од секој од изворите може да резултира со драматично различни „јаглородни отпечатоци“, така што, на пример, виното од Бордо, во Франција, е „позелена“ опција за жителите на Њујорк од калифорниските вина. „Сорсмап“ (SourceMap, <http://sourcemap.com/>), проект осмислен во школата за бизнис на Институтот за технологија на Масачусетс (Massachusetts Institute of Technology – MIT), користи дијаграми на текови за еден ригорозен преглед на глобалните набавки на индустриски производи, нивните компоненти и потребните сировини за производство. Благодарение на обемното истражување, корисниците сега можат да извршат пребарување по различни производи, од чевлите на производителот „Еко“ (Есо) до сок од портокали и да пронајдат од кој дел на светот потекнуваат и каков е нивниот „јаглороден отпечаток“.

Прикажување на хиерархиските структури

Во 1991 година, истражувачот Бен Шнајдермен (Ben Shneiderman) измисли нова форма на визуелен приказ наречена „мапа во форма на дрво“ (treemap, <http://www.cs.umd.edu/hcil/treemap-history/>), која се состои од повеќе полиња поставени концентрично едно во друго. Површината на полето ја претставува прикажаната количина за самото дрво и како збир од другите полиња содржани во него. Без оглед на тоа дали се работи за визуелизација на распределбата на националниот буџет по тела и нивни оддели, визуелизација на берзата на хартии од вредност по различни сектори и по индивидуални компании, или на еден програмски јазик по класи и поткласи, „дрвото“ претставува компактен и интуитивен преглед за мапирање на некој субјект и на неговите составни делови. Друг одличен формат е „дендрограмата“ (dendrogram), или „дрво-дијаграм“, што повеќе изгледа како вообичаен организациски дијаграм во кој поткатегориите се одвојуваат како гранки од изворниот труп.



Слика 6-8. OpenSpending.org (Фондација „Отворено знаење“)

Прегледување големи бази на податоци

Иако понекогаш визуелизациите се ефикасни во прикажувањето познати информации во ново светло, прашањето е што се случува кога имате сосема нови информации што луѓето сакаат да ги разгледаат? Ерата на податоците во која живееме секој ден носи нови откритија, од брилијантната географска анализа на фотографии на „Фликр“ (Flickr snapshots, <http://bit.ly/flickr-analysis>) на Ерик Фишер (Eric Fischer), до објавувањето илјадници претходно доверливи оценки на наставниците од страна на градските власти на Њујорк (<http://projects.wsj.com/nyc-teachers/>).

Таквите групи податоци имаат најголема моќ кога корисниците можат да ги „прекопуваат“ и да „ријат“ по нив за да дојдат до онаа информација која е релевантна за нив лично.

Во почетокот на 2010 година, „Њујорк тајмс“ доби пристап до, во нормални услови, доверливите податоци за тоа кои филмови на сервисот „Нетфликс“ (Netflix) најмногу се изнајмуваат во определени делови на градот. Кога „Нетфликс“ одби да ги обезбеди суровите бројки, „Њујорк тајмс“ креира привлечна интерактивна база на податоци (<http://nyti.ms/interactive-database>) која на корисниците им дозволуваше да ги прегледаат 100-те најпопуларни филмови што биле изнајмувани во 12 метрополи во САД, разложени до поштенскиот број од каде доаѓа изнајмувачот. „Термичка мапа“, шифрирана со различни бои, исцртана преку мапата на секоја поединечна заедница, им овозможуваше на корисниците набрзина да прегледаат и да видат каде определен филм уживал најголема популарност.

Кон крајот на истата година, „Њујорк тајмс“ ги објави резултатите од пописот на населението во САД што се врши еднаш на десет години (<http://nyti.ms/census-explorer>), само неколку часа откако официјално беа објавени. Корисничкиот интерфејс, изработен во „Adobe Flash“ нудеше повеќе можности за визуелизација и им дозволуваше на корисниците да ги прегледуваат податоците до ниво на основните пописни единици (на национално ниво има 8,2 милиони такви единици) и да видат како е распределено населението по раса, приходи и ниво на образование. Податоците беа прикажани во толку фина резолуција што, прегледувајќи ја базата во првите неколку часа по објавувањето, моравте да се прашате дали не сте првиот човек на светот што разгледува некој дел од визуелизацијата.

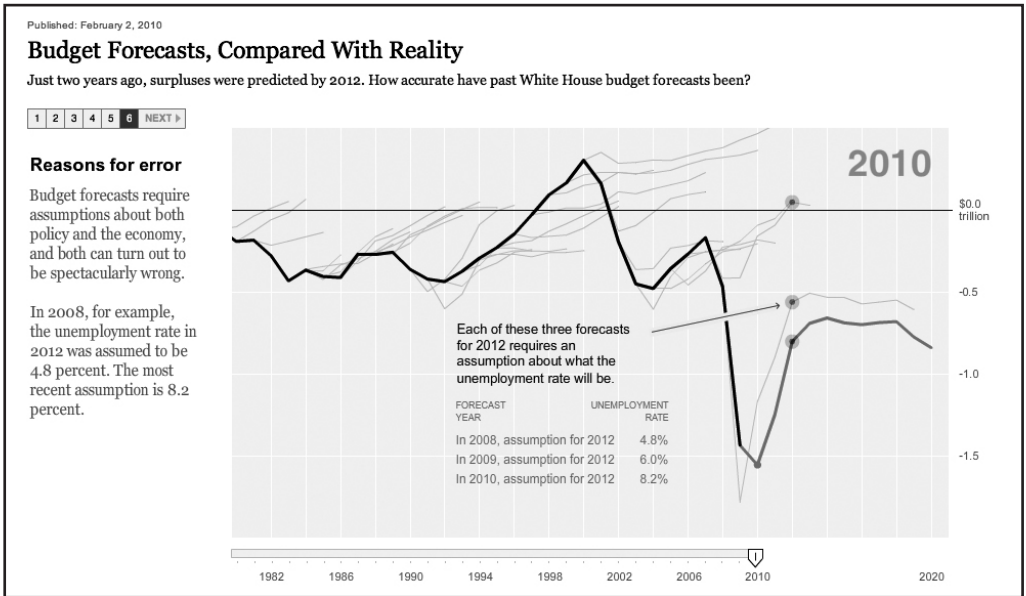
Слични пофални примери на користење визуелизации како насловни страници на бази на податоци се и истражувањето на смртните случаи во сообраќајни незгоди на Би-Би-Си и повеќето обиди за брзо индексирање огромни збирови податоци како што беше објавувањето на воените дневници од Ирак и Авганистан што ги објави „Викиликс“.

Замислување на поинакви резултати

Линеарниот дијаграм на Аманда Кокс (Amanda Cox) од „Њујорк тајмс“, на кој беа прикажани трагикомично оптимистичките предвидувања за трговскиот дефицит на САД во период од повеќе години, покажува дека понекогаш тоа што не се случило може да биде многу поинтересно од она што се случило. Трескавичната линија на Кокс која го отсликува забрзаниот раст на дефицитот по цела деценија војни и даночни олеснувања, покажува колку нереални можат да бидат очекувањата за иднината.

Брет Виктор (Bret Victor), искусен дизајнер на интерфејси за „Епл“ (Apple) и родоначалник на теоријата за визуелизацијата како средство за пренесување квантитативни информации „убиј ја математиката“ („kill math“ theory), подготви прототип за еден вид реактивен документ. Во неговиот пример, идеите за штедење на енергијата вклучуваат премиси што можат да се менуваат, така што едноставна постапка како што е гаснењето на сијалиците во празните соби може на Американците да им заштеди енергија еднаква на производството на 2 до 40 термоелектрани на јаглен. Промената на процентите наведени во средината на еден параграф автоматски повлекува ажурирање на остатокот од текстот!

За повеќе примери и сугестии, погледнете го списокот на различни начини на употреба (<http://bit.ly/ericson-links>) на визуелизации, мапи и интерактивни графички решенија, подготвен од Метју Ериксон (Matthew Ericson) од „Њујорк тајмс“.



Слика 6-10. Сјоредба помеѓу проектираниите и реализираниите буџети („Њујорк тајмс“)

Кога не треба да се користи визуелизација

Во крајна линија, квалитетната визуелизација на податоци зависи од добрите, чисти, точни полнозначни информации. Во истата мера во која квалитетното наративно новинарство се потпира на добрите изјави, факти и описи, и визуелизацијата на податоците ќе биде онолку добра колку што се добри податоците што ја придвижуваат.

Кога приказната ќе биде подобро раскажана со текст или со мултимедијски решенија

Понекогаш податоците сами за себе не можат да ја направат приказната доволно уверлива. Ако еден едноставен дијаграм може да биде корисен во опишувањето еден тренд или резимирани статистички податоци, наративот што се однесува на реалните последици од едно прашање може да има подиректно и посилно влијание врз читателите.

Кога имате многу малку статистички точки

Некој рекол дека „бројот сам за себе не значи ништо“. Најчестиот одговор што уредниците на вести го даваат кога некој ќе наведе некаква статистика е прашањето „споредено со што?“. Дали трендот е нагорен или надолен? Која е нормалната состојба?

Кога има многу малку варијации во податоците, кога не е видлив јасен тренд или заклучок

Понекогаш, додека ги внесувате податоците во „Excel“ или во некоја друга апликација за изготвување графикони, ќе забележите дека во информациите има многу шум, флуктуации или трендовите се релативно „рамни“. Дали основата ќе ја поместите од нула до најниската вредност што ја имате за линијата да добие подобра форма? Не! Изгледа дека податоците се двомислени и им треба уште „копање“ и анализа.

Кога мајката не е мајка

Понекогаш просторните елементи немаат значење, не се убедливи или го одвлекуваат вниманието од позначајните нумерички трендови, како што е промената во тек на времето или прикажувањето сличности помеѓу области кои не се допираат.

