

COMPUTATION + JOURNALISM IN THE PUBLIC INTEREST

Sarah Cohen, DeWitt Wallace Center

Public interest reporting defined

- Information of importance to the public that powerful institutions would prefer to be kept hidden or secret.
- A method of comparing what *ought* to happen with what *actually* happens
 - ▣ Breaking laws or rules
 - ▣ Taking advantage of the helpless
- Comforting the afflicted and afflicting the comfortable.

Common traits

Reporting shares some traits with other investigative or exploratory fields :

- Use of multiple sources with varying accuracy and collection methods
- Practitioners have little or no technical ability or interest
- Much of the data is in the form of unstructured text
- Start with broad, ill-defined questions, not discrete goals
- Combines field work and lab work (street reporting and data or document analysis)

Uncommon traits

- Rarely spend much time on a given data source or question
- Reporters look for tips and examples, not evidence or statistical patterns.
- Work from primary sources
- Analysis is not mission-critical
- Unique identifiers are censored.
- Little or no control over the form of the data: it's up to the source

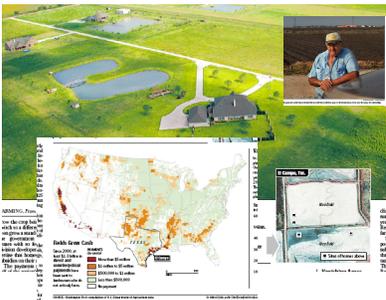
Typical data sources

- ▣ Interview notes, both on the record and for background only. Very little is taped, especially early in a story.
- ▣ Taped or videotaped hearings, sometimes transcribed already
- ▣ Calendars, correspondence, investigative and inspection reports
- ▣ Emails and press releases
- ▣ Excel spreadsheets created for printing, not analysis
- ▣ Structured databases
- ▣ Court documents: complaints, indictments, settlements, discovery.
- ▣ Online document collections, such as regulatory enforcement actions or audits.
- ▣ Handwritten forms, such as inspections or police incident reports, usually heavily redacted

Transparency myths

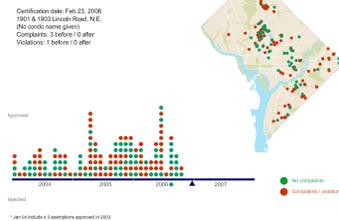
- ▣ Government records are easy to get in a usable form? Not so much.
- ▣ Examples, even when it is:
 - ▣ [Recovery Act](#)
 - ▣ [Campaign finance](#)
 - ▣ [Crime](#)
- ▣ Most government information is still hidden, secret or difficult to access
 - ▣ data.gov: Promise of 100,000 datasets. Currently have 728, half of which are annual state-level cuts of existing EPA data that has been available for years. No records, mainly just aggregated statistics. Almost nothing on it that wasn't already readily available.

Harvesting Cash

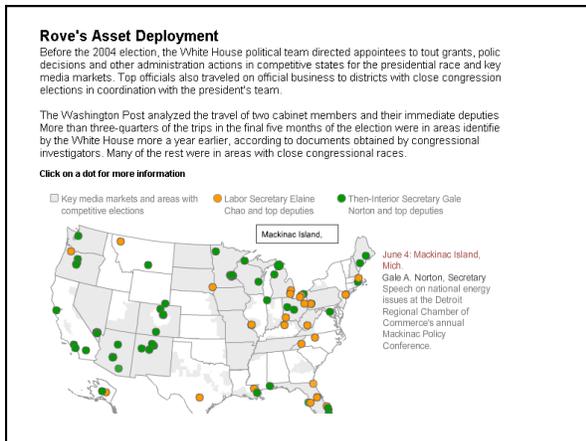
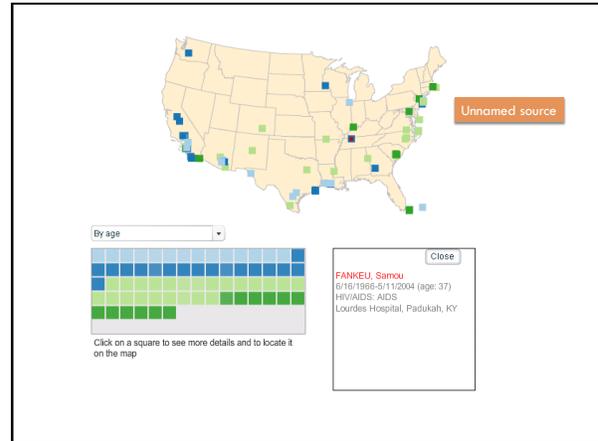
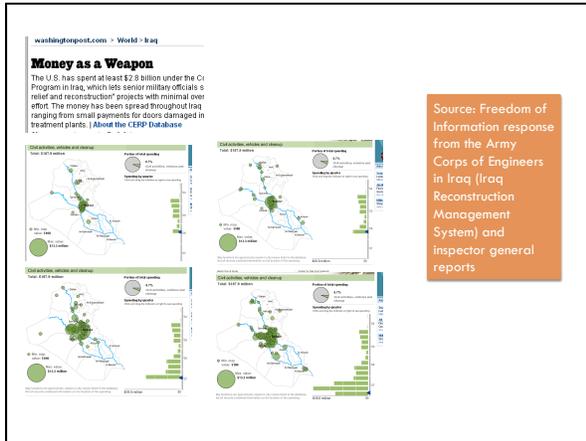


