

Big Data

a glimpse of today's state of play

Sanjana Hattotuwa

TED Fellow

Special Advisor, ICT4Peace Foundation

40 ZETTABYTES

(Estimated volume of data will be created by 2020, an increase of 300 times from 2009)

6 BILLION PEOPLE
Have cell phones

WORLD POPULATION: 7 BILLION

Volume

SCALE OF DATA

It's estimated that **2.5 QUINTILLION BYTES** (2.5 quadrillion bytes) of data are created each day



Most companies in the U.S. have at least **100 TERABYTES** (100 billion bytes) of data stored

The New York Stock Exchange captures **1 TB OF TRADE INFORMATION** during each trading session



Velocity

ANALYSIS OF STREAMING DATA

Modern data feeds come at **100 SECONDS** that require instant action

By 2014, it is projected there will be **18.9 BILLION NETWORK CONNECTIONS** - almost 2.5 connections per person on earth



The FOUR V's of Big Data

Volume, velocity, variety, and veracity are the four V's of big data. Volume refers to the amount of data, velocity to the speed at which data is generated and processed, variety to the different types of data, and veracity to the reliability of the data.

Processing all this data requires sophisticated tools and techniques. As data volumes increase, so do the challenges of storage, management, and analysis. Organizations must invest in robust infrastructure and skilled personnel to effectively manage and derive value from their big data.

By 2015, **4.4 MILLION IT JOBS** will be needed to manage the growth of big data, according to the International Data Corporation (IDC).



As of 2011, the growth rate of data at Facebook was estimated to be **100 X ANNUAL**

100 X ANNUAL



30 BILLION PIECES OF CONTENT are shared on Facebook every month



Variety

DIFFERENT FORMS OF DATA

By 2014, it's estimated there will be **400 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

4 BILLION+ HOURS OF VIDEO are uploaded to YouTube each month



100 MILLION TWEETS are sent per day by about 200 million people on the earth



1 IN 3 BUSINESS LEADERS don't trust the information they use to make decisions



More than 90% of the data we produce is non-structured

Veracity

UNCERTAINTY OF DATA

How much quality costs that US economy grows **\$1.1 TRILLION A YEAR**





global pulse

A global nervous system

Defending online freedoms

As global connectivity increases, we must make sure the internet is used to increase, not undermine, human progress



Hillary Clinton

[theguardian.com](#), Friday 22 January 2010 14:11 GMT

 [Jump to comments \(103\)](#)



Hillary Clinton delivers a speech on internet freedom at the NewsEum in Washington DC. Photograph: Allison Shelley/EPA

The spread of information networks is forming a new nervous system for our planet. When something happens in [Haiti](#) or Hunan, the rest of us learn about it in real time – from real people. And we can respond in real time as well. Americans eager to help in the aftermath of a disaster and the [girl trapped in the supermarket](#) are connected in ways that were not even imagined a year ago, even a generation ago. That same principle applies to almost all of humanity today.

MetaLayer: sentiment analysis on twitter

<http://dev.geosprocket.com/vtirenecarto.html#>

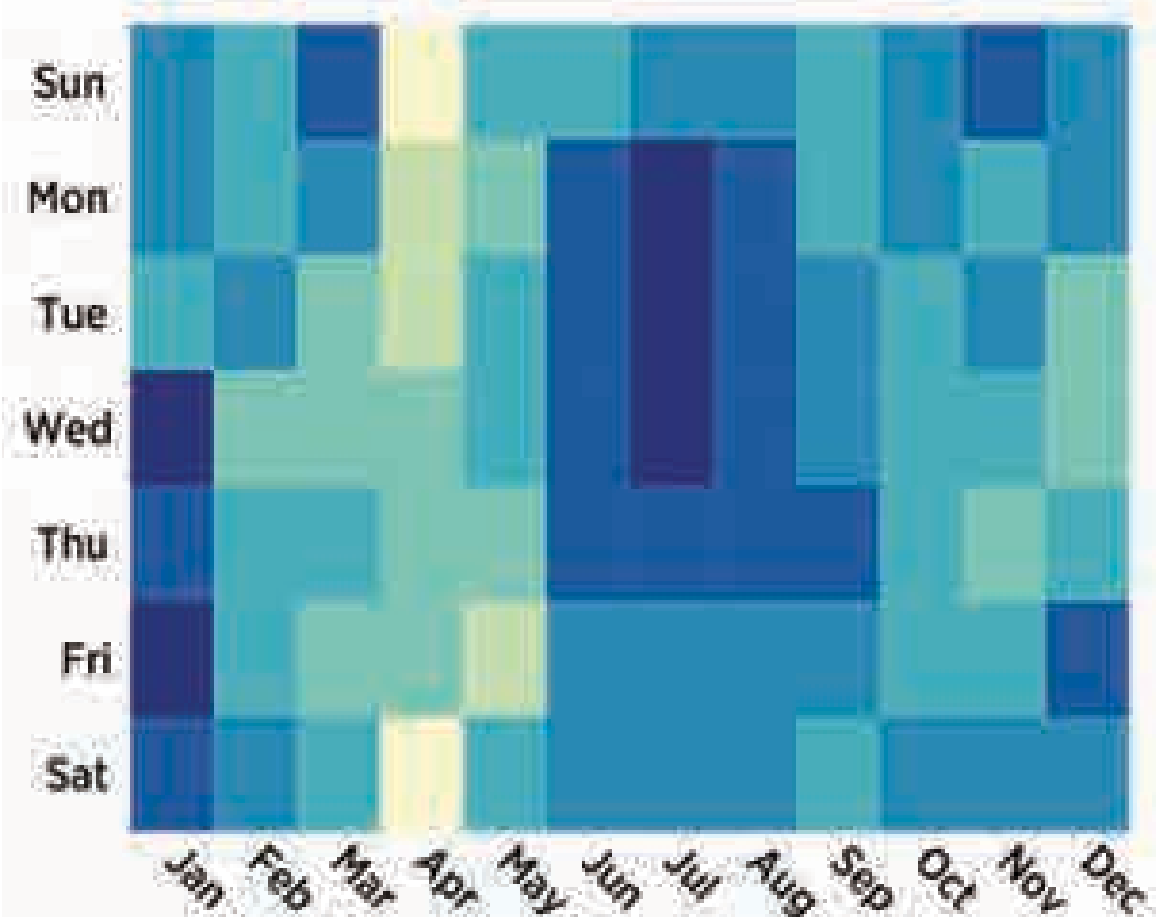


twitter sentiments

<http://mashable.com/2014/03/10/twitter-data-hungover/>

When will you be “late for work”?

Tweets using the phrase “late for work” by day, 2013

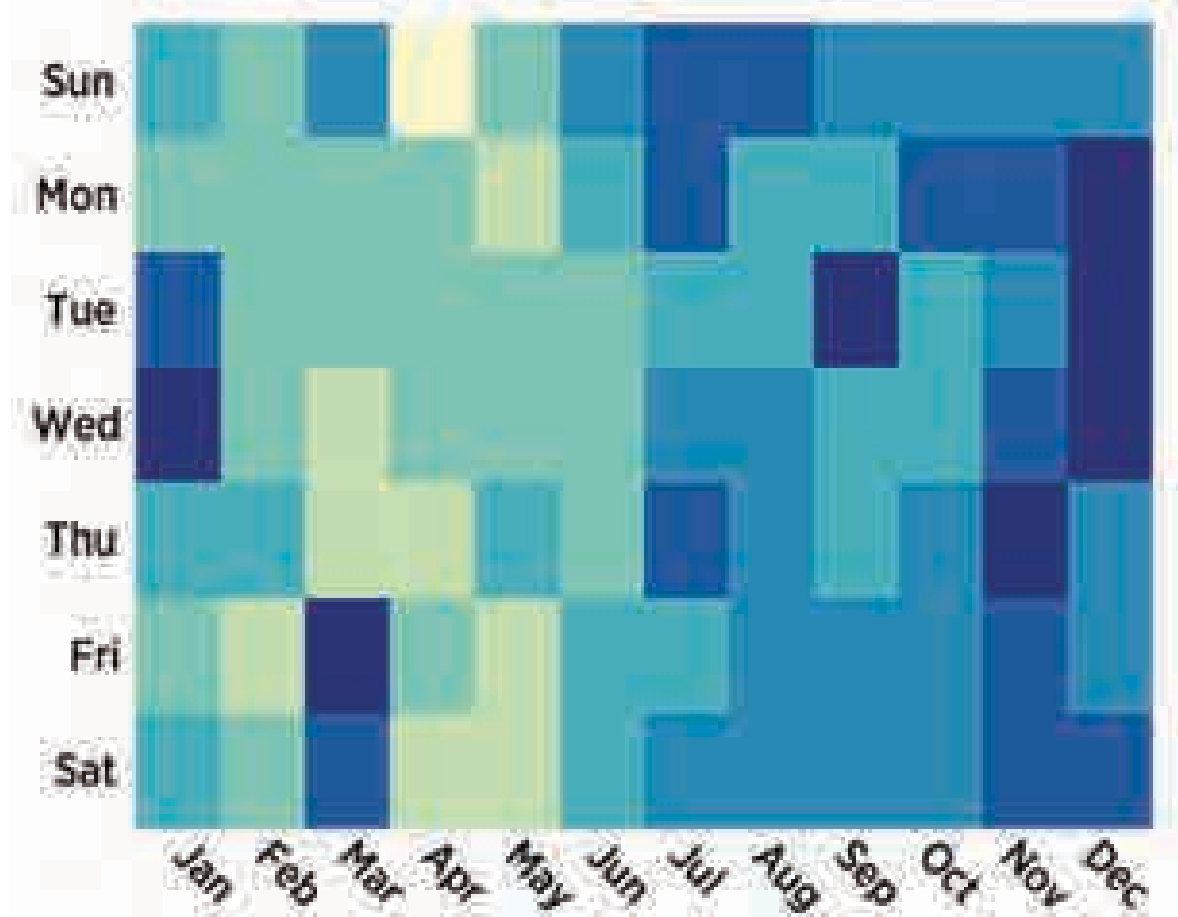


Fewer Tweets More Tweets



When do you “feel happy”?