Кога и најобична табела ќе заврши работата

Ако имате релативно малку статистички точки, а сепак имате информации што можат да им бидат од корист на некои од вашите читатели, размислете за можноста податоците да ги претставите само во форма на табела. Табелите се чисти, лесни за читање и не создаваат нереални очекувања од „приказната“. Всушност, табелите можат да обезбедат ефикасен и елегантен приказ на основните информации.

– Геоф Мекџи, Универзитет „Стивенс“ (Geoff McGhee, Stanford University)

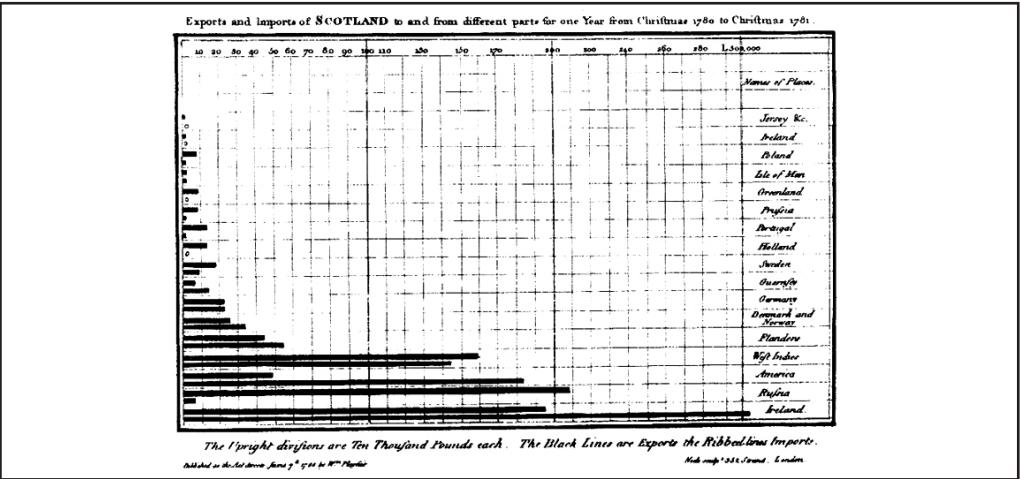
Различни графикони раскажуваат различни приказни

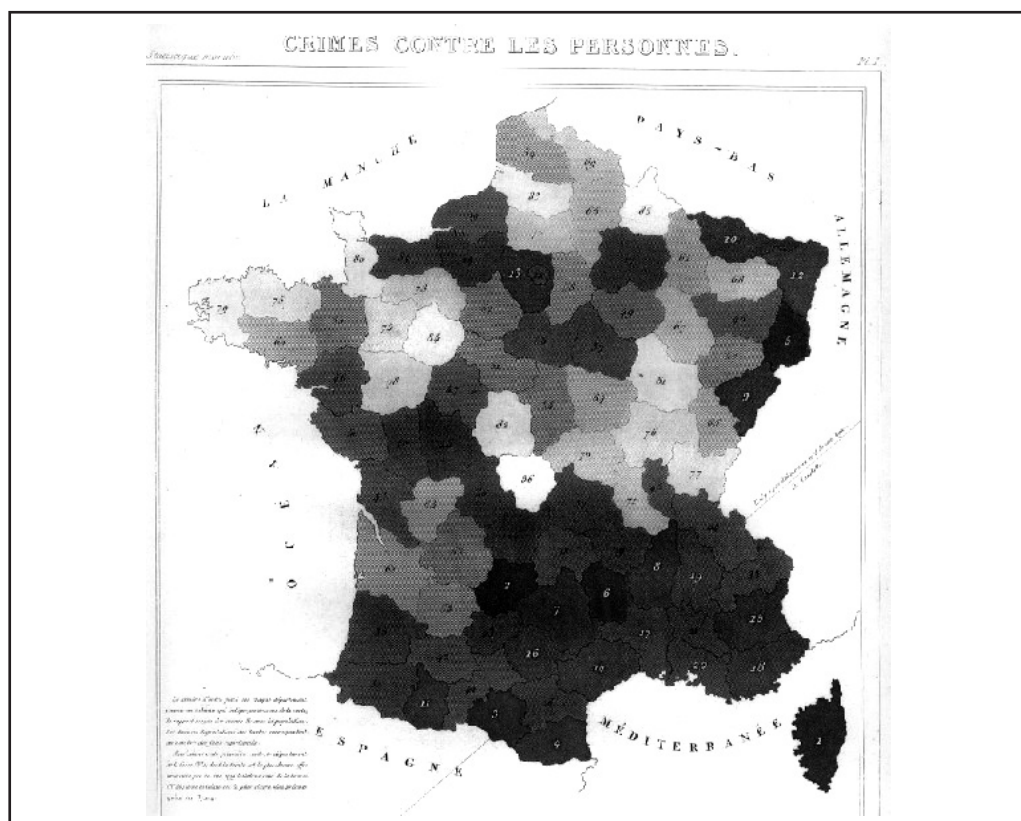
Во дигиталниот свет во кој живееме, свет што ветува бескрајни „3Д“ искуства, почнуваме да забораваме дека претходно, со векови, на располагање ги имавме само мастилото и хартијата. Денес, на тој статичен, рамен медиум гледаме како на граѓанин од втора класа, но факт е дека во стотиците години во кои пишувавме и печатевме, успеавме да достигнеме извонредно ниво на знаења и практики за претставување на податоците на хартија. Иако денес се зборува само за интерактивните графикони, визуелизации и инфографики, тие се само продолжение на неколку стари практики. Само ако се навратиме наназад, низ историјата на успешните дијаграми и графикони, можеме да разбереме какво богатство носиме со себе во новите медиуми.

Некои од најпознатите дијаграми и графикони произлегуваат од потребата за подобро објаснување табели со набиени податоци. Вилијам Плејфер (William Playfair) е шкотски полиглот што живеел во втората половина на 18. и првата половина на 19 век. Тој своерачно создал многу од дијаграмите и графиконите што ги користиме до денешен ден. Во својата книга „Трговски и политички атлас“ (Commercial and Political Atlas), објавена во 1786 година, Плејфер го вовел столбестиот дијаграм за појасно да ги прикаже увозот и извозот на Шкотска на нов визуелен начин.

Подоцна, во книгата „Статистички брeвијар“ (Statistical Breviary), од 1801 година, го популаризирал озлогласениот кружен (пита) дијаграм. Потребата од такви нови форми на дијаграми и графикони прво се појавила во трговијата, но со текот на времето се јавуваат и други форми, понекогаш користени буквално за спасување човечки животи. Во 1854 година, Џон Сноу (John Snow) ја создава денес славната „Карта на колерата во Лондон“, на која додал мала црна прачка над секоја адреса на која била забележена појава на болеста. Со текот на времето, зоните со поголема или помала густина на заболени станале видливи и биле преземени дејства за решавање на проблемот.

Оние што ги употребуваа новите дијаграми и графикони стануваа се похрабри и продолжуваа со експериментите, постојано поместувајќи ги границите на медиумот. Андре-Мишел Гери (André-Michel Guerry) беше првиот што ја предложи идејата за мапа на која одделните региони би се обоиле со различни бои, врз основа на некоја променлива вредност. Во 1829 година, тој го објави првиот „хороплет“ (choropleth, повеќебојна мапа, од старогрчки „хоро“ – регион и „плетос“ – мноштво) на регионите на Франција прикажани во различни нијанси на сива боја што ги означуваа различните стапки на криминал. Денес таквите мапи се користат за прикажување на изборните резултати по региони, кој за кого гласал, распределбата на богатството и други географски дефинирани променливи вредности. Изгледа како едноставна идеја, но дури и денес е тешко да се усоврши и да се разбере ако не се користи правилно.

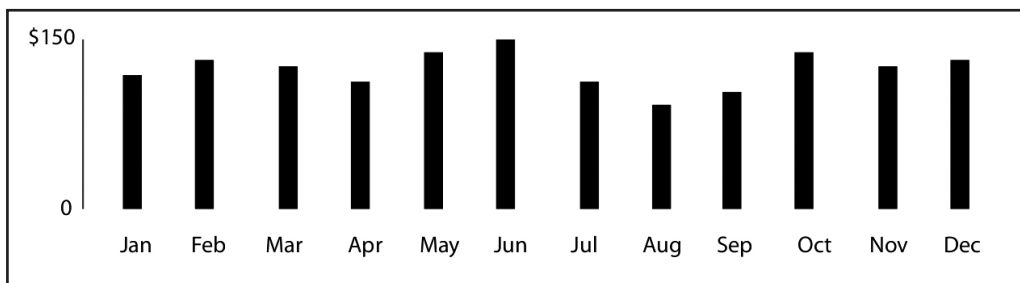




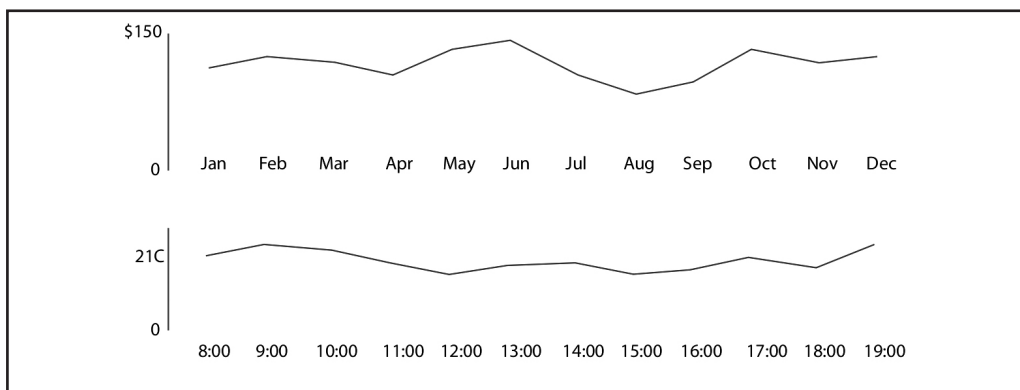
Слика 6-13. Хоройлейска мапа на Франција на која се прикажани ситајкиите на криминал (Андре-Мишел Гери)

Постојат повеќе алатки што добрите новинари треба да ги совладаат и да ги вбројат во алатките што ги користат во изработката на визуелизации. Значајно е да се усвојат основните знаења за дијаграмите и графиконите. Би кажалe, да се почувствува дното под нозете, а не да се скокне веднаш во длабокиот дел од базенот. Сè што создавате треба да потекнува од некаква серија на основни дијаграми и графикони. Дури кога ќе ги совладате основите можете да преминете на конструирање покомплексни визуелизации што се составени од основните единици.