- Sources:
- *People
 - *database of 200 million farm subsidy payment
 - * crop estimates
 - *weather records
 - *emergency records
 - *property ownership records (local)
 - *letters from members of Congress

Forced Out



- Sources:
- *File cabinets (some missing)
 - *Spreadsheet where the color meant different things.
 - *200,000 or so housing code complaints and violations
 - *Paper records from court
 - *People (residents, activists, government officials)
 - *Physical inspection of properties
 - *Landlord-tenant court disputes
 - *Property assessments and deeds



Department	Edited	FactChecked	LinkToOrig	TravelDate	OfficialName	OfficialTitle	Location/City/Locati	Description
Interior	Yes	Yes	http://www.dola.gov/newsroom/06/06_1706	9/2/2004	Craig Manson	Assistant Secretary	Bayfield, WI	At Apostle Islands National Lakeshore, announced the completion of its wilderness study.
Labor	Yes	Yes	http://www.dola.gov/newsroom/06/06_1706	9/29/2004	Dave D. Lauriski	Assistant Secretary	Las Vegas, NV	With the National Mining Association, presented awards to nine operators for their safety records.

Resulting graphics at: <http://www.washingtonpost.com/wp-srv/politics/obama/100days/>

Reversing Predecessors' Actions

Selected reversals:

- Obama's first unilateral actions were a series of orders directing the government to open records in agencies and to post transcripts.
- A series of executive orders concerning labor rights in federal contracts have been removed with each change a party control of the White House since Ronald Reagan.
- The Obama administration has started the process of rescinding at least four sets of "mangled rules" instituted in the waning days of the Bush administration on issues including the "transcience clause" for medical workers and financial disclosures required from union officials.

How the 20 reversals break down:

- 7 actions
- Executive orders and presidential memorandums
- 16
- 3 legislative reversing and policies
- Related to court cases

KEY PLEDGES

Before presidency

- Report torture without expunction; close Guantanamo Bay and secret prisons.
- Armed executive orders to make sure that all contact between all White House staff and sections on regulatory matters is open.
- Sign legislation restoring some rights to sue against pay discrimination and authorizing a health care program for disadvantaged children.

What he has done

- All of Obama's promises on detainee treatment have been abandoned. The Guantanamo prison remains in operation while the administration considers how to close it.
- No executive order has been issued. The OMB has reported one meeting with outside groups related to regulations. The transcript posted all recommendations from outsiders on its Web site.
- Obama signed the Lilly Ledbetter Act on his second week in office, and the children's health-care bill on Oct. 4.

SOURCE: Blueprint for Change | The Washington Post - April 29, 2009

Sources: Daily press releases and "pool" reports from the White House press corps, hand-coded

POTUS Tracker

Every day President Obama meets with key members of his administration, Congress, foreign dignitaries, interest groups and regular citizens. Use our interactive database to track how Obama is spending his time, what issues are getting the most attention and who is influencing the debate. Subscribe to daily schedule via RSS.

EVENTS BY MONTH/YEAR

999

BOXES SIZED TO NUMBER OF EVENTS

Foreign Policy, Economy, Health Care, etc.

HEAD COUNT

Tracking Obama's Appointments

WHO'S WHO

James L. Jones, Appointed, Assistant to the President for National Security Affairs, White House

Senate-Confirmed Positions, By Agency

Obama has 514 Senate-approved positions to fill

Affiliated with the Kennedy family

ABOUT THIS GROUP

Members of the extended Kennedy family network.

- Melody C. Barnes
- Greg Craig
- William J. Lynn III
- Shawn Maher
- Carmel Martin

CHOICES AT A GLANCE

All Senate-approved Positions

13 65 282

100% 100% 100%

Theories of open data

- Governments prefer to lead you to the “right” answer, so they like colorful and “user-friendly” websites that convey their message.
- Transparency advocates look for government to first “wholesale” data, then work on their own website. They’ve had little luck.

New directions in newsgathering

- *What's New?* Smart scraping and text mining – indictments, settlements and regulatory actions across agencies; lawmaker releases and grant announcements.
- *The Real Story*: Anti-aggregation of stories and news from blogs, news sites, RSS feeds, government agencies grouped by story, not by source.
- *Chronologies and social network tools* to organize notes and see new connections
- *Text mining of government documents: Jigsaw, Meandre, Document Cloud*
- *Audio and video analysis*

Opportunities for Computer Science students

- DukeEngage project (look on the DeWitt Wallace Center’s website for a link) to create newsgathering tools
- “Middle layer” of sense-making on unexplored local and national datasets
 - Stimulus
 - Contracts and grants
 - Dispersed records (calendars, etc.)
- Entity extraction, geocoding tools
- Visualizations that say something new