Tweets using the phrase “feeling happy” by day, 2013. Rate per million Tweets sent



Fewer Tweets More Tweets

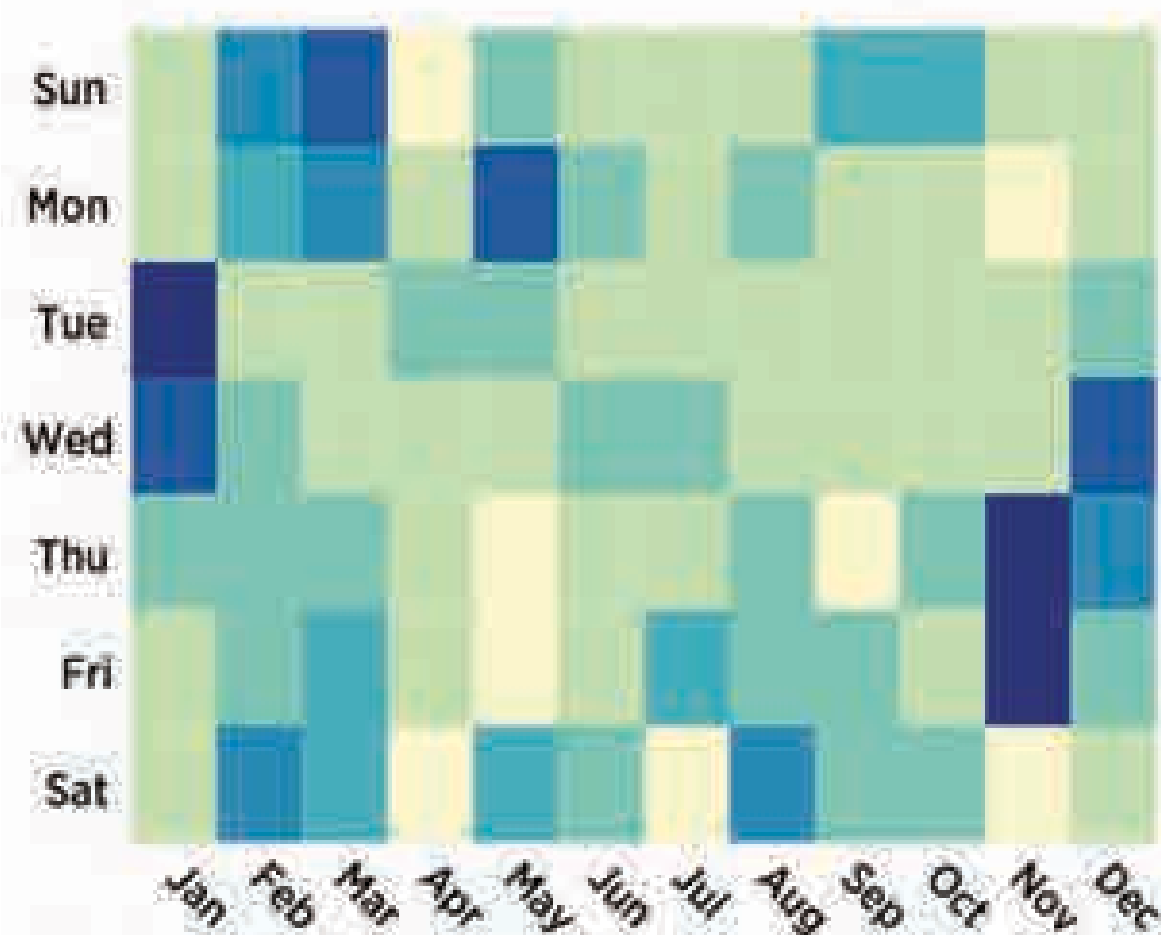


twitter sentiments

<http://mashable.com/2014/03/10/twitter-data-hungover/>

When are you “hungover”?

Tweets using the phrase “hungover” by day, 2013. Rate per million Tweets sent

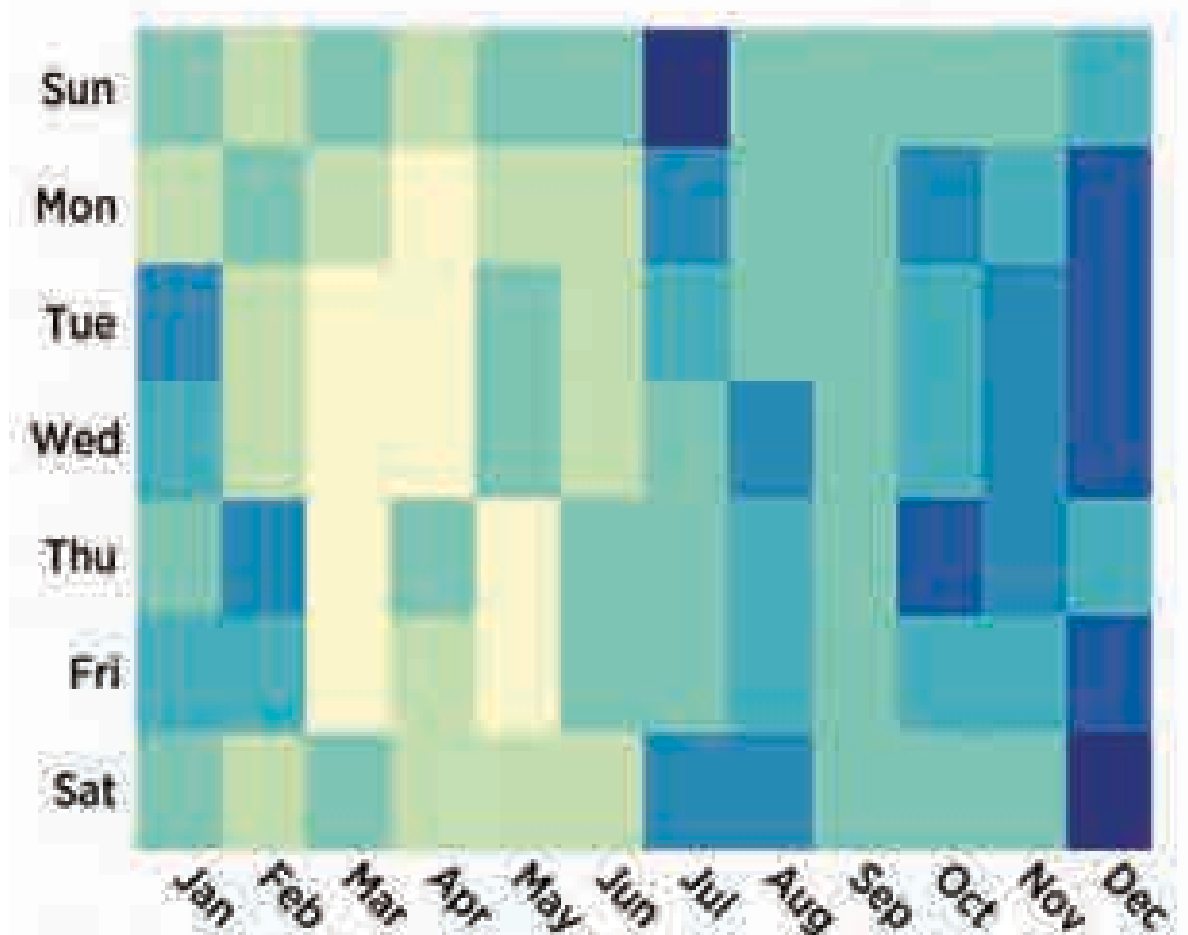


Fewer Tweets More Tweets



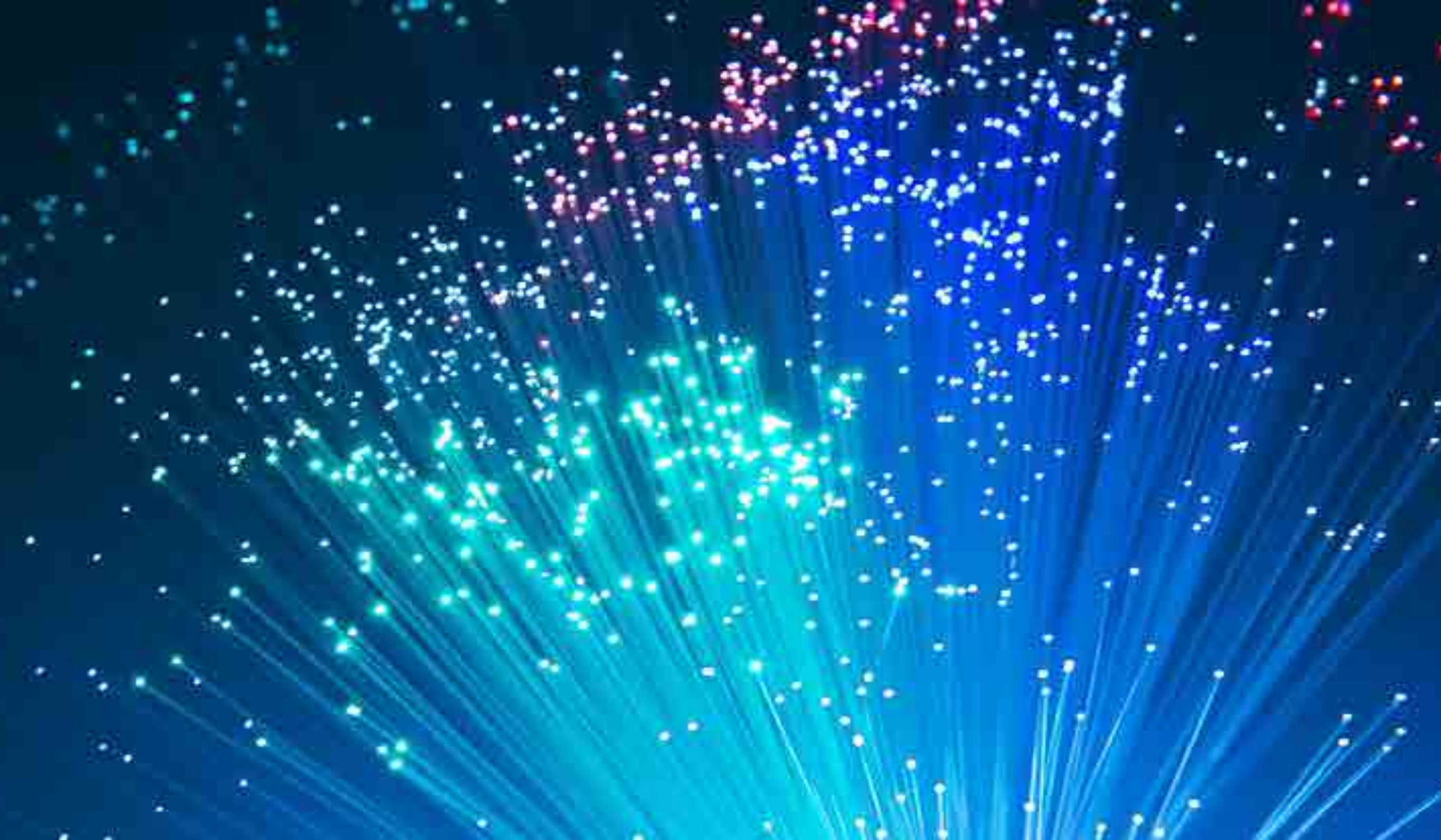
When do you “feel sad”?