Столбестите дијаграми и линиските дијаграми се два од неколкуте основни видови дијаграми. Иако се мошне слични по начинот на нивната употреба, тие можат многу да се разликуваат по своето значење. Да ги разгледаме, на пример, резултатите од продажбата на една компанија за секој месец од годината. Ќе добиеме 12 столба што ги претставуваат сумите пари заработени за секој месец одделно (Слика 6-14).



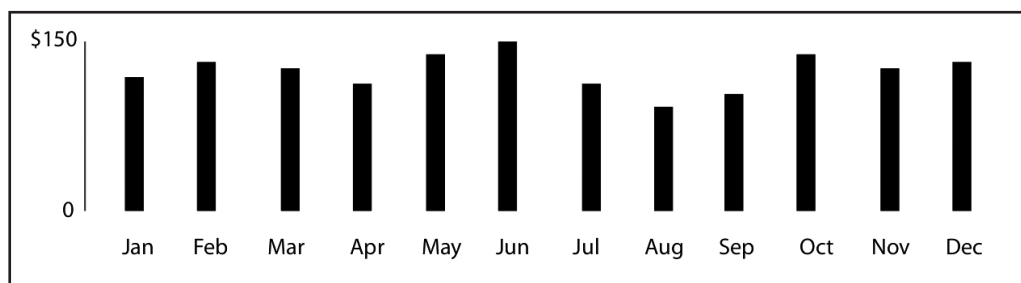
Слика 6-14. Едноставен столбест дијаграм: корисен за претставување на дискретни информации



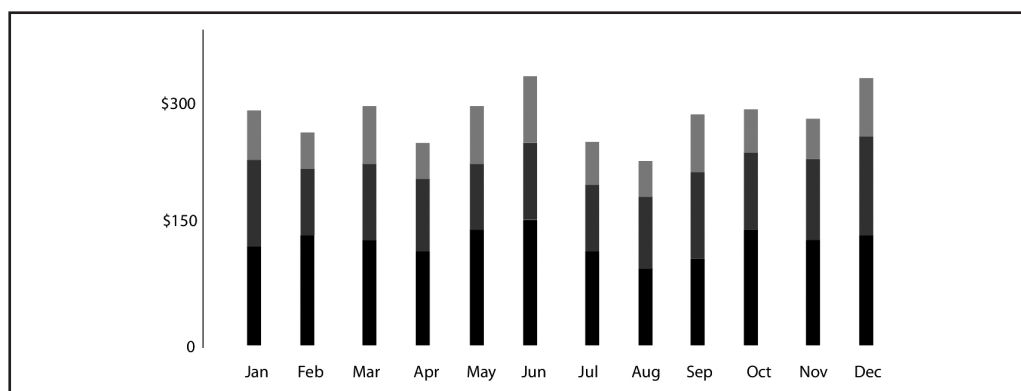
Слика 6-15. Едноставен линиски дијаграм, корисен за претставување на континуирани информации

Сега, да видиме зошто е подобро да користиме столбест отколку линиски дијаграм. Линиските дијаграми се идеални за податоци што течат континуирано. Во нашиот случај со резултатите од продажбата, вкупните износи за секој одделен месец не се континуирани податоци. Столпчињата ни кажуваат дека во јануари компанијата заработила 100 долари, а во февруари заработила 120 долари. Ако ги претставевме резултатите како линиски дијаграм, тој исто така ќе покажуваше 100, односно 120 долари секој први во месецот, но гледајќи ја линијата, можевме да процениме дека изгледа дека на 15 јануари компанијата заработила 110 долари. А тоа не е вистина. Столпчињата се користат за дискретни мерни единици, додека линиите се користат кога имаме континуирана вредност, каква што е температурата.

Гледаме дека температурата изнесувала 20 степени во 8.00 часот, а 22 степени во 9.00 часот. Ако не прашаат да ја погледнеме линијата и да процениме колку изнесувала температурата во 8.30 часот, ќе речеме 21 степен, што би била точна проценка затоа што температурата е континуирана големина каде што некоја



Слика 6-16. Групиран сѐтолбесѐ дијагра́м



Слика 6-17. Слоевий сѐтолбесѐ дијагра́м

точка на линијата не е само збир на други вредности; таа ја претставува точната вредност во тој момент, или проценетата вредност помеѓу две прецизни мерења.

И столпчињата и линиите имаат свои варијации во слоеви (Слика 6-17). Се работи за извонредна алатка за раскажување што може да функционира на повеќе начини. Да замислиме, на пример, една компанија што има три продавници.

За секој месец имаме по три столпчиња, по едно за секоја продавница, или вкупно 36 столпчиња за цела година. Кога ќе ги поставиме едно до друго (Слика 6-16), веднаш ќе видиме која продавница заработила најмногу во определен месец. Тоа е интересна и веродостојна приказна, но има уште една приказна што е скриена во истите податоци. Ако столпчињата ги наредиме едно врз друго за да добиеме само по еден столб за секој месец, веќе не можеме да видиме толку лесно која продавница заработила најмногу, но можеме да видиме во кои месеци компанијата во целина работела најдобро.

Во двата случаи се изложени истите информации, а сепак имаме две различни приказни што произлегуваат од истите почетни податоци. Како новинари, најзначајниот аспект на работата со податоци е најпрво да одлучите која приказна сакате да ја раскажете. Дали сакате да покажете кој месец е најуспешен или сакате да покажете која продавница предводи во продажбата? Ова е само еден едноставен пример, но тука всушност лежи целиот фокус на датановинарството – да се постави вистинското прашање пред да отидете предалеку. Приказната ќе ве води до одлуката која визуелизација да ја изберете.

Столбестиот дијаграм и линискиот дијаграм навистина се „леб и сол“ за секој новинар што работи со податоци. Оттука можете да преминете на хистограми, хоризонтални дијаграми, мали хоризонтални графикони („sparklines“), дијаграми на тек итн. Сите тие имаат слични својства и одговараат на донекаде различни ситуации – вклучувајќи го и количеството податоци или извори на податоци, како и местоположбата на графичкиот приказ во однос на текстот.

Мапите се едни од најчесто користените графички решенија во новинарството. Сите мапи прикажуваат неколку заеднички елементи – времето, количеството и географската местоположба. Секогаш сакаме да знаеме колку многу од нешто има во еден или друг регион, или на кој начин податоците се движат од еден регион кон друг. Дијаграмите на тек и „хороплетите“ се корисни алатки што треба да ги користите во изработката на новинарски визуелизации. Клучно е да знаете како правилно да ги користите боите на мапата без да ги водите читателите во заблуда или погрешно да ги претставите фактите. Кај политичките мапи, определени региони се бојат со една боја, како да се работи за ситуација на „победа на сè или ништо“, дури и кога едниот кандидат победил во тој регион со разлика од само 1 отсто од гласовите. Боенето не мора да биде претставено како избор помеѓу две можности. Користењето нијанси за означување различни групи е дозволено, но мора да се изведува со големо внимание. Знаењето како се читаат и разбираат мапите е важен сегмент од новинарската работа. Мапите го олеснуваат одговорот на едно од петте прашања, прашањето КАДЕ?

Кога ќе ја совладате работата со основните видови дијаграми и графикони, можете да преминете на изработка на посложени визуелизации. Ако не ги знаете основите, изграденото ќе стои на климави темели. Слично како кога учите да станете добар пишуваач – знаете веќе, кратки реченици, фокус на читателите, без комплицирање само за да звучите паметно, работата е да му го пренесете значењето на читателот итн. – и кога работите со податоци не смеете да одите предалеку. Најдобро ќе ја раскажете приказната ако почнете со малите нешта, надградувајќи ги само кога е неопходно.

Живото пишување е концизно. Реченицата не смее да содржи непотребни зборови, во параграфот не смее да има непотребни реченици, од истите причини поради кои во цртежот не смее да има непотребни линии или

непотребни делови во машината. Тоа не значи дека писателот мора да ги крати сите реченици, или да ги дава само контурите на субјектите, избегнувајќи ги деталите. Тоа значи дека секој збор треба да кажува нешто.

– Вилијам Странк Џомладиоџи, од книгата „Стилски елементи“, 1918 [William Strunk Jr., *Elements of Style* (1918)]

Сосема е во ред ако не ги употребите сите податоци што ги имате во приказната. Не би требало да барате дозвола од некогаш за да бидете концизни. Концизността би требало да биде правило.

– Брајан Суда (Brian Suda), *optional.is*

„Направи сам“ визуелизација: Најдобрите алатки според новинарите на „Гардијан“

Кои алатки за визуелизација на податоците - лесни за користење и бесплатни - се достапни на интернет? Ние, во „Датаблогот“ и во „Датастор“ (Datastore, „продавница за податоци“, <http://www.theguardian.com/data>) се трудиме што повеќе да ги користиме моќните бесплатни и слободни можности што се нудат на интернет.

