Visualizing A Web Of Data
Stamen Design

Eric Rodenbeck
Michal Migurski (me)
Shawn Allen
Tom Carden
Geraldine Sarmiento
Deborah Monaghan
Ben Cerveny
Sha Hwang
Data + Design

Open data and visualization are colliding, and feeding off one another to mutual benefit
Nate Silver maintained a set of simple information graphics. These were kept minimal: maps, pie charts, bar charts, spark lines. However, the constantly-shifting nature of the election made the site an addictive experience. Fans of the site visited it constantly, showing the emotional value of timely, important information presented simply.
These images, made after the election, show a hunger for a more nuanced way to present information.
This is the New York Times Casualties of War, again dealing with a constant flow of new incoming information. This piece debuted in 2005, and has been kept up to date through the present day.
The site links several variables about the Iraq war to a timeline and a map: date, location, age, race, military branch, and duty type of each American soldier killed in action.

News event connected to the war are also shown along the timeline, allowing for rich interactions with data.
The importance of this project is in its currency. As a news organization, the New York Times has committed to maintaining up to date information for several years.
How the Government Dealt With Past Recessions

Since the Great Depression, presidents have frequently experimented with Keynesian economics to combat recessions. Three economists chronicle the history of government policy during past recessions and explain what worked and what didn’t.
Live, Vast, and Deep

The iron triangle of information visualization

“Live”: our favorite projects demonstrate data that is, ideally, being generated as you watch it.

“Vast”: data can cover an enormous surface area, think Google Maps

“Deep”: data is dense, interlinked
Live, Vast, and Deep
MoveOn’s Virtual Town Hall was an early project we did with US-based liberal political group, MoveOn.org. In 2004, they were (unsuccessfully) trying to prevent a second term for President Bush.

They wanted to improve on the conference calls they were doing via regular phone connections – these were both expensive and not-very-good.
The first difference the project introduced was a move to the web. That’s the Real Audio plugin up there on every page. The application was used as a radio call-in show, with music and conversation broadcast centrally.
The second big difference the project introduced was that users identified themselves with a post code and a headcount of people with them.
The result: Virtual Town Hall knows where you are, and where everyone else is.
50K people joined us on for the first Town Hall!
This ability to reach through the map, to see other people around you, imparted a feeling of presence and excitement to the application.
Special guest, Director Michael Moore, could accept questions from the audience
The map showed a constant stream of input from people around the country. We heard from people in remote parts of the United States that they were able to see, for the first time, how many like-minded individuals were near them.
You could zoom in and see your local area – this is the San Francisco Bay Area, showing all the people attending nearby along with pie charts showing their answers to straw poll questions.
Steal From The Best

“A healthy musical continuum is one where everyone involved is listening to everyone else very, very closely, but they're not only listening to music from the people in their own scene, they're tuned in to other stuff, and then they use that stuff from outside as part of their arsenal in this sort of ongoing battle with the other producers or rivals.”

—Simon Reynolds, The Hardcore Continuum
This is not, strictly speaking, a new form of visual presentation.
This is a major influence on our work, a Canadian atlas published almost 40 years ago. What’s different now is the introduction of instant, live feedback via the internet.
Our work with social news website Digg.com demonstrates a different sort of liveness.
What’s interesting here is that the Labs section visually demonstrates up-to-the-moment live activity on the website. This is Stack, which shows current popular stories on the site. User activity falls from the top of the screen and builds up columns of popularity.

It’s an animated pie chart, and is designed to evoke Tetris.
You can expand individual stories to see what historical activity looks like.

Here’s one from earlier today – the chart across the bottom shows that this popular post-election story hit the front page about 9 hours previously, and you can see a burst of voting and comment activity.
This is a video of Digg Swarm, which shows the same underlying data but in a different visual form.

Here user activity is shown using a biological clustering concept.
Users swarm around stories, and stories grow connections based on shared voting activity.
As before, you can investigate the visual interface – read more about something that looks interesting.
Live, Vast, and Deep
Our work with San Francisco-based Real Estate information aggregator Trulia demonstrates the potential vastness of data online.

Trulia collects data about residential properties for the entire country, from government and private sources. They wanted to demonstrate the massive scale of this information collection by creating something that could be relevant and interesting to people who weren’t necessarily in the market for a home.
Hindsight is initially presented in the form of a blog, with posts about a variety of communities nationwide.
The interface shows colored dots for each home. You can immediately see construction patterns resulting from tract home development, where whole swaths of similar homes are laid down at once.

We used the standard slippy map interface, so it’s possible to move around the country and see construction for different zoom levels. Even when you look up your childhood home in an out-of-the-way town, you can see historical information, link to it, and share it.

This is the SF Bay Area, you can see how people started off living in SF and Oakland in the North in the late 19th century, and then slowly moved down the Bay filling in former farmland with homes. There is a massive bump immediately after World War II.
I’ve shown three projects with a fairly high degree of polish.

It’s important to note that these don’t spring from our heads fully-formed, there’s a lot of exploratory work and dead-ends that contribute to the final piece.

This was an early sketch we did in Processing, showing data animated by house price, from cheapest to most expensive. It beautifully demonstrated animated sweeps, but home prices as a metric seemed insufficiently “real” for a final piece.
This is the same visualization style, also with just properties in San Francisco, now animated by construction date. This was a more satisfying visualization, the information presented is more directly relevant to people, because it’s possible to use it to tell a story about the city.
We added satellite imagery to the visualizations before, and quickly ran into a problem of accidental visual resonance. Prior to the last five or so years, the place most people encounter data visualization is through science, through the military, and through disaster response.
Hindsight looked too much like this NASA image of the fires in the Oakland Hills in 1991. You can pick out each home, and neighbourhood, clearly. “Accidental Visual Resonance.” In response, we removed red, orange, and yellow colors to keep it from looking like a fire.
This shows Plano, Texas - an example of suburban housing development.