Tweets using the phrase “feeling sad” by day, 2013. Rate per million Tweets sent



Fewer Tweets More Tweets





big data

Massive, often open data

public data: free portals

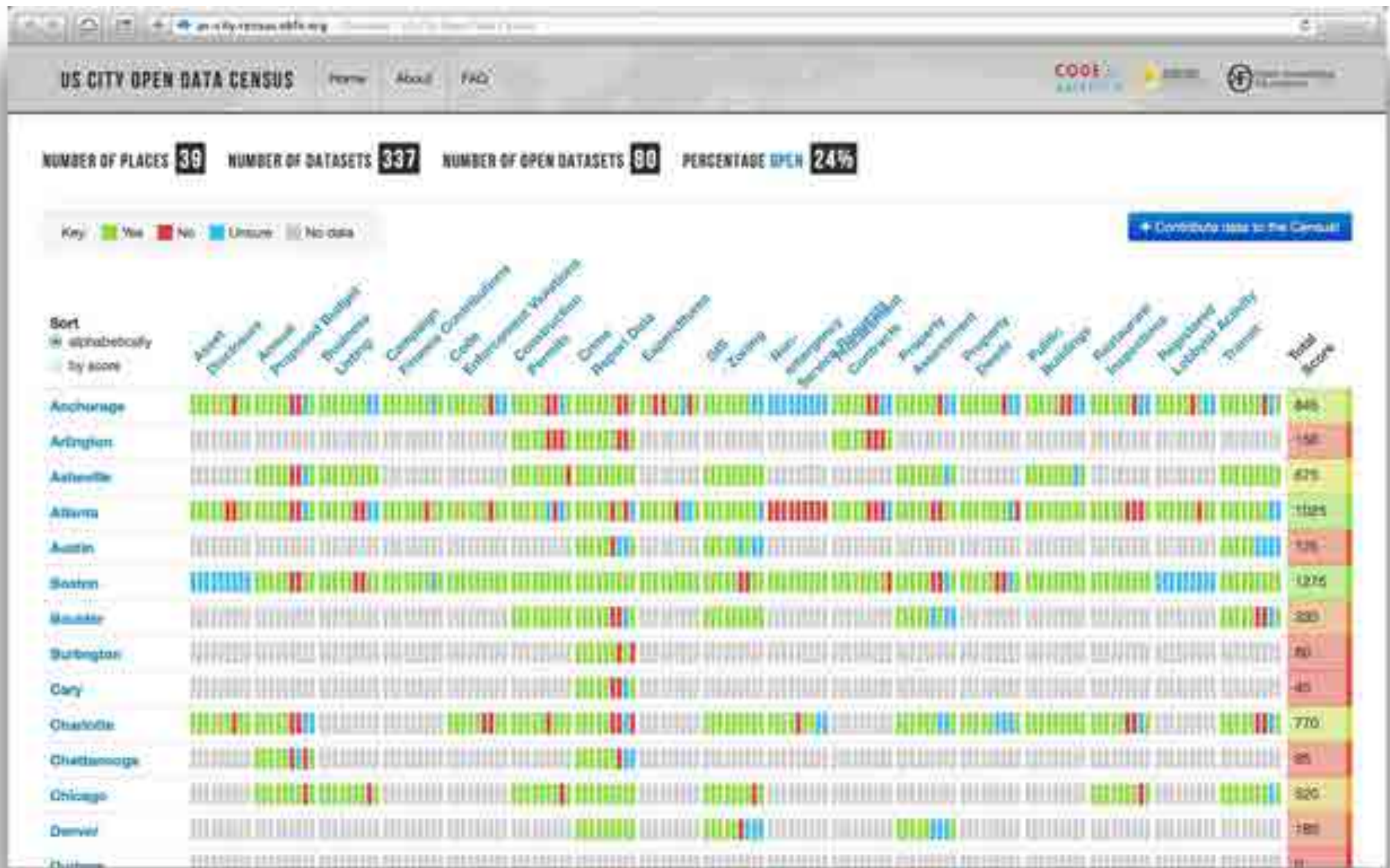


public data: cities



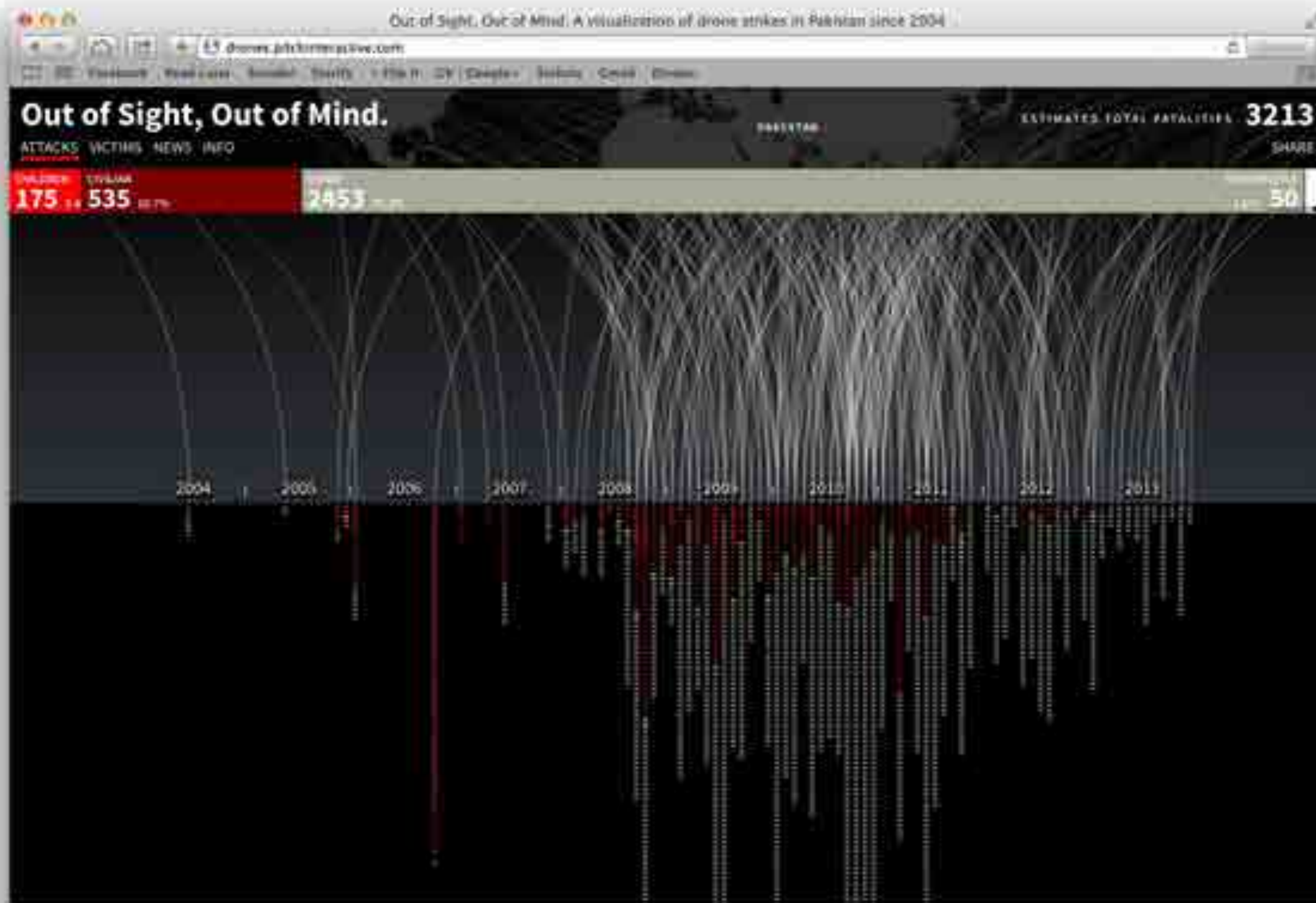
open cities data: US

<http://us-city.census.okfn.org>



drones: rendering new perspectives

<http://drones.pitchinteractive.com>





check-ins

locating self & others

New York



Europe



Flickr and Twitter: geo-tagged

<http://www.fastcodesign.com/1664462/infographic-of-the-day-using-twitter-and-flickr-geotags-to-map-the-world>



The screenshot shows a web browser window with the URL www.fastcodesign.com/1664462/infographic-of-the-day-using-twitter-and-flickr-geotags-to-map-the-world. The page features a large infographic of the world map where the landmasses are composed of numerous small, colorful dots representing geotagged photos from Flickr and Twitter. The browser's address bar and various navigation icons are visible at the top. Below the infographic, the title 'Infographic Of The Day: Using Twitter And Flickr Geotags To Map The World' is displayed, followed by a short paragraph: 'THERE'S TWO SUBTLE FORCES BEING MAPPED HERE: ONE, THE DIFFERING WAYS THAT INTERNET COMPANIES SPREAD, AND TWO, THE VAST DISPARITY BETWEEN A CITY SEEN BY A TOURIST VS. A NATIVE.' To the right of the text is a small advertisement for 'Slow Mac?' featuring a speedometer graphic.

Co.DESIGN | Co.EXIST | Co.CREATE | SUBSCRIBE | DESIGN DAILY | FASTCOMPANY

INFOGRAPHIC OF THE DAY

Infographic Of The Day: Using Twitter And Flickr Geotags To Map The World

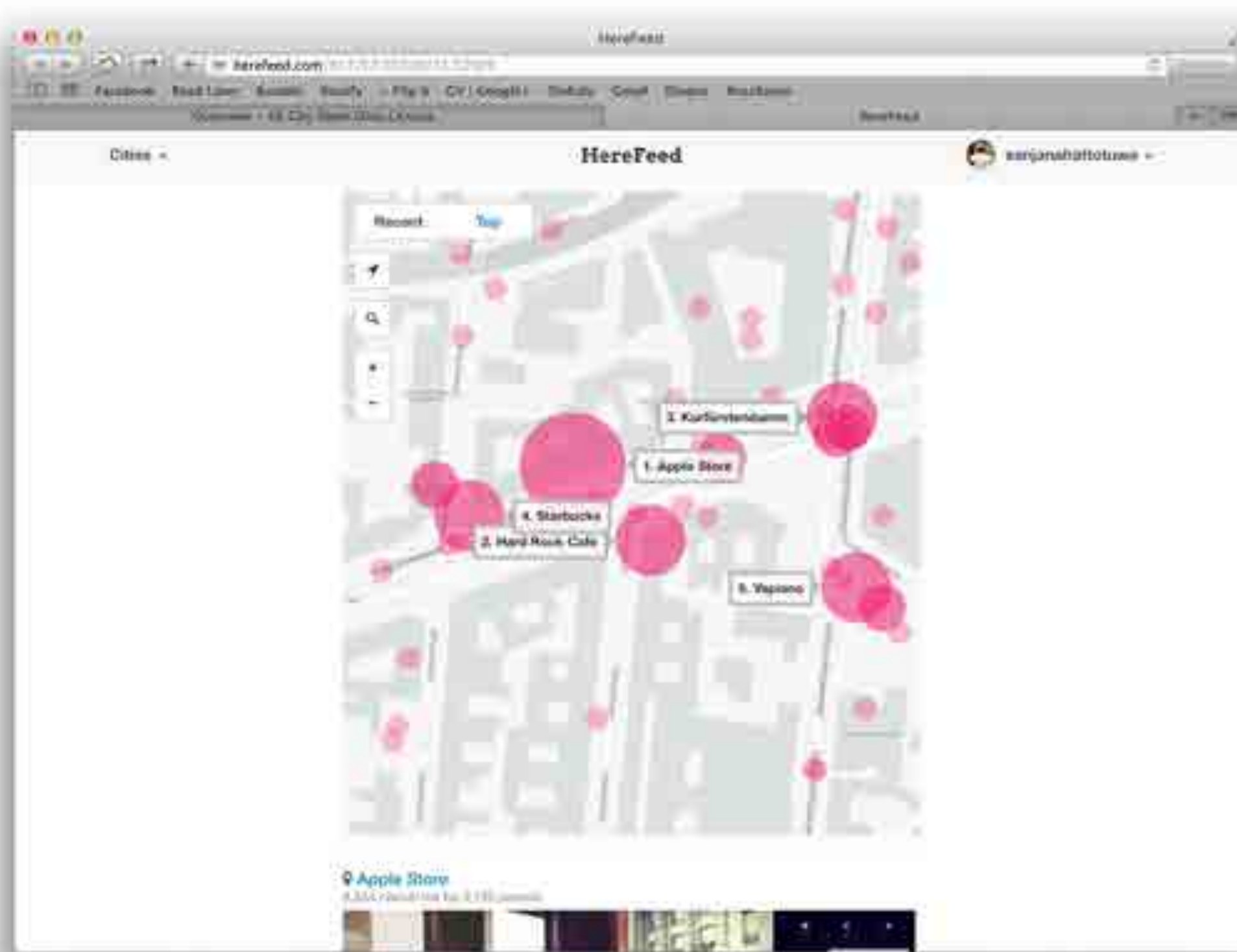
THERE'S TWO SUBTLE FORCES BEING MAPPED HERE: ONE, THE DIFFERING WAYS THAT INTERNET COMPANIES SPREAD, AND TWO, THE VAST DISPARITY BETWEEN A CITY SEEN BY A TOURIST VS. A NATIVE.

Slow Mac?
MacKeeper - Speed Up your Mac!