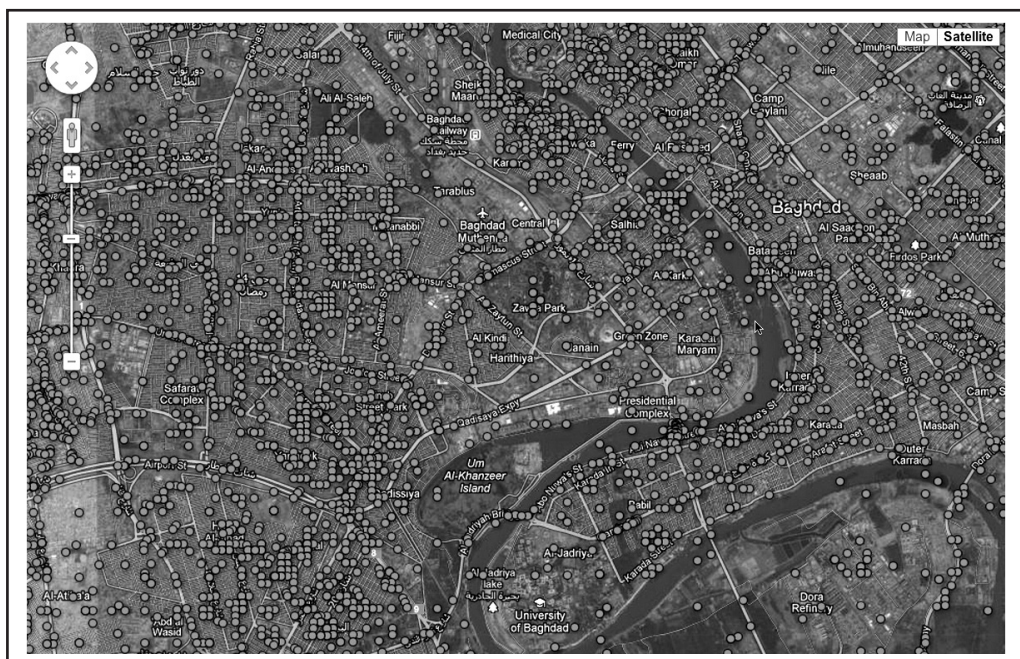
Тоа може да звучи помалку неискрено, знаејќи дека имаме пристап до извонредните тимови за графичка обработка и интерактивни прикази на „Гардијан“, особено кога имаме доволно време – како, на пример, во изработката на мапата на јавната потрошувачка [<http://bit.ly/guardian-spending>; изработена со „Адоуб илустратор“ (Adobe Illustrator)] или интерактивниот приказ на немирите на „Твитер“.

Сепак, во секојдневната работа, често се служиме со алатки што се достапни за сите и изработуваме графички решенија што сите можат да ги изработат.

Значи, кои алатки ги користиме?

Табелите на „Гугл Фјужн“ (Google Fusion Tables)

Оваа онлајн-алатка за работа со бази на податоци и мапирање полека прерасна во алатка што по автоматизам ја користиме кога треба набрзина да произведеме детални мапи, особено ако мапите треба да ја нудат и можноста за зумирање. Ја имате на располагање високата резолуција на „Гугл мепс“ (Google Maps) заедно со способноста да носи многу податоци, на пример, CSV-датотека со големина до 100 МВ (мегабајти). На прва рака, табелите во „Фјужн“ можат да изгледаат збркани и конфузни, но бидете упорни. Ние ги користевме за изработка на мапи какви што се мапата од војната во Ирак на Слика 6-18 и мапата на регионалната распределба на бездомништвото на Слика 6-19.



Слика 6-18. Воениџе дневници на „Викиликс“ („Гарџијан“)

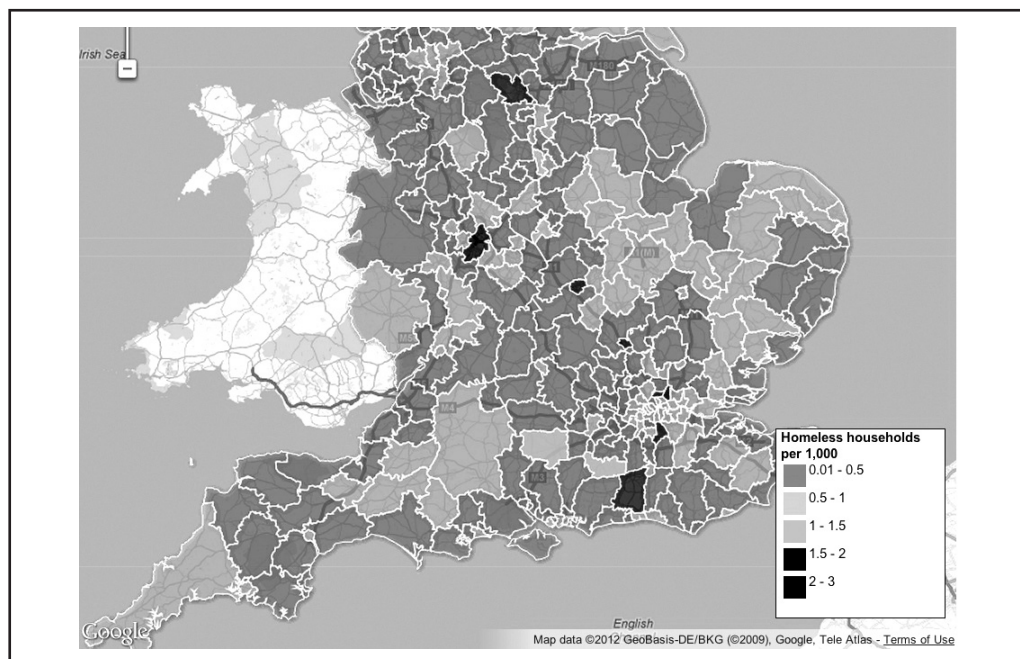
Основната предност што ја нудат е флексибилноста – можете, на пример, да внесете „KML“-датотека со границите на регионите или окупациите и со неа да ги соедините податоците од некоја табела. Во моментот се работи на нов кориснички интерфејс кој би требало да го олесни користењето.

Не мора да сте програмер за да изработите мапа – а алатката за работење во слоеви на „Фјуџн“ (Fusion layers tool) дозволува да интегрирате повеќе различни мапи или да додадете можност за пребарување и филтрирање, а потоа мапата да ја вградите во блогерски пост или на веб-страница.

На почетниците од голема помош ќе им бидат одличните упатства за користење (<http://bit.ly/fusiontables-tutorial>) што ги подготви Катрин Харли (Kathryn Hurley) од „Гугл“.



Со помош на наредбата „shpescape“ трансформирајте ги официјалните .shp датотеки во „Фјуџн“-табели што можете да ги користите понатаму. Избегнувајте ги премногу комплицираните мапи – „Фјуџн“ може да се справи со не повеќе од милион точки по ќелија.

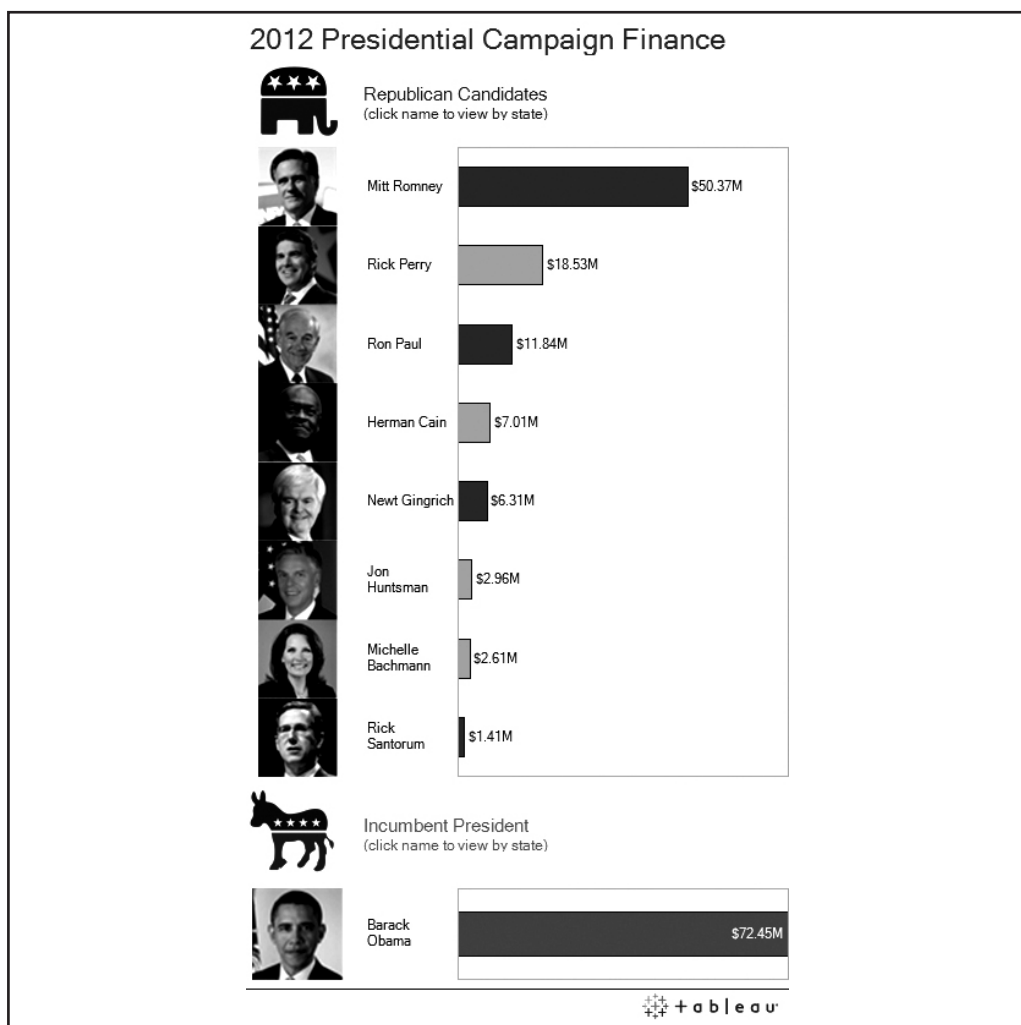


Слика 6-19. Мапа на бездомниџе семејства („Гардијан“)

„Табло паблик“ (Tableau Public)

„Табло паблик“ е бесплатен за користење, освен ако не ви треба неограничениот капацитет што го нуди неговата професионалната верзија. Овозможува изработка на мошне сложени визуелизации, што користат до 100.000 редови податоци, на лесен и едноставен начин. Ние го користиме кога треба да соединиме различни графикони, како во примерот на мапата за највисоките даночни стапки во светот (<http://bit.ly/guardian-top-tax>; мапата користи и столбест дијаграм).