At the bottom you can see a timeline of development - it’s quite short, because these homes were all built around the same time, quite rapidly.
Cloudmade

New cartography for OpenStreetMap

http://maps.cloudmade.com
OpenStreetMap is the new story here
In four years of work, the volume of data in OpenStreetMap has progressed enormously.
Baghdad
Appearances Matter
Intentional Visual Resonance (pursue!)
Live, Vast, and Deep
The SFMOMA ArtScope is our most recent project.

This is a deep dive into the Museum Of Modern Art’s collection, most of which is not on display in the museum.

It doesn’t look like a map, but we’ve been exploring the use of development and publishing ideas from online maps and satellite images for other purposes.
This is an image from Microsoft’s Terraserver, from 2004. Prior to the release of Google Maps in 2005, this and MapQuest were the only good ways to get geographical information online. Interactions were slow, form–based, and limited. Imagery and map content was limited and poorly–designed.
Google’s maps were a massive step forward, because they allowed you to move the map in the page, zooming freely from one place to another.
The infinite, continuous road maps and satellite imagery are available over a regular broadband connection because Google serves them to you as small square images.
This idea can be applied to other kinds of deep navigation.
This is ArtScope

Kenneth Josephson

**Chicago**

1961

photograph | gelatin silver print

**Keywords:**

- men
- women
- shadows

8 1/16 in. x 11 15/16 in. (20.48 cm x 30.32 cm)

Acquired 1980

Collection SFMOMA. Gift of the Sidney and Diana Avery Trust

© Kenneth Josephson
You can move the magnifying glass around to see what’s in the Museum of Modern Art’s extensive collection, or use tags and text searches to the right if you know what artist you’re looking for, or are interested in browsing artwork by medium or subject.
Really though, it’s a kind of Google Maps for Modern Art.

All the familiar interaction metaphors – panning, zooming, small overview map – are there. There are certain interaction ideas popular on the web that are ripe for application to new contexts.
Maps, Generally

Maps online are swiftly becoming the focal point for new thinking about data
Andy Allan: Open Cycle Map
This is an award–winning custom render of OpenStreetMaps that illustrates why making your own maps from scratch is important: there are a lot of different kinds of geographic context out there, beyond just freeways and pizza places.
GraphServer: Routing

Brandon Martin-Anderson: shortest-path tree of San Francisco transit (red) and walking (black).
This is just transit, with an indication of walk + ride time.
Shawn Allen: San Francisco contours and bike routes
Shawn overlaid bike routes on contoured hill in SF, using data directly published by city government.
Heat Maps, Underneath

Oakland Crimespotting heat maps
Meanwhile, this is a heat map of crime in West Oakland with a thin layer of street context from OSM.
Black Box/Reach Through

“...scientific and technical work is made invisible by its own success. When a machine runs efficiently, when a matter of fact is settled, one need focus only on its inputs and outputs and not on its internal complexity. Thus, paradoxically, the more science and technology succeed, the more opaque and obscure they become.”

—Bruno Latour, Pandora’s Hope
Oakland Crimespotting is another recent project that demonstrates depth of interpretation through a variety of views on information.
This is the state of the art in mapping crime information for cities in the US. Maps are designed by programmers and not fully resident on the web.

Generally speaking, it’s hard to link to things and hard to explore information laterally.
I’ve borrowed the title of this talk from Tom Coates’s Native to a Web of Data.

“A web of data sources, services for exploring and manipulating data, and ways users can connect them together”
“Designing Data For Reuse”
Matt Biddulph, 2005
http://www.hackdiary.com/slides/xtech2005/

Another influential piece of writing in this space is from Dopplr CTO Matt Biddulph.
Matt describes a series of best practices around open data formats and permanent, guessable URL’s.
Views

How many ways are there to look at data?

Oakland Crimespotting combines practices around usable data formats, web-friendly bookmarkable addresses, and visualization as a interpretation layer above data.
Murder

Wednesday, Mar 5, 2008 5:48am

MURDER

Case Number 08-016924-003, Police Beat 06X.

3200 block of San Pablo Ave, Emeryville, CA 94608, USA

Scroll down to see related and nearby reports, or to leave a comment.

View an interactive map of this report.

Related Reports

These reports are directly related to the one above and may be part of the same incident.

First, there is a page for every individual crime report in the system.

This is the most basic building block of the site. Each one has a sensible URL that includes a map, connected reports, and nearby neighborhood information.
Second, there are listing pages that show collections of reports: by day, by type of crime, by police beat.
We're intentionally trying to stretch the definition of "API" here: the classic, Flickr–driven concept of an XML web service is definitely one Web 2.0 compliant way of looking at things, but Excel files and permanent URLs right there on the website is a broader concept that invites members of the non–geek public to join in. These have all been API–like "handles" that visitors can connect with.
Third, everything is presented on a map.
The map shows global context.
As well as the relative isolation of just one city of data.
Lessons

What can be learned from visual data?

The project also demonstrates additional layers of information that can be derived from raw data.
Here we see reports of just one particular type, isolated on the map.
These reports almost all happened on one day – prostitution in particular following this “sting” pattern.
Violent crime is spread all over – there isn’t as much of a focus in West Oakland as you’d expect
Most of these narcotics reports are in beats 06X and 07X
Design + Data

Open data and visualization are colliding, and feeding off one another to mutual benefit
Thank You

mike@stamen.com