PERFORMANCE level

herefeed | instagram geo-tagged

<http://herefeed.com>



city map

<http://www.flickr.com/photos/walkingsf/5926358324/sizes/l/in/photostream>



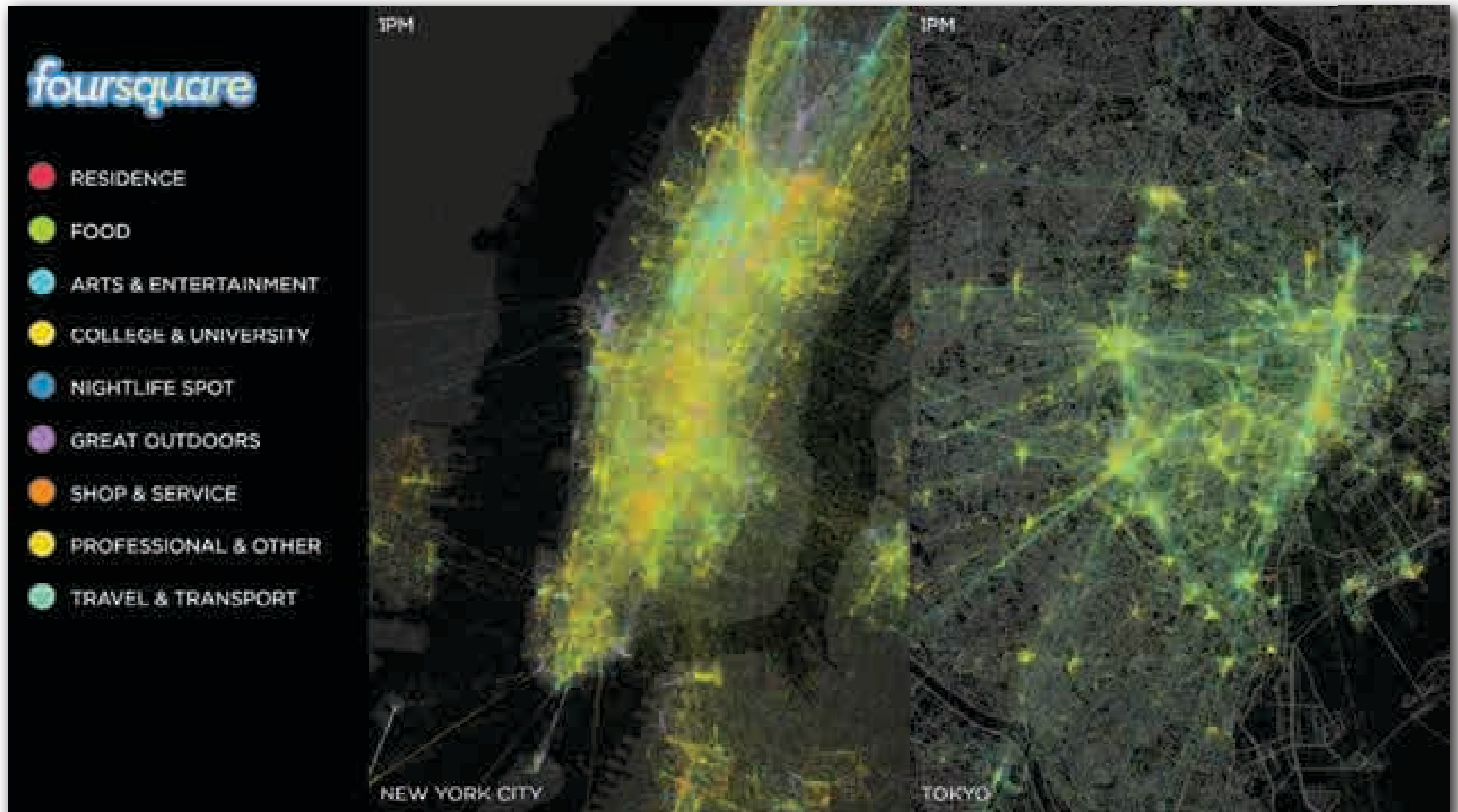
foursquare: check-ins

<https://foursquare.com/infographics/500million?>



foursquare checkins: geo-tagged pulse

<http://vimeo.com/62289901>





big data & media

data driven journalism

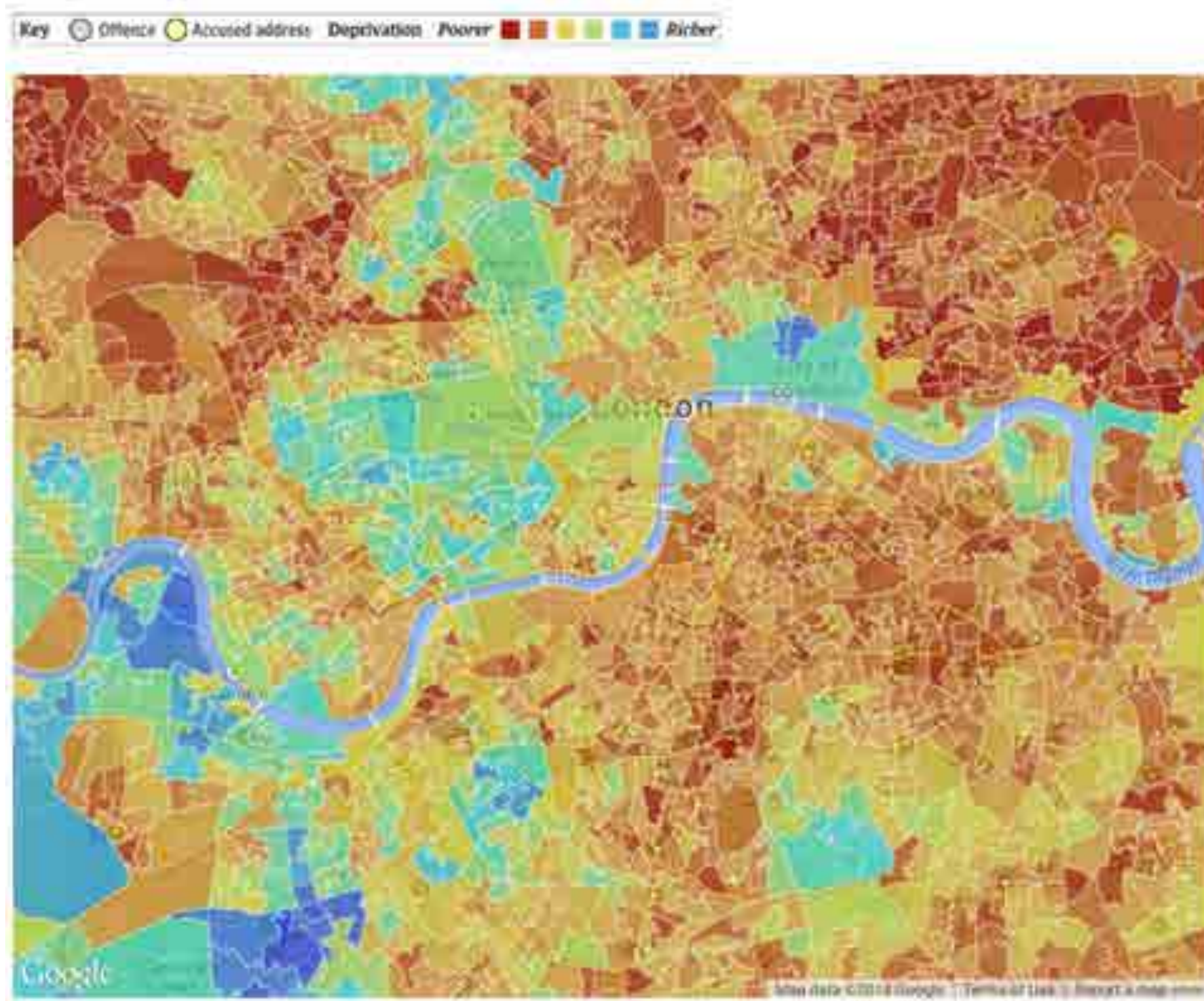
data driven journalism: Guardian

<http://www.guardian.co.uk/data>

The screenshot shows the Guardian Data Store website. At the top, the Guardian logo is on the left, and a search bar is on the right. Below the logo is a navigation menu with links to News, Sport, Comments, Culture, Business, Money, Life & style, Travel, Environment, Tech, TV, Video, Dating, Offers, and Jobs. A secondary menu includes News, Datablog, Data A-Z, Show & Tell, Big Data, Data Journalism, Twitter, Flickr, Facebook, Tumblr, Staff, and OUR BOOK. The main heading is "DATA STORE" in large orange letters, with the tagline "Facts are sacred" below it. The page is divided into several sections: "Latest from the Datablog" featuring a map of the Boston area with the headline "Boston explosions: information and misinformation" and a sub-headline "No arrests have yet been made and no groups have claimed responsibility for the explosions at the Boston marathon. Social media, both good and bad, has tried the information flood"; "Editor's picks" featuring a portrait of Margaret Thatcher with the headline "How Britain changed under Margaret Thatcher. In 15 charts" and a sub-headline "What kind of Britain did Margaret Thatcher leave behind?"; "Law" featuring a headline "Death penalty statistics, country by country" and a sub-headline "The latest regional punishment statistics are out. Find out who uses the death penalty today"; "Military" featuring a headline "South v North Korea: how do the two countries compare? Visualised"; and "Highlights" featuring two smaller articles: "Public spending: Your share spent" and "Datablog: More & better". A blue banner on the right side of the page reads "10% plus avantageux avec Help Point Plus!" and "Calculated to promote maintenance!". Below this is a Zurich Connect logo.

poverty and unrest: Guardian

<http://www.theguardian.com/news/datablog/interactive/2011/aug/16/riots-poverty-map>



social media & rumours: Guardian

<http://www.theguardian.com/uk/interactive/2011/dec/07/london-riots-twitter>

How riot rumours spread on Twitter

Analysis of 2.6 million tweets shows Twitter is adept at correcting misinformation - particularly if the claim is that a tiger is on the loose in Primrose Hill

[Read about how we made this interactive](#)



Guardian interactive team, Rob Procter, Farida Vli and Alex Voss
theguardian.com, Wednesday 7 December 2011 15:34 GMT

Select a rumour to see how unsubstantiated claims are spread on Twitter before being confirmed or denied