Можете да го користите и за прегледување податоци, како што сторивме ние со податоците за трошоците на сојузните избори во САД, прикажани на Слика 6-20 (<http://bit.ly/guardianelections-us>; во еден момент дури го искористивме целиот мемориски капацитет што го нуди бесплатната јавна верзија... нешто на што мора постојано да внимавате). „Табло“, исто така, бара податоците да бидат форматирани на посебен начин за да можете да го извлечете најдоброто од нив. Кога и тоа ќе го завршите, ќе имате одлична, функционална и интуитивна алатка. Аргентинскиот весник „Ла насион“, на пример, целосно се потпира на „Табло“ во сите активности на полето на дата-новинарството.



Слика 6-20. Финансирањето на кампањата за време на Прејседателските избори 2012 година („Гардијан“)

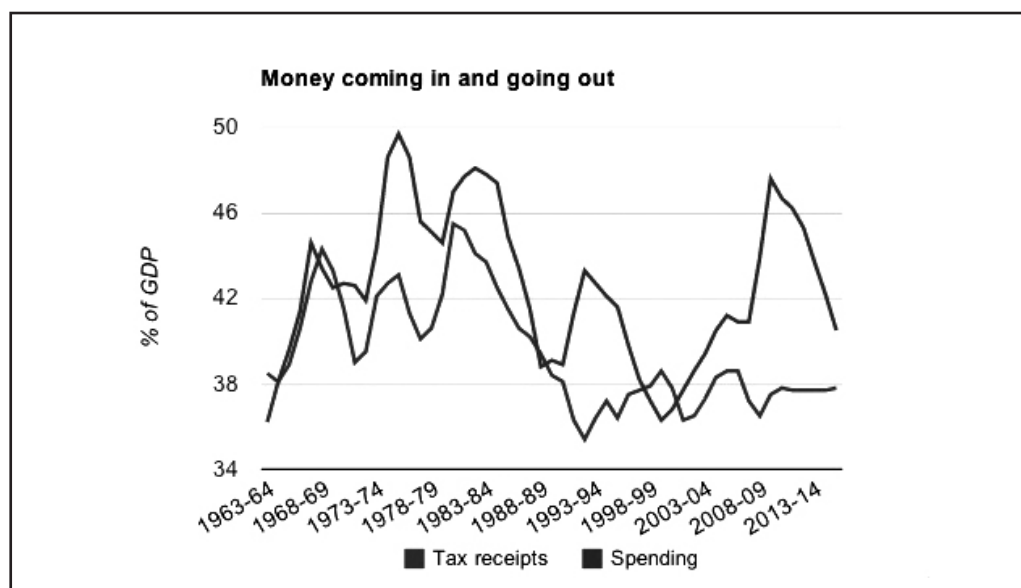
„Табло“, исто така, нуди одлични упатства за користење за почетници, достапни на <http://www.tableausoftware.com/learn/training>.



„Табло“ е наменет за „PC“ компјутери, а во моментот се работи и на верзија за „Мак“ (Mac) на „Епл“ (Apple). Ако работите на „Мак“, можете да користите некоја од огледалните страници (mirror-site).

Дијаграмите на „Гугл спредшит“ (Google Spreadsheet Charts)

Алатката е достапна на адресата <http://www.google.com/google-d-s/spreadsheets/>



Слика 6-21. Јавната појрошувачка и собирањето на даноци во Велика Британија („Гардијан“)

Освен едноставните графикони и дијаграми (столбести и линеарни дијаграми, или секторски (пита) дијаграми), ќе откриете дека сметководствените табели на „Гугл“ (што ќе ги изработите во секцијата за работа со документи на вашиот кориснички налог) можат да произведат мошне привлечни графички решенија – вклучувајќи ги и анимираните меурчиња што Ханс Рослинг (Hans Rosling) ги користеше за неговиот „Гепмајндер“ (Gapminder, <http://www.gapminder.org/>). За разлика од интерфејсот за програмирање графички апликации (API), не мора да се грижите за програмерските аспекти или за пишување компјутерски код. Како и кај „Excel“, само ги означувате податоците и кликувате на „виџетот“ за креирање графикони. Не е лошо и да ги разгледате можностите за корисничко приспособување, каде што можете да ги менувате боите, насловите и размерите. Дизајнот е неутрален, што е корисно, особено кај помалите графикони и дијаграми. И линиските дијаграми нудат неколку одлични опции, вклучувајќи ја и можноста да додадете свои коментари.



Посветете и малку време на опциите за корисничко приспособување на графиконите за да изработите сопствена палета на бои.

„Датамаркет“ (Datamarket)

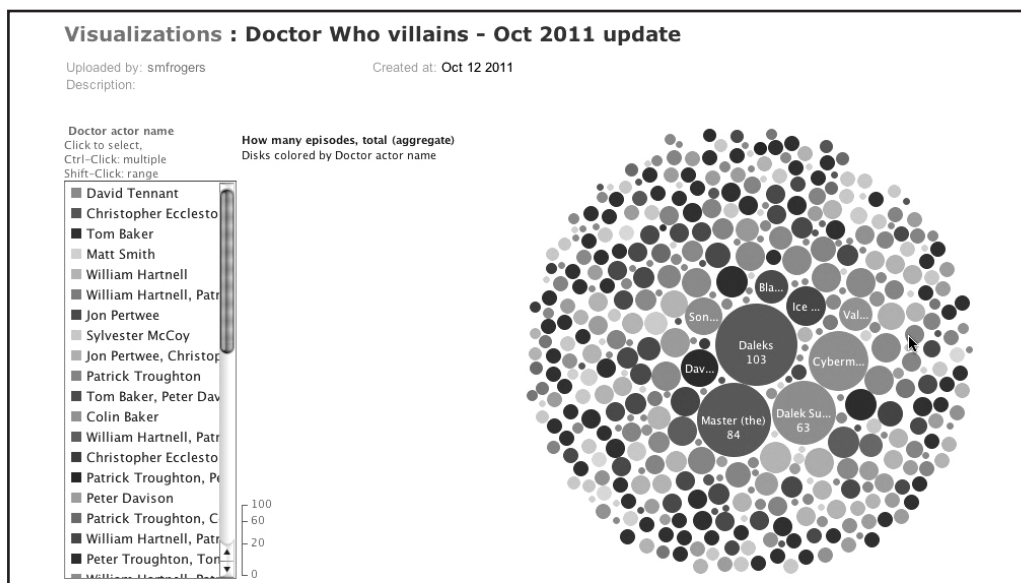
Иако е попознат како алатка за обезбедување податоци, „Датамаркет“ истовремено е и сосема добра алатка за визуелизација на бројки (<http://bit.ly/datamarket-explore>). Можете да внесете сопствени податоци или некои од многубројните групи на податоци од понудата на „Датамаркет“. Секако, на располагање ќе имате многу повеќе и подобри опции ако отворите кориснички налог за професионалци.



„Датамаркет“ работи најдобро со податоци распоредени во временски серии. Сепак, проверете ја и богатата понуда на податоци на нивната веб-страница.

„Мени ајс“ (Many Eyes)

Ако треба да избереме веб-страница што изгледа запуштено, тоа секако ќе биде „Мени ајс“ (буквалниот превод гласи „многу очи“) на Ај-Би-Ем (IBM). Изработена од Фернанда Б. Виегас (Fernanda B. Viégas) и Мартин Ватенберг (Martin Wattenberg), во моментот кога почна со работа претставуваше единствена можност за едноставно поставување и визуелизирање податоци. Денес неговите автори работат за „Гугл“, а веб-страницата, со придушените бои, изгледа запоставена и веќе подолго време не нуди ништо ново на полето на визуелизацијата.



Слика 6-22. Непријателите на Доктор Ху (Doctor Who, лик од популарната истоимена научно-фантастична телевизиска серија, забележ.) („Гардијан“)



Морате да бидете сигурни дека сè што сакате, затоа што по поставувањето на податоците, нема да можете да ги дополнувате или да внесувате исправки.

„Готвач за бои“ (Color Brewer)

„Color Brewer“ (<http://colorbrewer2.org/>) не е алатка за визуелизација во вистинската смисла на зборот, туку повеќе е алатка што овозможува бирање на боите што ќе ги користите на вашата мапа. Вие ја бирате основната боја и „Color Brewer“ ќе ви ги даде шифрите за сите бои во саканата палета.

Уште неколку алатки

Ако ниту една од претходно наведените алатки не одговара на вашите потреби, можете да ја прочитате оваа статија на „Дејли тек“ (DailyTekk, <http://bit.ly/dailytekk-infographic>) која нуди уште повеќе можности. Алатките наведени погоре не се единствените, туку се алатките што најчесто ги користиме. Постојат многу други алатки, на пример:

- „Чартсбин“ (Chartsbin, <http://chartsbin.com>), алатка за изработка на мапи на светот во кои може да се работи со компјутерски глушец (clickable).
- „АјЧартс“ (iCharts, <http://www.icharts.net>), специјализирана алатка за мали интернет-уреди (widgets) за работа со графикони
- „Геокомонс“ (Geocommons, <http://geocommons.com>), алатка за размена на податоци и податоци за границите за изработка на глобални и локални мапи
- Тука е и веб-страницата piktochart.com, со готови обрасци за најпопуларните визуелизации на текст или броеви во определен момент.

— Сајмон Роџерс, „Гардијан“

Како „Верденс ганг“ ги прикажува податоците

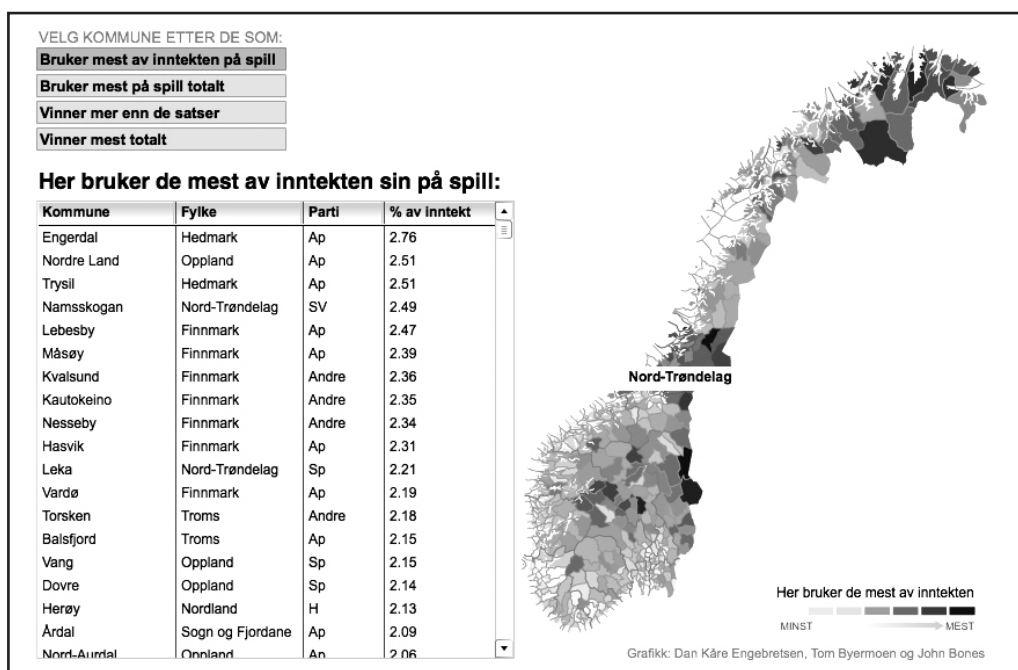
Суштината на информативното новинарство е на читателите да им се пренесат најновите информации на најбрзиот можен начин. Целта – брзото пренесување на информациите – е иста и за визуелизациите. Новите алатки за работа со податоци им овозможуваат на новинарите да откријат приказни што инаку не би можеле, како и да ги пренесат приказните на нови начини. Следат неколку примери за тоа како ние, во најтиражниот и најчитаниот дневен весник во Норвешка „Верденс ганг“ [Verdens gang (VG)], ги „послужуваме“ податоците за нашите читатели.

Броеви

Приказната што можете да ја пронајдете на линкот <http://bit.ly/vg-lotto> како основа ги има податоците на Норвешкиот завод за статистика, податоците за даночните обврзници и податоците од националниот монопол за организирање лотарија и игри на среќа. Интерактивниот графикон им нуди на читателите различни информации за сите општини и окрузи во Норвешка. Табелата што служи како основа за графиконот покажува колкав процент од своите приходи трошат граѓаните на игри на среќа. Во изработката користевме „Access“, „Excel“, „MySql“ и „Flash“.

Мрежи

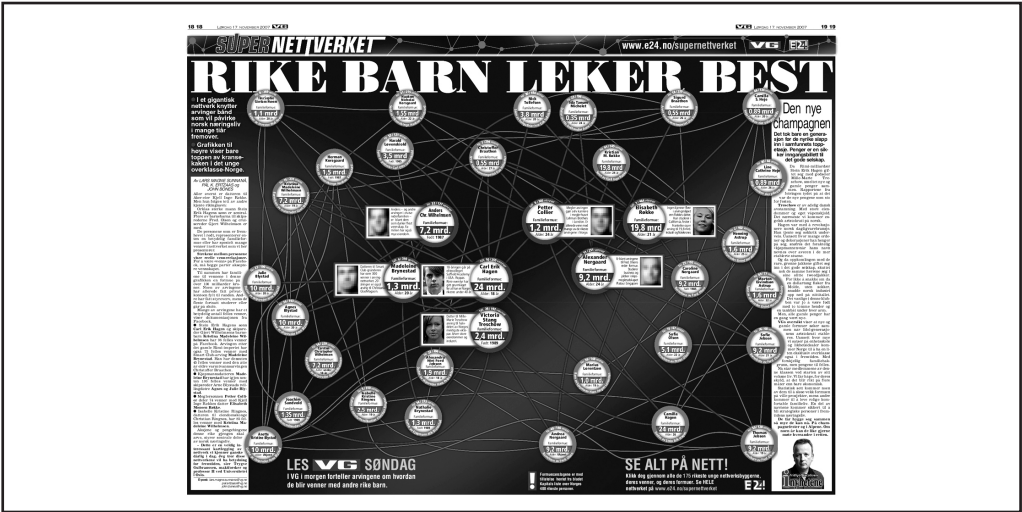
Преку анализа на социјалните мрежи ги прегледавме меѓусебните врски на 157 синови и ќерки на најбогатите луѓе во Норвешка. Анализата покажа дека наследниците на најбогатите Норвежани од родителите ги наследуваат и своите врски во општеството. Сè на сè, идентификувавме повеќе од 26.000 врски, а графичките решенија беа доработени рачно со користење „Фотошоп“ (Photoshop). Во изработката користевме „Access“, „Excel“, „Notepad“ и алатката за анализа на социјалните мрежи „Јусинет“ (Ucinet, <https://sites.google.com/site/ucinetsoftware/>).



Слика 6-23. Майирање на податоциите на даночните обврзници и податоциите на државната лотарија („Вергенс ганг“)

Мапи

На анимираната термичка мапа, комбинирана со едноставен столбест дијаграм, можете да ги видите, час по час, кривичните дела извршени во центарот на Осло за време на викендите во неколку последователни месеци. На дното на мапата можете да го видите и бројот на полицајци на должност во моментот на вршење некое кривично дело. Во изработката го користевме софтверот „Арквју“ (ArcView, <http://www.esri.com/software/arcgis/arcview>) заедно со алатката за просторна анализа „Спејшал аналитист“ (Spatial Analyst, <http://www.esri.com/software/arcgis/extensions/spatialanalyst>).



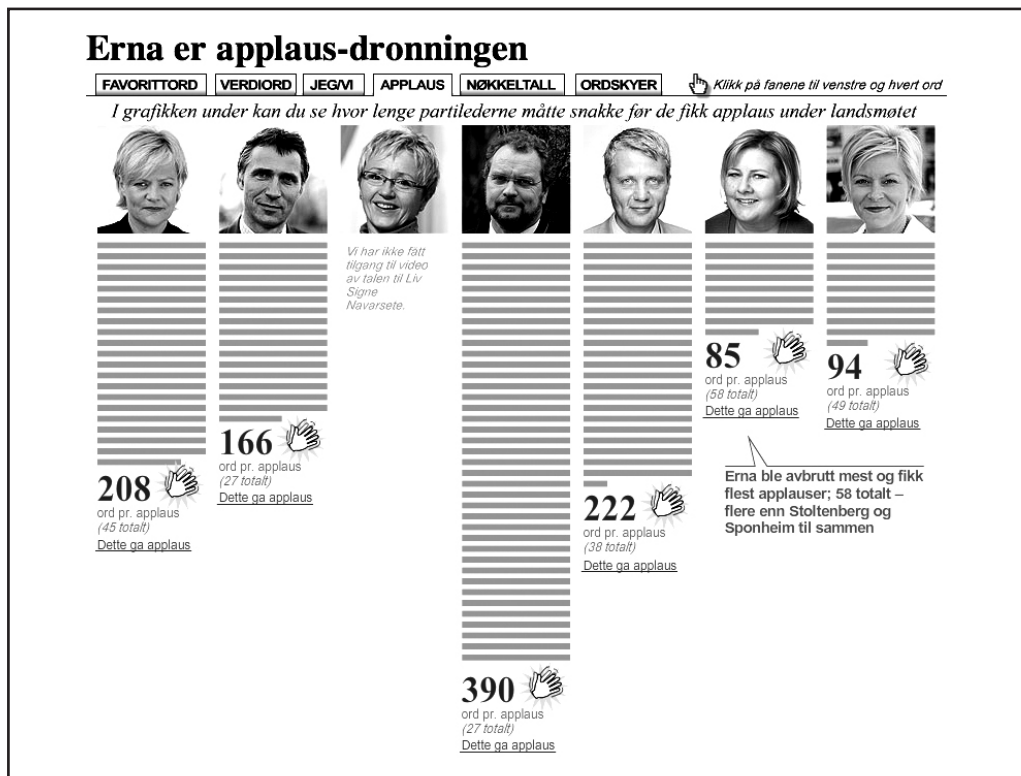
Слика 6-24. Богајишките „џиџики“ одај во своето јато („Верденс ганг“)



Слика 6-25. Анимирана термичка мапа („Верденс ганг“)

Анализа на текст

За долната визуелизација ги анализираме говорите на партиските конвенции на лидерите на седум политички партии во Норвешка. Анализата на говорите обезбеди повеќе агли на гледање од кои произлегоа неколку приказни. Сите приказни беа поврзани со графиконот што на читателите им овозможи да го проучат јазикот што го користат политичарите. Користевме „Excel“, „Access“, „Flash“ и „Illustrator“. Да го подготвувавме графиконот во 2012 година, најверојатно ќе изработевме интерактивен графикон во „JavaScript“.