Rioters attack London zoo and release animals



Rioters cook their own food in McDonald's



Police beat a 15-year-old girl



London Eye set on fire



Rioters attack a children's hospital in Birmingham



Army deployed in Bank



Miss Selfridge set on fire



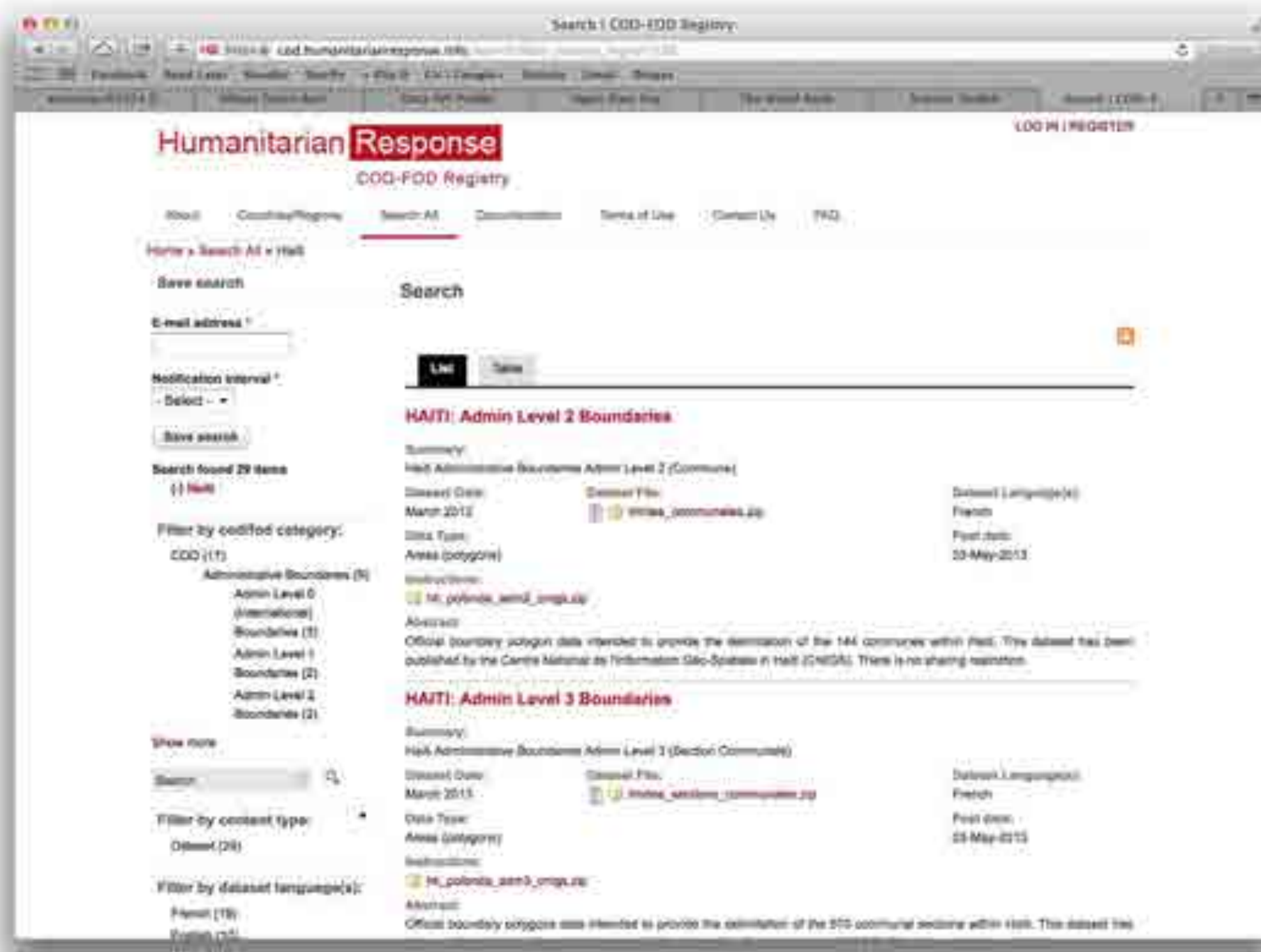
big data & aid

Data driven relief ops



humanitarian response: CODs & FODs

<http://cod.humanitarianresponse.info>



situational awareness

<http://corp.geofeedia.com>



mobile network data & cdr's

http://www.unglobalpulse.org/Mobile_Phone_Network_Data-for-Dev

CALLER ID	CALLER CELL TOWER LOCATION	RECIPIENT PHONE NUMBER	RECIPIENT CELL TOWER LOCATION	CALL TIME	CALL DURATION
X76VG588RLPQ	2°24' 22.14", 35°49' 56.54"	A81UTC93KK52	3°26' 30.47", 31°12' 18.01"	2013-11- 07T15:15:00	01:12:02

mobile network data & cdr's

http://www.unglobalpulse.org/Mobile_Phone_Network_Data-for-Dev

- Population migration during the Haiti earthquake
- Mapping malaria in Kenya
- Combatting H1N1 flu in Mexico
- Socio-economic indicators in the UK, disaggregated by region
- Poverty levels in Cote d'Ivoire
- Optimising transportation networks in Abidjan

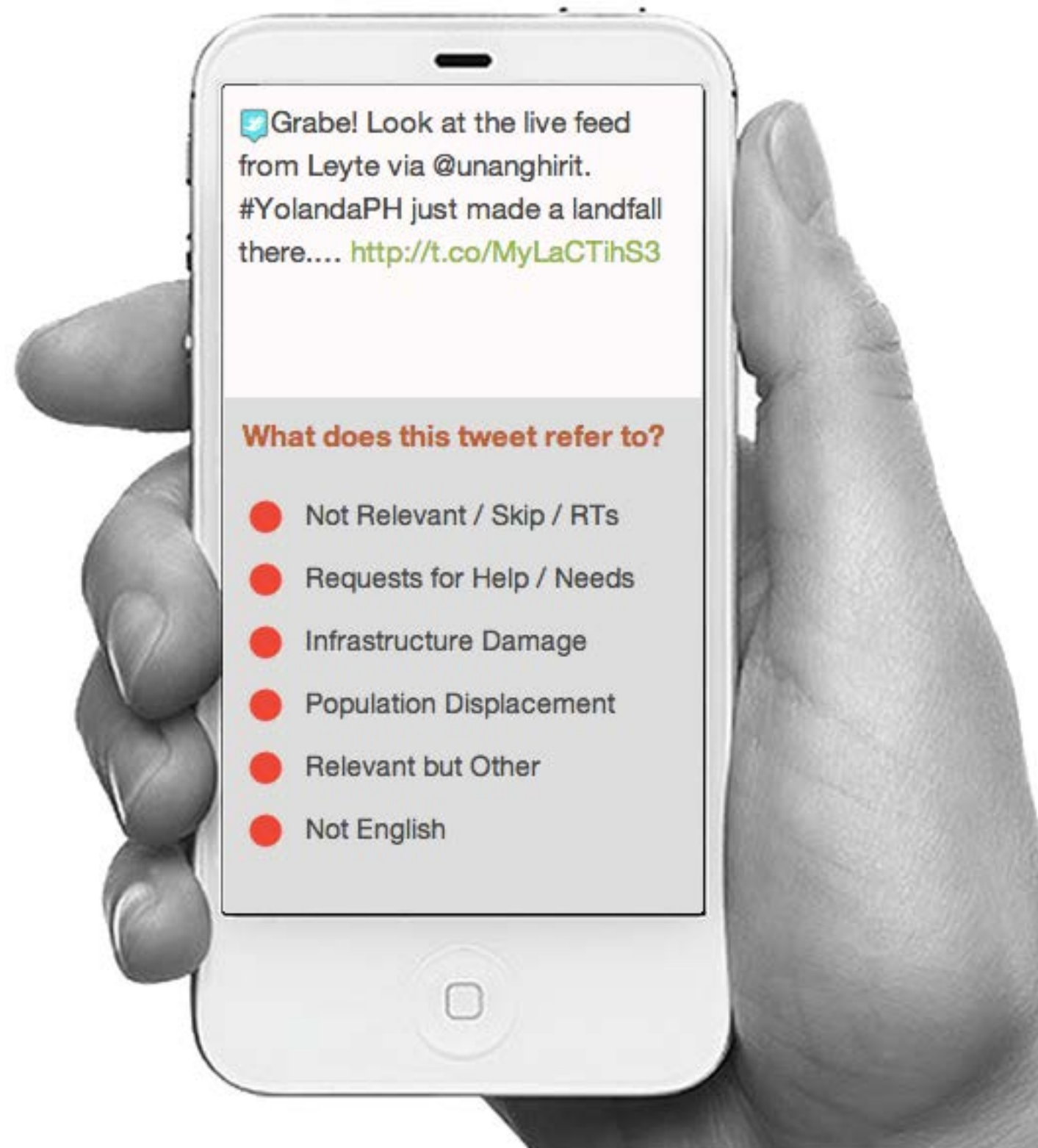
digital humanitarians | new communities

<http://digitalhumanitarians.com>



digital humanitarians

<http://digitalhumanitarians.com>



digital humanitarians

<http://digitalhumanitarians.com>



digital humanitarians

<http://bit.ly/lS7yaE>

7,000+ images

105,000+ clicks (tweets + images)

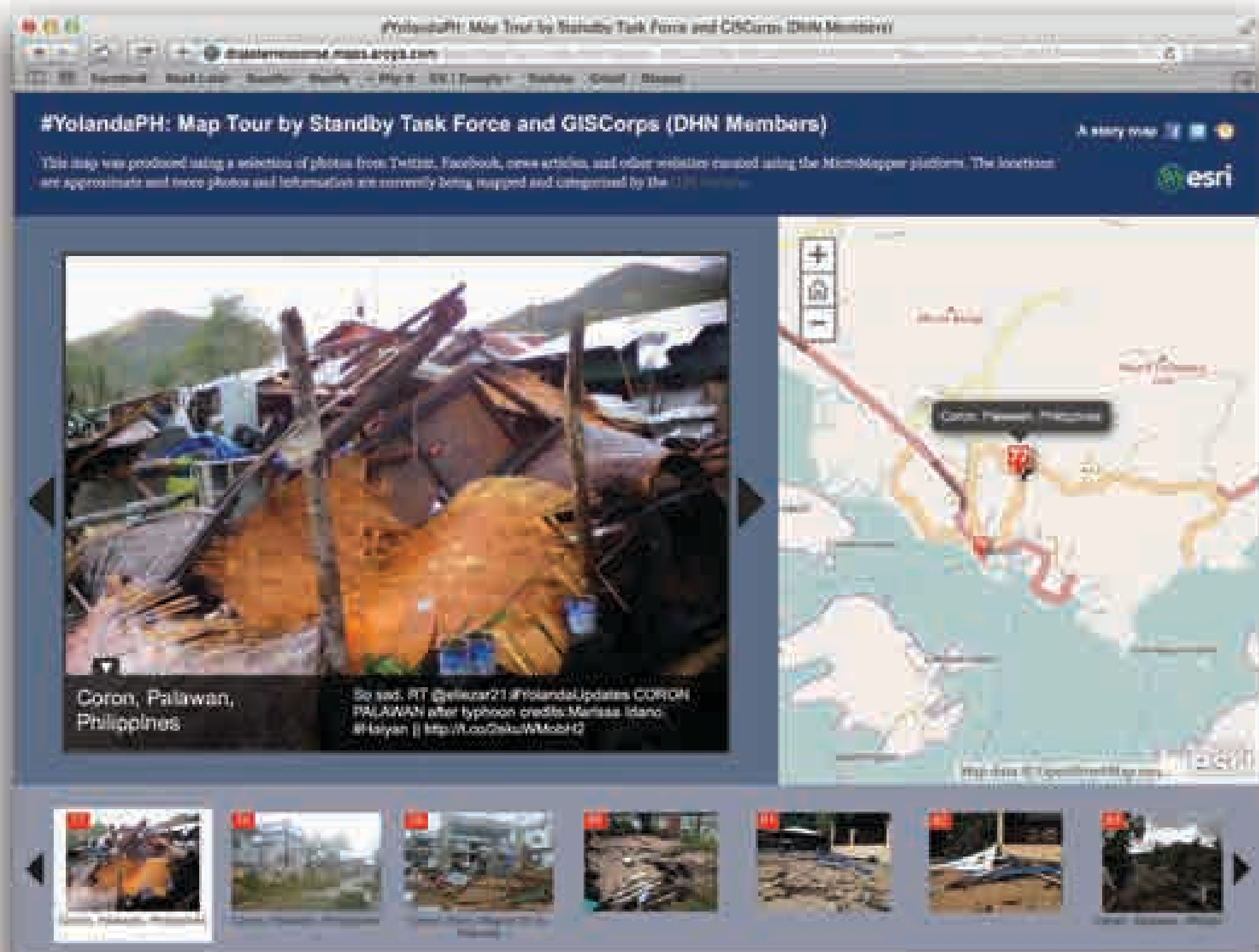
Makati (**Philippines**), Cape Town (**South Africa**), Canberra & Gold Coast (**Australia**), Berkeley, Brooklyn, Citrus Heights & Hinesburg (**US**), Kamloops (**Canada**), Paris & Marcoussis (**France**), Geneva (**Switzerland**), Sevilla (**Spain**), Den Haag (**Holland**), Munich (**Germany**) and Stokkermarke (**Denmark**)

No prior experience or training necessary.

Anyone who knows how to use a computer mouse can become a digital humanitarian.

digital humanitarians

<http://bit.ly/lS7yaE>



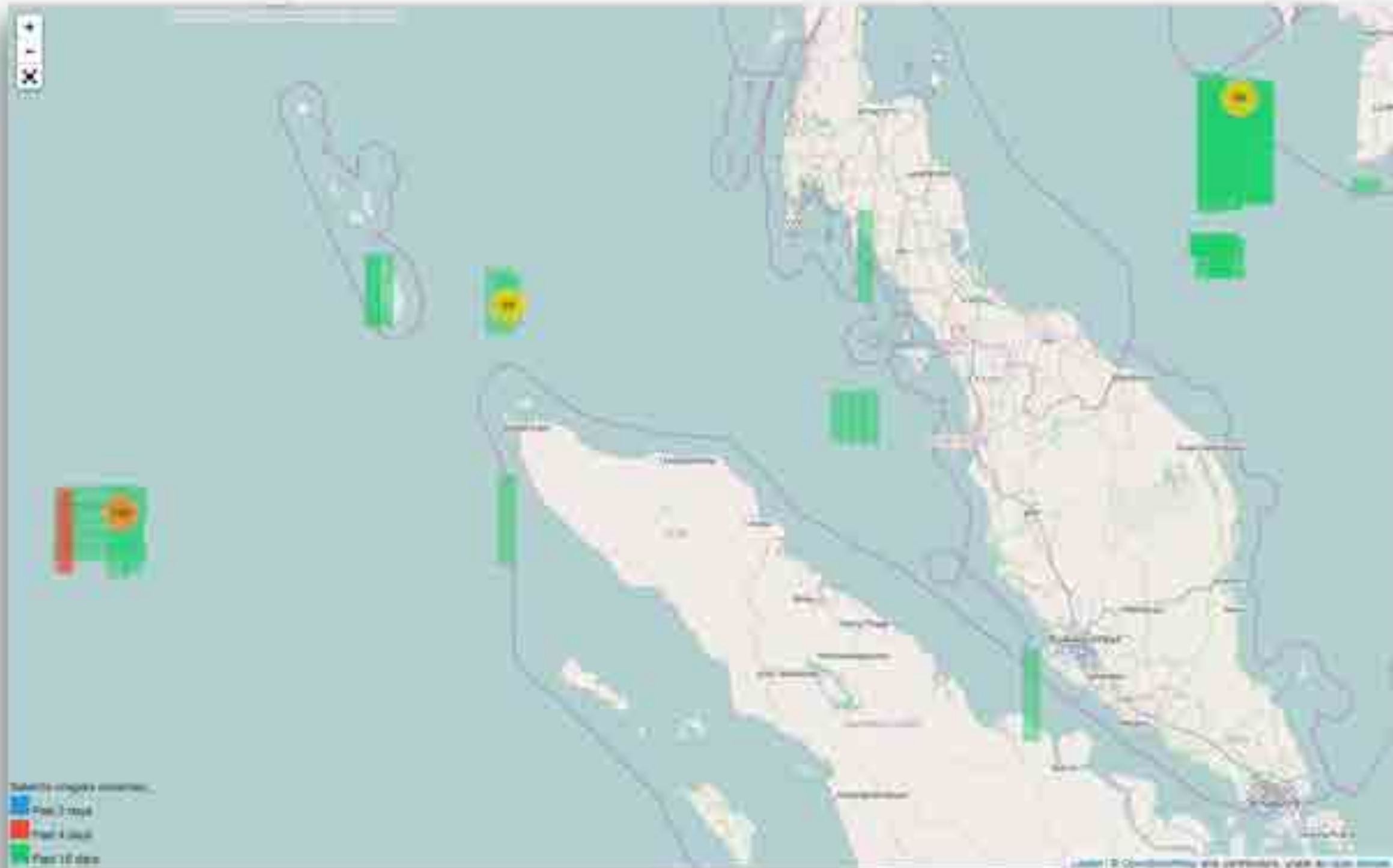
parcelling out large data analytics

<http://tomnod.com>



parcelling out large data analytics

<http://tomnod.com>



digital command operations centre

<http://i.dell.com/sites/doccontent/corporate/case-studies/en/Documents/2012-red-cross-10011336.pdf>

The screenshot shows a web browser window displaying a Dell case study. The title is "The American Red Cross and Dell collaborate on first-ever digital operations center focused on humanitarian efforts". The main image shows the interior of the "Digital Operations Center" with the American Red Cross logo and the text "Powered by Dell". To the right, there is a "Customer profile" section with the American Red Cross logo and a table of details. Below that is a "Challenge" section, followed by a "Solution" section, and finally a "Benefits" section.

The American Red Cross and Dell collaborate on first-ever digital operations center focused on humanitarian efforts

Customer profile	
	American Red Cross
Company	American Red Cross
Industry	Humanitarian Aid
Country	United States
Employees	35,000
Web site	redcross.org

Challenge

The American Red Cross wanted to boost its social media efforts to be able to respond immediately to disasters using information gathered by monitoring the social Web.

Solution

The organization partnered with Dell to build a Digital Command Center at American Red Cross headquarters in Washington, D.C., with six large screens that show a variety of data visualizations of relevant public conversations on the social Web. Staff members who make up the social engagement team monitor the screens and shared volunteers are brought in as reinforcements during an active disaster.

Benefits

digital command operations centre

<http://i.dell.com/sites/doccontent/corporate/case-studies/en/Documents/2012-red-cross-10011336.pdf>