Слика 6-26. Анализа на говорите на норвешките лидери („Верденс жанџ“)

Заклучоци

Кога навистина е потребно да визуелизираме една приказна? Во повеќето случаи нема потреба, па сепак го правиме тоа за да им помогнеме на читателите. Визуелизацијата често е потребна кај приказните што содржат големи количества податоци. Одлуката кои податоци ќе ги прикажеме, од нас бара

да бидеме строги во проценката. Ние може да знаеме сè и сешто кога извествуваме, но што е тоа што е поврзано со приказната што читателите навистина треба да го знаат? Можеби ќе биде доволно да понудите само табела, или едноставен дијаграм што го покажува развојот од годината А до годината В. Поентата на дата-новинарството не мора да биде да се прикаже огромно количество податоци. Поентата е во новинарството!

Во последниве две до три години постои изразен тренд за креирање интерактивни графикони и табели што ќе им овозможат на читателите да „копаат“ по различни теми. Добрата визуелизација е како добра илустрација. Ви треба само секунда или две за да разберете за што се работи. Што подолго гледате во визуелизацијата, повеќе ќе видите. Ако читателите не знаат од каде да почнат и кога да престанат, ако визуелизацијата е преоптоварена со детали, тогаш имаме лоша визуелизација. Можеби обичен краток текст подобро би ја завршил работата?

– Џон Боунс, „Верденс џанџ“ (John Bones, Verdens Gang)

Јавните податоци стануваат општествено прашање

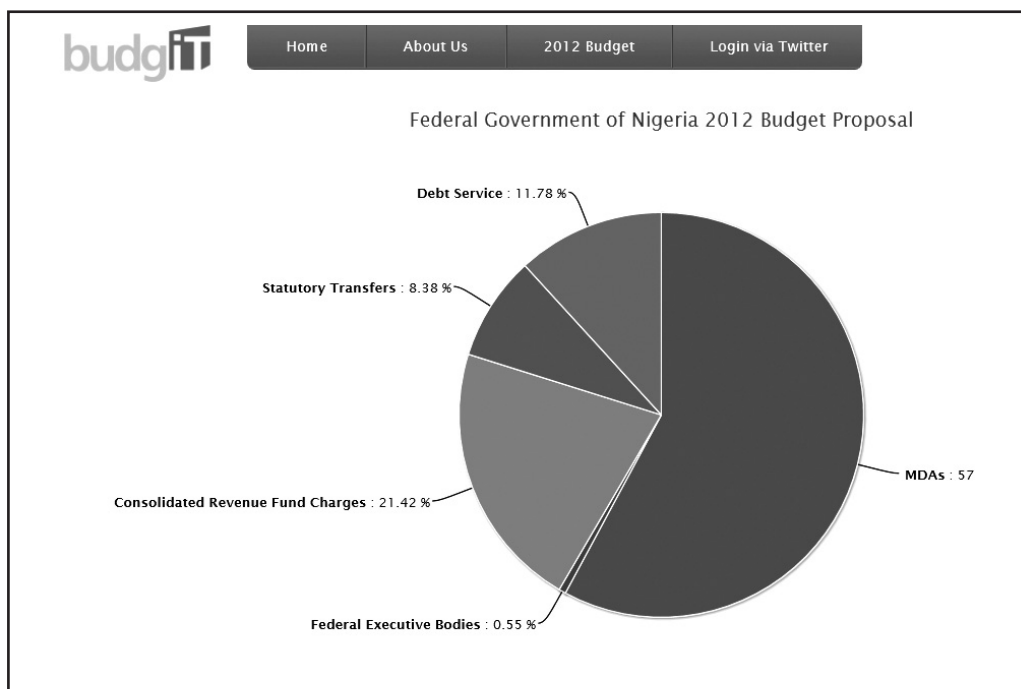
Податоците се од непроценливо значење. Пристапот до податоците може да ги расветли прашањата на начин што ќе доведе до резултат. Сепак, лошото ракување со податоците може од фактите да создаде непроѕирна структура што не пренесува никаква порака и низ која не може да се проникне. Податоците што нема да успеат да поттикнат јавна дискусија или да обезбедат разбирање на контекстот, имаат само ограничено значење за јавноста.

По многу години под воена власт, демократијата во Нигерија беше обновена во 1999 година. Обидот да се проникне во фактите што се кријат зад податоците се сметаше за форма на отпор кон власта и за обид за оспорување на и онака извалканата репутација на хунтата. Законот за службени тајни ги обврзуваше јавните службеници да не ги споделуваат владините информации. Дури и денес, 13 години по обновата на демократијата, понекогаш е тешко да се пристапи до јавните податоци. На мнозинството граѓани што немаат скоро никакви познавања на полето на финансиското сметководство или сложените аритметички операции, податоците за јавната потрошувачка не им значат ништо.

Порастот на употребата на мобилните уреди и сè поголемиот број корисници на интернет во Нигерија, за нас, во организацијата „БаџИТ“ (BudgIT, се работи за игра на зборови – „BudgIT“ се изговара исто како и „Budget“, „буџет“, заб.прев.; <http://www.yourbudgit.com/>), отвори извонредна можност за користење на технологијата за визуелизација на податоци за да им објасниме на луѓето за што се ра-

боти и да ги поттикнеме да се ангажираат во поглед на прашањето на јавната потрошувачка. За да го постигнеме тоа, требаше да стигнеме до корисниците на сите платформи и да стигнеме до граѓаните преку невладините организации. Целта на проектот е од јавните податоци да направиме предмет од општествен интерес и да изградиме широка мрежа што ќе се залага за општествени промени.

Ако сакаме успешна соработка со корисниците, мораме да знаеме што е тоа што тие го сакаат. Што сакаат граѓаните на Нигерија? На кој начин го чувствуваат влијанието на информатичкиот јаз? Како да постигнеме податоците да станат релевантни за нив? Во моментот, нашата цел во BudgIT се просечните писмени Нигеријци што се присутни на форумите на интернет и на социјалните медиуми. За да го привлечеме ограниченото внимание на корисниците што се внесени во своите разновидни интереси (игри на среќа, читање, дружење), ќе треба податоците да ги прикажеме кратко и концизно. Објавувањето кадар (snapshot) од податоците на „Твитер“ или во форма на инфографика, отвора можност за посуштински ангажман со поинтерактивни искуства кои на корисниците ќе им ја понудат пошироката слика.



Слика 6-27. Апликација за распределба на буџетот во Нигерија (BudgIT Nigeria)

Кога визуелизираме податоци, значајно е да го знаеме нивото на писменост во областа на работата со податоци што го поседуваат нашите корисници. Ви-

зуелизациите можат да бидат убави и софистицирани колку што сакате, но ако корисниците немале претходно искуство во толкување податоци, тие можеби нема да препознаат никаква порака во комплексните дијаграми и интерактивните апликации. Добрата визуелизација им зборува на корисниците на јазик што го разбираат и нуди приказна со која лесно се идентификуваат.

За работата со националниот буџет ангажиравме повеќе од 10.000 Нигеријци, што ги поделивме во три категории за да се осигураме дека од нив ќе ја добиеме оптималната вредност. Следи кратко објаснување за секоја од трите категории:

Повремени корисници

Во оваа категорија спаѓаат корисниците што сакаат брза и кратка информација. Сакаат да видат слика од податоците, но не и детални анализи или статистики. Со нив комуницираме преку „твитови“ или интерактивни графики.

Активни корисници

Корисници што се вклучуваат во дискусиите и ги користат податоците за да стекнат нови знаења во некоја област или да ги оспорат претпоставките што произлегуваат од податоците. На таквите корисници сакаме да им понудиме механизми за давање повратни информации и можност за споделување на нивните сознанија со сличните на нив преку социјалните мрежи.

Собирачи на податоци

Корисниците од оваа група бараат податоци за визуелизација или анализа. Ним едноставно им ги даваме податоците за да ги користат како што сакаат.

Ангажманот на корисниците на „BudgIT“ се заснова на следното:

Пополнување расправата за јавноесекторските трендови

„BudgIT“ ги следи дискусиите на интернет и во реалниот свет со цел да ги поткрепи со податоци за темите за кои се разговара. На пример, во случајот на штрајковите поврзани со цената на горивото во јануари 2012 година, демонстрантите постојано ја спомнуваа потребата од повторно воведување субвенции за бензинот и намалување на непотребната и екстравагантна јавна потрошувачка. Ја следевме дискусијата на социјалните медиуми и, по 36 часови напорна работа, на граѓаните им понудивме апликација што им овозможуваше да го реорганизираат државниот буџет на Нигерија.

Добри механизми за собирање повратни информации

Корисниците ги вклучуваме во нашите активности преку дискусии канали и преку социјалните медиуми. Многу корисници сакаат да дознаат кои приказни се кријат во податоците и сакаат да го слушнат нашето мислење. Се трудиме во одговорите на нивните прашања да понудиме директно објаснување за фактите зад податоците, без влијание на нашите лични или политички ставови. Каналите за повратните информации треба да останат отворени, треба активно да одговараме на коментарите и да бидеме креативни во комуникацијата со корисниците за да обезбедиме одржливост на заедницата што ја изградивме околу податоците.

Дејствиуваме локално

Кога имаме податоци наменети за определена група корисници, ние во BudgIT се трудиме да ги локализираме содржините и да понудиме канал за дискусија што се однесува на потребите и интересите на одделни групи корисници. Особено сме заинтересирани за можноста да ги инволвираме корисниците со користење на пораки по мобилен телефон (SMS).

По објавувањето на податоците за јавната потрошувачка на yourbudget.com, им се обраќаме на граѓаните преку различни невладини организации. Планираме да развиеме партиципаторна рамка, во која граѓаните и владините институции ќе се среќаваат во зградите на општините за да ги дефинираат клучните прашања на кои треба да им се даде приоритет во буџетирањето.