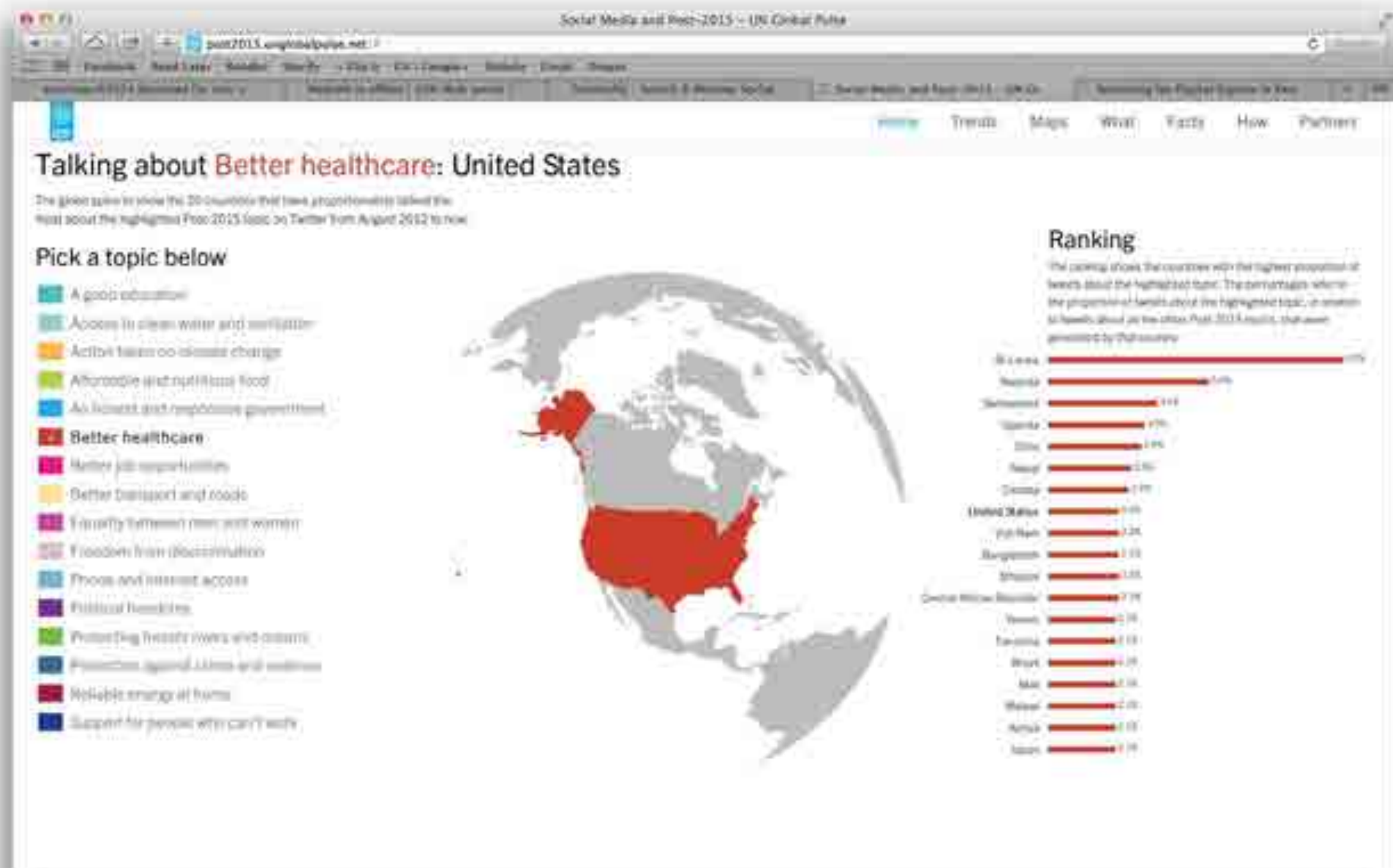


polycymaking

Data driven governance

Post-2015 Conversations on Twitter

<http://post2015.unglobalpulse.net>





challenges

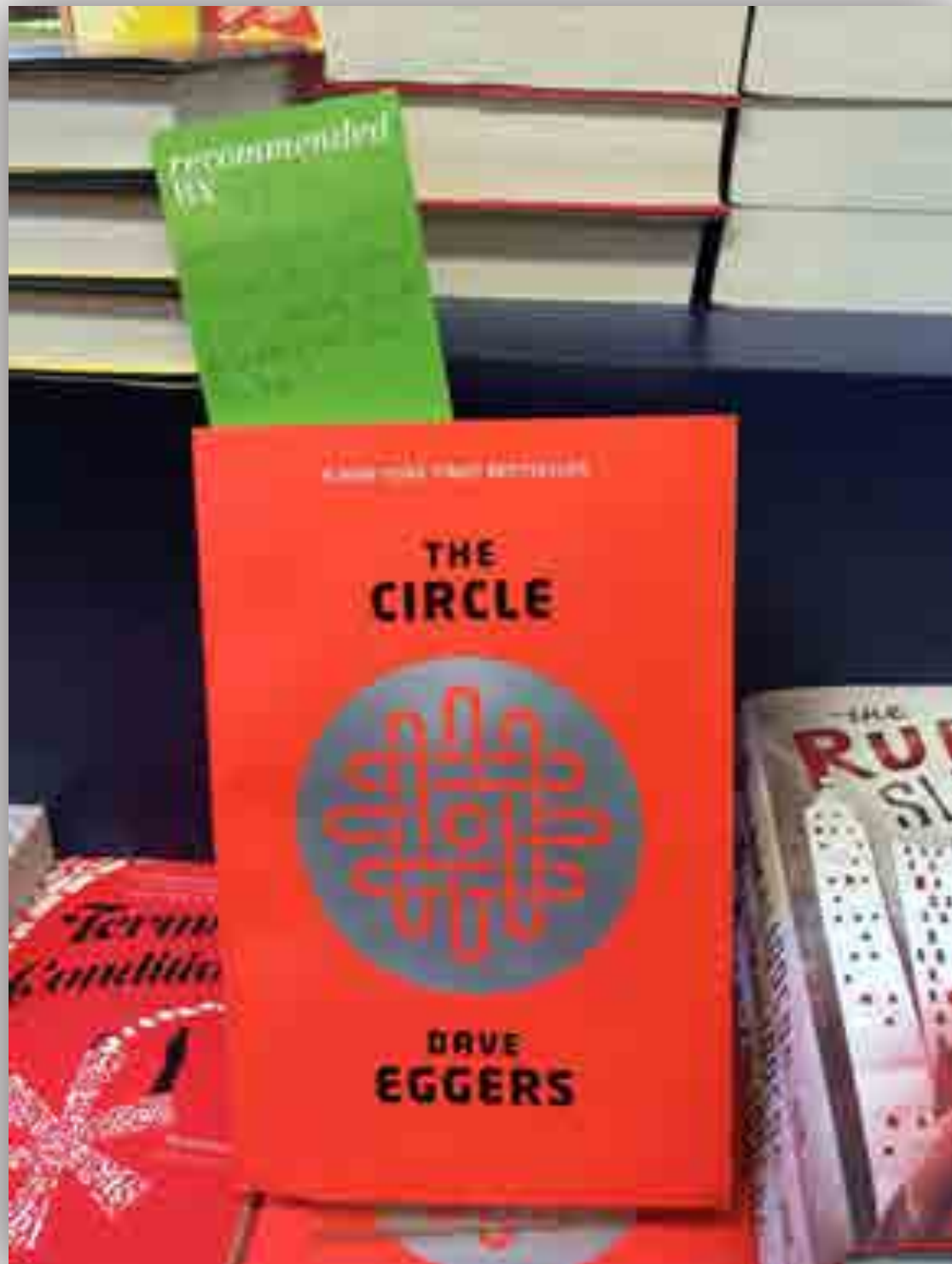
big data brings big challenges

The challenge: from information to intelligence

- velocity
- volume
- variety
- veracity

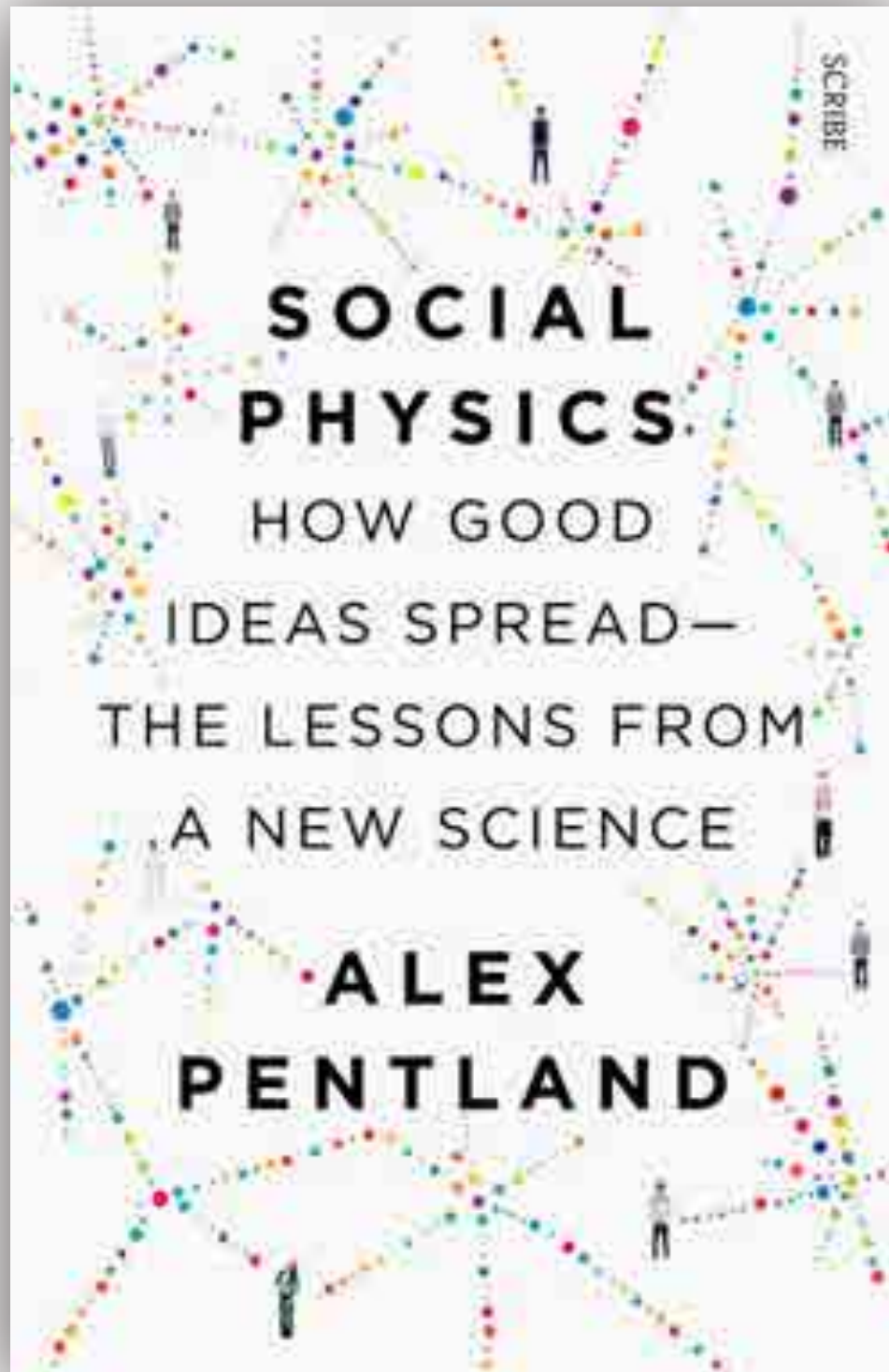
a dystopian future **the circle**

<http://www.amazon.com/The-Circle-Dave-Eggers/dp/0385351399>



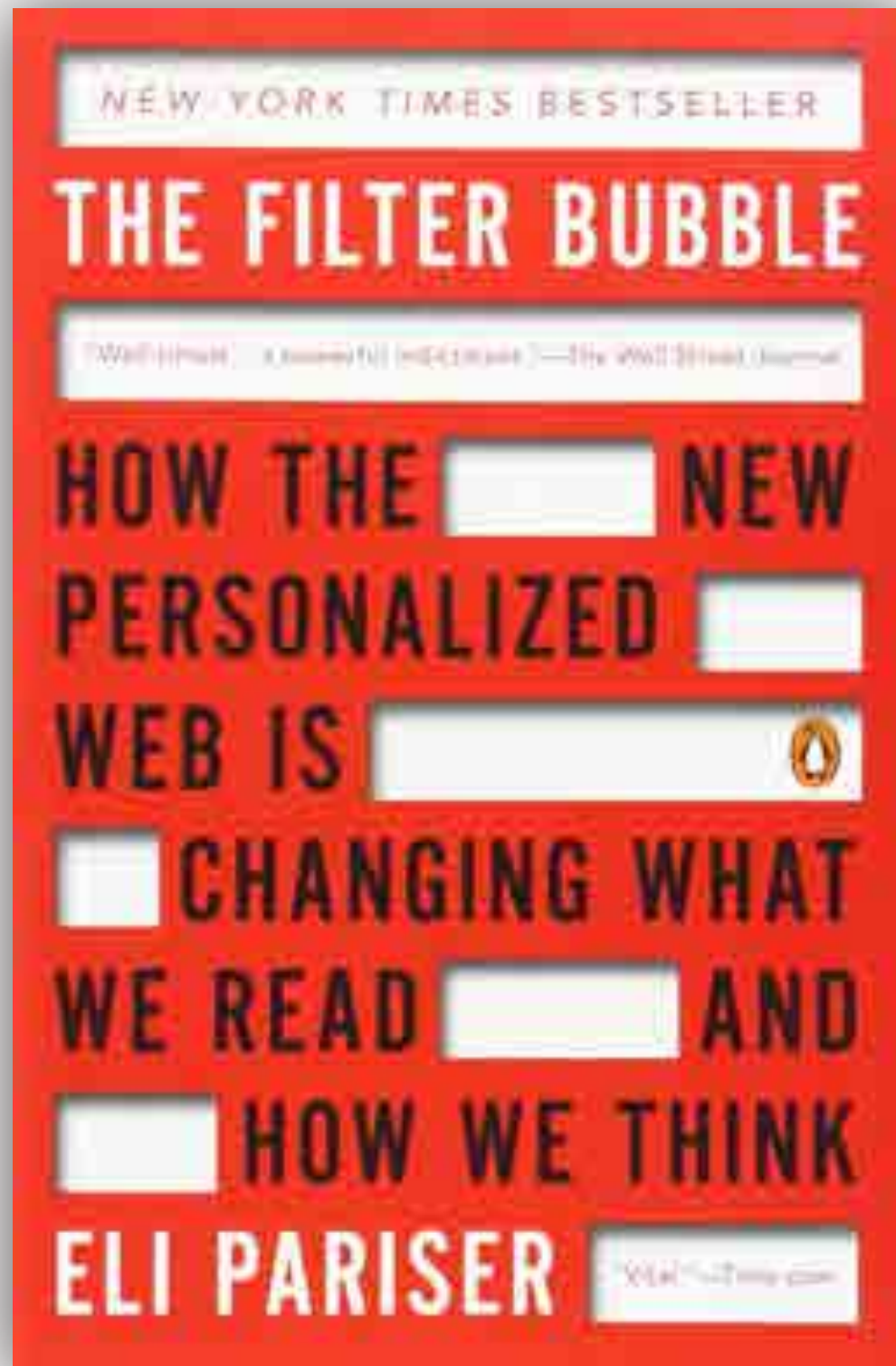
a sobering present **social physics**

<http://www.amazon.com/Social-Physics-Spread%C2%97The-Lessons-Science/dp/1594205655>



the filter bubble **algorithmic tailoring**

<http://www.amazon.com/The-Filter-Bubble-Personalized-Changing/dp/0143121235>



big data vs. rights

more data, more access vs. bounded use, sunset clauses

privacy as a luxury

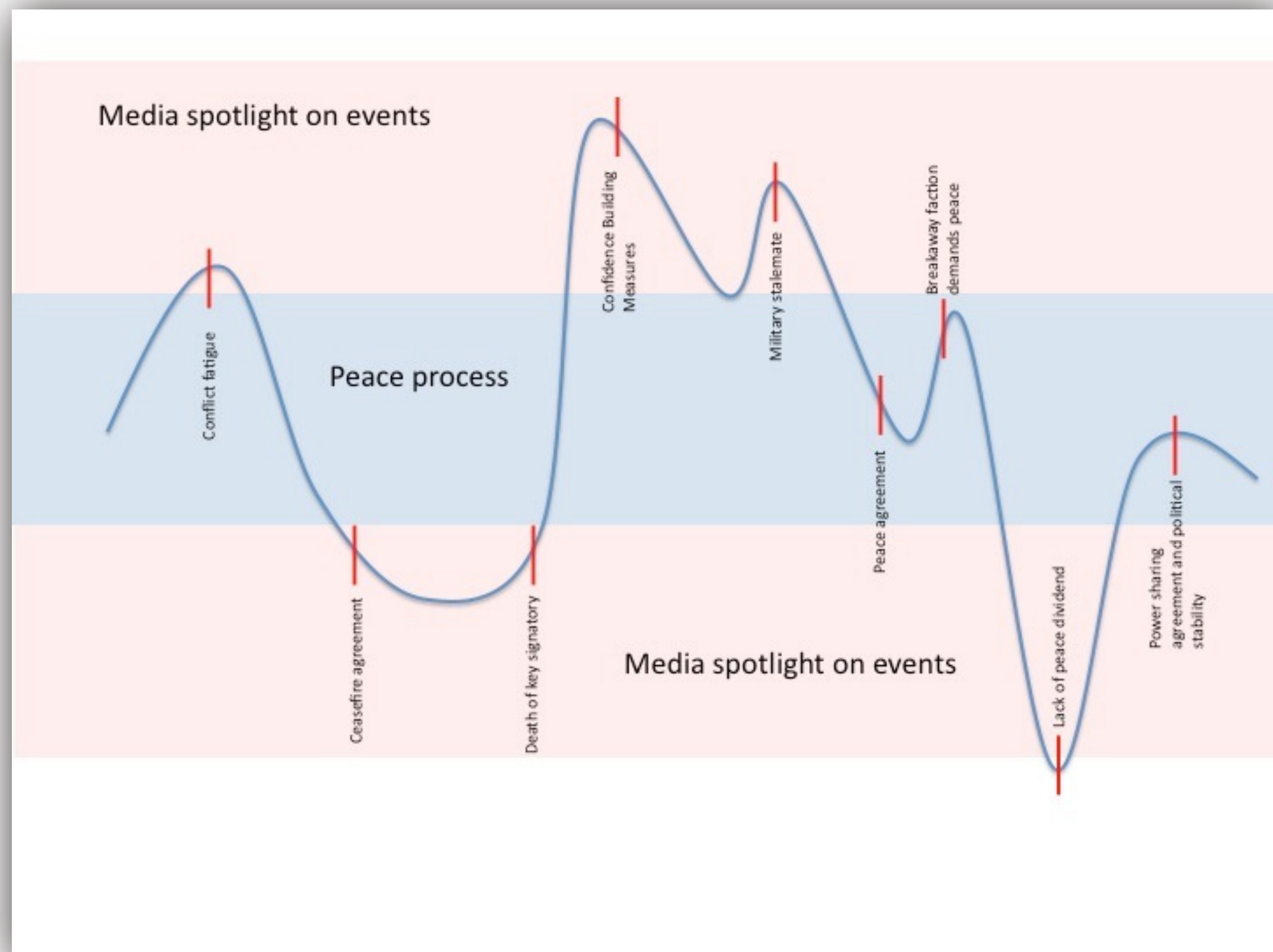
commodification of human rights

platforms vs end users

cost of access determines ways of use

process

event



frequency vs. value

Is more always better?

@ReallyVirtual



Sohaib Athar

@ReallyVirtual

Follow

Helicopter hovering above Abbottabad at 1AM (is a rare event).

7:58 PM - 01 May 11

3,440 RETWEETS 2,565 FAVORITES



Sohaib Athar

@ReallyVirtual

Follow

Go away helicopter - before I take out my giant swatter :-/

8:05 PM - 01 May 11

968 RETWEETS 814 FAVORITES



Sohaib Athar

@ReallyVirtual

Follow

A huge window shaking bang here in Abbottabad Cantt. I hope its not the start of something nasty :-S

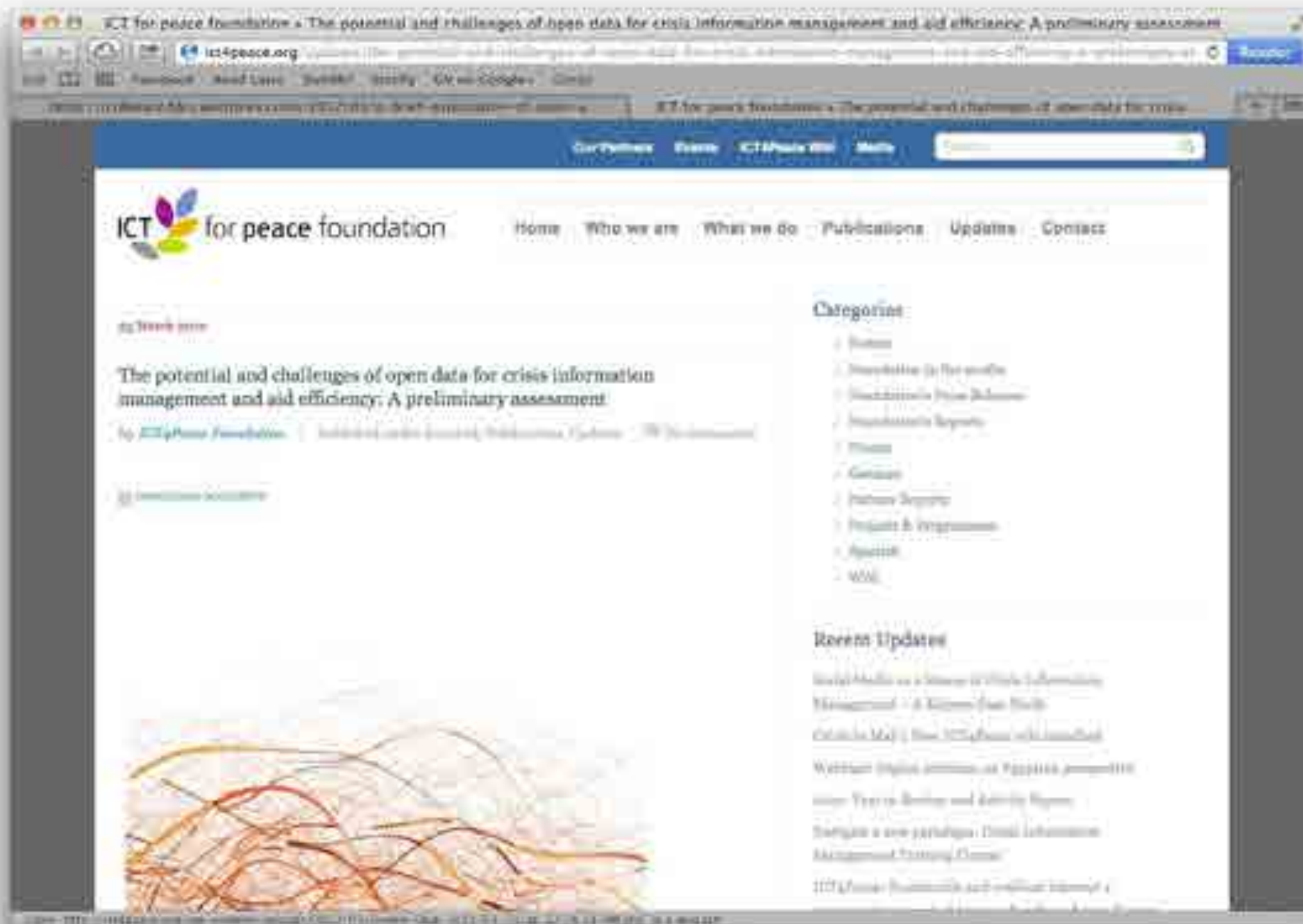
8:09 PM - 01 May 11

1,944 RETWEETS 1,100 FAVORITES



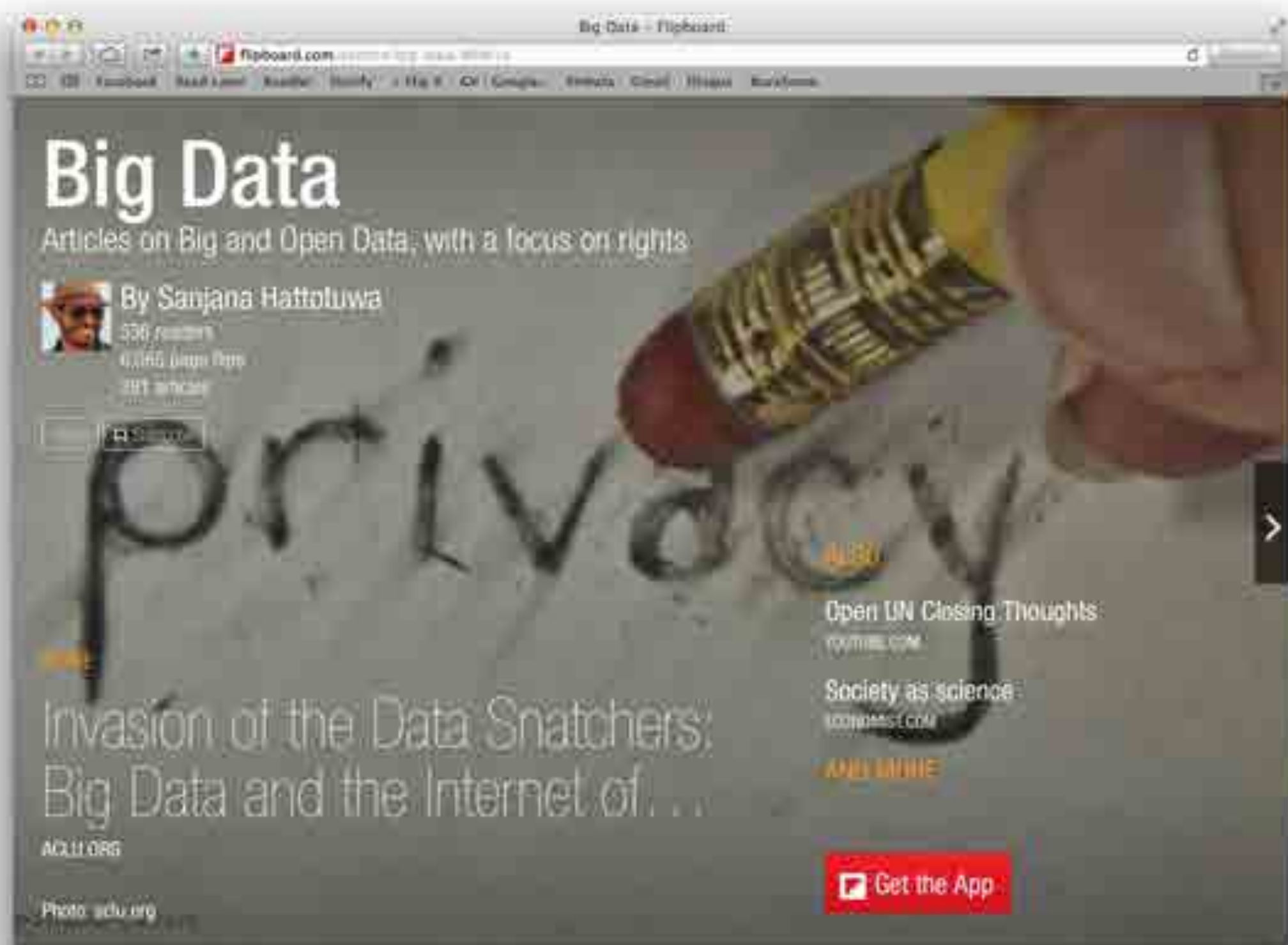
ict4peace report

<http://ict4peace.org/updates/the-potential-and-challenges-of-open-data-for-crisis-information-management-and-aid-efficiency-a-preliminary-assessment>



flipboard magazine on big data

<https://flipboard.com/section/big-data-b94t1n>



tiki toki timeline on big data

<http://www.tiki-toki.com/timeline/entry/132920/Big-Data>



iccm 2013 panel on big data

<https://www.youtube.com/watch?v=9M8Hq948Gv8>



ICCM 2013 Panel: What's so big about Big Data? - YouTube

www.youtube.com

Facebook, Read Later, Email, Print, The TV, CP, Google, Twitter, Sound, Embed

Recommended for you: Welcome to YouTube (2:28:54), Connecting - Google & YouTube, Big Data Symposium, ICCM 2013 Panel: What's so...

YouTube

ICCM 2013 Panel: What's so big about Big Data?

276 views

Published on Jan 13, 2014

Host: ICT4Peace Foundation
Moderator: Benjamin Hattis (ICT4Peace Foundation)
Presentation: Jon Gruber (DMA Corp), Respondents: Arun-Asale
Kacust (Hemmes), & Emmanuel Jafard (California Berkeley)

Big Data is the new black. Newsweek reports, schools and even
winter books are devoted to how Big Data, variously defined, offers
new ways to better our lives. Humanitarian aid just lists

Recommended videos:

- Maker Faire 2011 - The Future of Education Panel
- BBC, November 2011 The Age of Big Data
- 'Is Happiness up to us?' panel at Happiness & Its Causes 2013
- Introduction of Big Data Open Source: Map reduce, Hive, Pentaho
- Raffaello D'Andina: The astounding athletic power of quadcopters
- Seetha Menon / Heaven by Gaurav Dagaonkar (Synchronicity) OFFICIAL
- Big Data Analytics: The Revolution Has Just Begun
- Introducing Apache Hadoop: The Modern Data Operating System
- Accelerating Big Data with Hadoop (HDFS, MapReduce and HBase) and

thank you

sanjanahattotuwa@ict4peace.org