За проектот известуваа и локалните и странските медиуми, од „Селебратинг прогрес Африка“ (Celebrating Progress - CP-Africa, <http://bit.ly/cp-africa-budget>) до Би-Би-Си (<http://bbc.in/africa-budget>). За Јинка Ибукун (Yinka Ibukun), новинар на „Асошијетед прес“ (Associated Press), изработивме ревизија на буџетите за безбедносниот сектор за периодот од 2002 до 2011 година. Повеќето медиумски организации се „собирачи на податоци“ и често бараат да им доставиме податоци за нивното известување. Планираме и во месеците што доаѓаат да ја продолжиме соработката со новинарите и новинските организации.

– Олусеун Онигбинде (Oluseun Onigbinde), „BudgIT“ Ниџерија

Како да ги заинтересирате луѓето за вашите податоци

Реакцијата на вашата публика е скоро подеднакво значајна како и самото објавување на податоците. И вие сте само луѓе. Тоа значи дека од време на време ќе правите грешки, нешто ќе пропуштите, за нешто ќе стекнете погрешна претстава. Вашата публика е еден од најкорисните ресурси што ги имате на располагање. Тие можат да ги проверат фактите и да ви посочат нешто што можеби сте го пропуштиле во работата.

Сепак, ни најмалку не е лесно да ја натерате публиката да се вклучи во работата. Си имате работа со луѓе што долгогодишната употреба на интернет ги условила да скокаат од една веб-страница на друга, оставајќи зад себе само саркастични коментари. Клучна е изградбата на довербата помеѓу вас и вашите корисници. Тие треба да знаат што им нудите, како можат да реагираат и да ви понудат повратни информации, и треба да ги убедите дека таквите повратни информации ќе бидат земани сериозно.

Сепак, најпрво треба да размислите каква публика имате, или каква публика сакате да имате. Тоа ќе влијае врз податоците со кои работите, но истовре-

мено ќе трпи и влијание од нив. Ако вашата публика е ограничена на определен сектор, ќе треба да ги истражите специфичните начини на комуникација со тој сектор. Постојат ли струкови тела на кои можете да им пристапите, а кои би биле подготвени да ги претстават вашите ресурси и тоа што го работите на некоја поширока публика? Постои ли некој форум или веб-страница на заедница со која можете да стапите во контакт? Постојат ли специјализирани стручни изданија што би известувале за приказните што ги откривате во податоците?

Социјалните медиуми, исто така, се значајна алатка, иако тоа повторно зависи од видот на податоци со кои работите. Ако, на пример, се интересирате за статистичките податоци за глобалниот поморски транспорт стоки, мала е веројатноста дека на „Фејсбук“ или на „Твитер“ ќе пронајдете група на членови заинтересирани за тоа што го работите. Од друга страна, ако ги просејувате индексите на корупција во светот, или локалните криминални статистики, таквите активности веројатно ќе побудат интерес кај многу поширока публика.

Кога станува збор за „Твитер“, се покажува дека најдобар пристап е да контактирате некоја позната личност, накратко да ѝ објасните зошто е значајно тоа што го работите и да ѝ испратите линк до вашата веб-страница. Со малку среќа, таа ќе ве „ретвитуира“ па за вас ќе дознаат и нејзините читатели. Тоа е одличен начин со минимален труд да обезбедите максимална видливост за она што го работите. Сепак, внимавајте, не додевајте им премногу на луѓето!

Кога вашата веб-страница ќе почне да привлекува публика, треба да почнете да мислите како публиката ќе се однесува кон она што го работите. Навистина, луѓето можеби ќе ја прочитаат приказната што сте ја напишале, ќе ги разгледаат инфографиките и мапите, но од особено значење е да им дадете можност да реагираат. Ништо не може да ви даде подобар увид во темата за која пишувате или подобри информации за идната работа на тоа поле.

Прво, се подразбира дека треба да ги објавите и суровите податоци заедно со статиите што произлегле од нив. Податоците чувајте ги како „CSV“-датотеки со обичен текст (Comma-Separated Values plain text), или на некој независен сервис каков што е „Google Docs“. Така ќе обезбедите дека ќе има само една верзија на податоците што можете да ја ажурирате по потреба, ако откриете грешки во податоците што треба да се исправат. Всушност, најдобро ќе биде да ги чувате и на двете места. Олеснете го, колку што можете, пристапот до суровите информации.

Потоа размислете дали има други начини за подобрување на комуникацијата со публиката. Проверете ги статистиките за користењето на вашата веб-страница за да видите кои сегменти се најпосетени – местата што привлекуваат најмногу сообраќај може да кријат нешто што претходно сте го пропуштиле. На пример, можеби вие не сте им обрнале внимание на статистичките податоци за сиромаштијата на Исланд, но ако многумина се интересираат за тие делови од табелата, можеби во нив има нешто што вреди да се погледне.

Размислувајте што можете да направите а да не се работи само за отворање на можноста за коментирање. Можете ли да прикачите коментари до индивидуални ќелии во сметководствената табела? До определен регион на инфографиката? Иако повеќето вметнати системи за објавување не ја нудат таквата можност, не е лошо да видите дали нешто може да се направи ако работите „по мерка“. Не е можно да се потценат придобивките што тоа ќе ги донесе за вашите податоци.

Овозможете и другите корисници да можат да ги прочитаат тие коментари. Тие, во многу случаи, се еднакво вредни колку и изворните податоци. Од друга страна, ако одлучите таквите информации да ги чувате за себе, истовремено ја лишувате публиката од нивната вредност.

Најпосле, можеби некој друг ќе сака да објави своја инфографика и приказна врз основа на истите извори на податоци. Треба да размислите како да ги поврзете таквите луѓе меѓу себе и како ќе ги опишете нивните трудови. На пример, можете да користите некој „хаштаг“ (#, „тараба“) што е специфичен за таа група податоци, а ако се работи за многу фотографии и илустрации, тогаш можете да го споделите преку корисничка група на „Фликр“ (Flickr).

Може да се покаже корисно да пронајдете начин што ќе обезбеди поголема тајност на размената на информации. Понекогаш јавното споделување податоци може да ги доведе во опасност оние што дале свој придонес. Можеби оние што сакаат да соработуваат со вас не сакаат тоа да го прават јавно. Тие можеби ќе претпочитаат да ги достават своите информации преку електронска пошта, или дури преку анонимни коментари.

Најважно од сè е вашите податоци да ги споделувате и разменуваат што повеќе и со што повеќе луѓе. Користа за вашата работа како новинар и за читателското искуство на вашата публика – ако на читателите им дозволите да го проверуваат вашиот труд, да ги пронајдат грешките што сте ги направиле, да пронајдат нешто што сте пропуштиле – ќе биде огромна.

— Данкан Гип, *Wired.co.uk*

Белешка за уредниците

Џонатан Греј е Директор на одделот за заедници и култура на Фондацијата „Отворено знаење“ (<http://okfn.org>), наградувана непрофитна организација што работи на промоција на отворените податоци, отворените содржини и јавното добро во повеќе различни области. Греј е основач на неколку проекти на полето на дата-новинарството на Фондацијата, вклучувајќи ги и OpenSpending.org, проект за мапирање на јавната потрошувачка во светот, и „Европска енергија“, проект што обезбедува контекст за енергетските цели на Европската Унија. Во моментот работи на истражување на филозофијата и историјата на идеите на колеџот „Ројал Холовеј“ на Универзитетот на Лодон (Royal Holloway, University of London). Повеќе информации можете да пронајдете на неговата веб-страница jonathangray.org.

Лилијана Бунегру е уредник на веб-страницата DataDrivenJournalism.net и проектен менаџер за дата-новинарство на Европскиот центар за новинарство (<http://www.ejc.net/>). Таа е координатор на „Наградите за дата-новинарство“ (Data Journalism Awards) и е дел од тимот што го уредува „Прирачникот за дата-новинарство“ (Data Journalism Handbook). Лилијана, исто така, работи како истражувач на истражувачкиот проект „Електронски мапи за помош на јавната наука“ (Electronic Maps to Assist Public Science – EMAPS) и е кандидат за истражувач на магистерските студии од областа на медиумите на Универзитетот на Амстердам. Нејзината истражувачка работа е фокусирана на влијанието што технологијата го има на медиумите, културата и општеството. Нејзиниот блог е достапен на адресата lilianabounegru.org.

Луси Чејмберс е координаторка на заедниците во Фондацијата „Отворено знаење“. Таа работи на проектот Open Spending.org и на „Приказни за потрошувачката“ [Spending Stories, добитник на наградата „Најт њус челинџ“ (Knight News challenge) за 2011 година], проект што им нуди помош на новинарите во обезбедувањето контекст и во проверката на податоците за јавната потрошувачка. Луси ги координира активностите на Фондацијата на полето на новинарството водено од податоци и организира сесии за обука на новинарите како да ги пронаоѓаат, обработуваат и презентираат податоците.

Белешка за координаторите на проектот

Европскиот центар за новинарство (<http://www.ejc.net>) нуди тренинзи за подобрување на квалитетот на известувањето за актуелните прашања во Европа и стратегиска поддршка за европските медиуми.

Фондацијата „Отворено знаење“ (<http://okfn.org>) се залага за свет во кој отвореното знаење е сеприсутно и рутинско – и на интернет и во реалниот свет – и работи на промоција на отвореното знаење поради неговиот потенцијал да донесе далекусежни придобивки за општеството.

Како новинаријте да ги искористат податоцијте за квалитетно известување

Прирачник за новинарство засновано на податоци

Приреден од Џонатан Греј, Лилијана Бунегру и Луси Чејмберс



**ЦЕНТАР за
РАЗВОЈ на
МЕДИУМИ**