

Data-driven journalism

Mar Cabra, [ICIJ](#), Fund. [Civio](#) - [@cabralens](#)
OSCE training, Chisinau, Moldavia

**Have you done any
data projects?**

What is it? (I)

Data-driven journalism (DDJ), Computer-assisted reporting (CAR)

"I am not talking here about statistics or numbers in general, because those are nothing new to journalists. When I talk about data, I mean **information that can be processed by computers**"

- Paul Bradshaw

What is it? (II)

Gather and analyze big amounts of information and detailed data to make them understandable to the audience through articles, visualizations or applications

*To know more: [Data Journalism Handbook](#)

The future of journalism

Richard Gingras, head of News Products at Google - [video](#)

"How do we rebuild **trust**? ... I want to know the reporter"

"The investigative report of tomorrow is not 15.000 words of narrative about a story, but it's actually a **persistent investigative report that is written in query strings and fusion tables**"

Not just the future... (I)

DAY SCHOOLS.— <i>Establishments.</i>	Boys	Girls	Total	Exp.	Remarks.
Grammar School	155	155	310	1000	Treated, clothed and boarded. Taught, clothed and boarded. Largest ordinary money; i.e. do. do.
Blue Coat do.	80	80	160	500	
Green Coat do.	50	50	100	300	
Collegiate Church do.	50	50	100	300	
St. George's do.	10	10	20	100	
St. Mary's do.	12	12	24	40	Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded.
St. John's do.	9	9	18	40	
St. Paul's do.	9	9	18	40	
Ladies' Jubilee	30	30	60	500	
St. Peter's do.	21	21	42	60	
National School, Greenway-st.	194	119	313	600	Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded.
St. James's do.	200	170	370	600	
St. George's do.	823	361	1184	£510	
St. Mary's do.	400	395	795	400	
St. John's do.	176	121	297	300	
St. Paul's do.	800	361	1161	£554	Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded.
St. George's do.	201	262	463	100	
St. Mary's do.	120	110	230	100	
St. John's do.	170	143	313	100	
St. Paul's do.	68	71	139	100	
St. George's do.	314	381	695	100	Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded.
St. Mary's do.	141	119	260	100	
St. John's do.	110	143	253	100	
St. James's do.	102	98	200	100	
St. Michael's do.	234	332	566	£100	
St. Peter's do.	121	120	241	100	Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded.
St. Clement's do.	90	90	180	100	
St. George's do.	305	1071	1376	100	
St. Mary's do.	181	207	388	100	
St. John's do.	120	143	263	100	
St. James's do.	102	98	200	100	Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded.
St. Michael's do.	234	332	566	£100	
St. Peter's do.	121	120	241	100	
St. Clement's do.	90	90	180	100	
St. George's do.	305	1071	1376	100	
St. Mary's do.	181	207	388	100	Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded.
St. John's do.	120	143	263	100	
St. James's do.	102	98	200	100	
St. Michael's do.	234	332	566	£100	
St. Peter's do.	121	120	241	100	
St. Clement's do.	90	90	180	100	Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded.
St. George's do.	305	1071	1376	100	
St. Mary's do.	181	207	388	100	
St. John's do.	120	143	263	100	
St. James's do.	102	98	200	100	
St. Michael's do.	234	332	566	£100	Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded.
St. Peter's do.	121	120	241	100	
St. Clement's do.	90	90	180	100	
St. George's do.	305	1071	1376	100	
St. Mary's do.	181	207	388	100	
St. John's do.	120	143	263	100	Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded.
St. James's do.	102	98	200	100	
St. Michael's do.	234	332	566	£100	
St. Peter's do.	121	120	241	100	
St. Clement's do.	90	90	180	100	
St. George's do.	305	1071	1376	100	Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded. Treated, clothed and boarded.
St. Mary's do.	181	207	388	100	
St. John's do.	120	143	263	100	
St. James's do.	102	98	200	100	

1821 - The Guardian

Manchester schools, students and cost

Not just the future... (II)

- Detroit protests 1967 Philip Meyer



More [here](#) and also [here](#).

Data Journalism Awards (I)

Let's look at the winners!

DATA-DRIVEN INVESTIGATIONS

National/international (about - court cases)

EXCLUSIVE! terrorists for the FBI
Inside the bureau's secret network that surveils and entraps Americans

SEE THE MAP | WATCH THE VIDEO | READ INTERNAL FBI DOCUMENTS

PROFILES IN TERROR
Click on an image to view information about each terrorism defendant, or [browse the whole database](#).

The Informants
The FBI has built a massive network of spies to prevent another domestic attack. But are they busting terrorist plots—or leading them? —**Trevor Aaronson**

Locked Up Abroad—for the FBI
How our government enlists foreign regimes to detain and interrogate American citizens. —**Nick Baumann**

CRUNCH THE NUMBERS
We analyzed the prosecutions of 508 alleged domestic terrorists. View them by alleged terrorist group or state, or play with the [full data set](#).

<All States> Go

Data Journalism Awards (II)

Local/regional (about - crossing databases)

METHADONE AND THE POLITICS OF PAIN

Silent deaths

Politics of pain

A troubled clinic



Sara Taylor, mother of former King County 911 dispatcher Angeline Burrell, describes her daughter's battle with disabling chronic pain and her overdose death from methadone. At left is Taylor's husband, Dale.

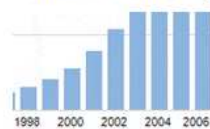
State pushes prescription painkiller methadone, saving millions but costing lives

To cut costs, Washington steers Medicaid patients to a narcotic that costs less than a dollar a dose. The state insists methadone is safe. But hundreds die from it each year — and more than anyone else, it's the poor who pay the price.

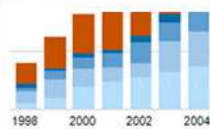
MORE IN THE SERIES:

- Update: 'Preferred' pain drug now called last resort
- Update: State plans emergency warning on risks of methadone
- 📺 Video: Methadone's toll
- 📅 Timeline: State defends methadone as deaths rise
- Silent deaths: source documents
- How we did it: our analysis
- Seattle Times methadone investigation wins 2012 Pulitzer Prize
- 🏆 Times awarded Selden Ring Award for Investigative Reporting

Methadone consumption and deaths: Numbers tell the story



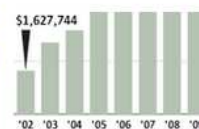
Interactive: Use has skyrocketed since state put drug on list



Interactive: More deaths from methadone than other drugs



Interactive: In Washington, victims are clustered in poorer areas



Graphic: The costs of hospital care linked to methadone have skyrocketed



Interactive: Nationwide, Washington stands out as one of the worst states

Data Journalism Awards (III)

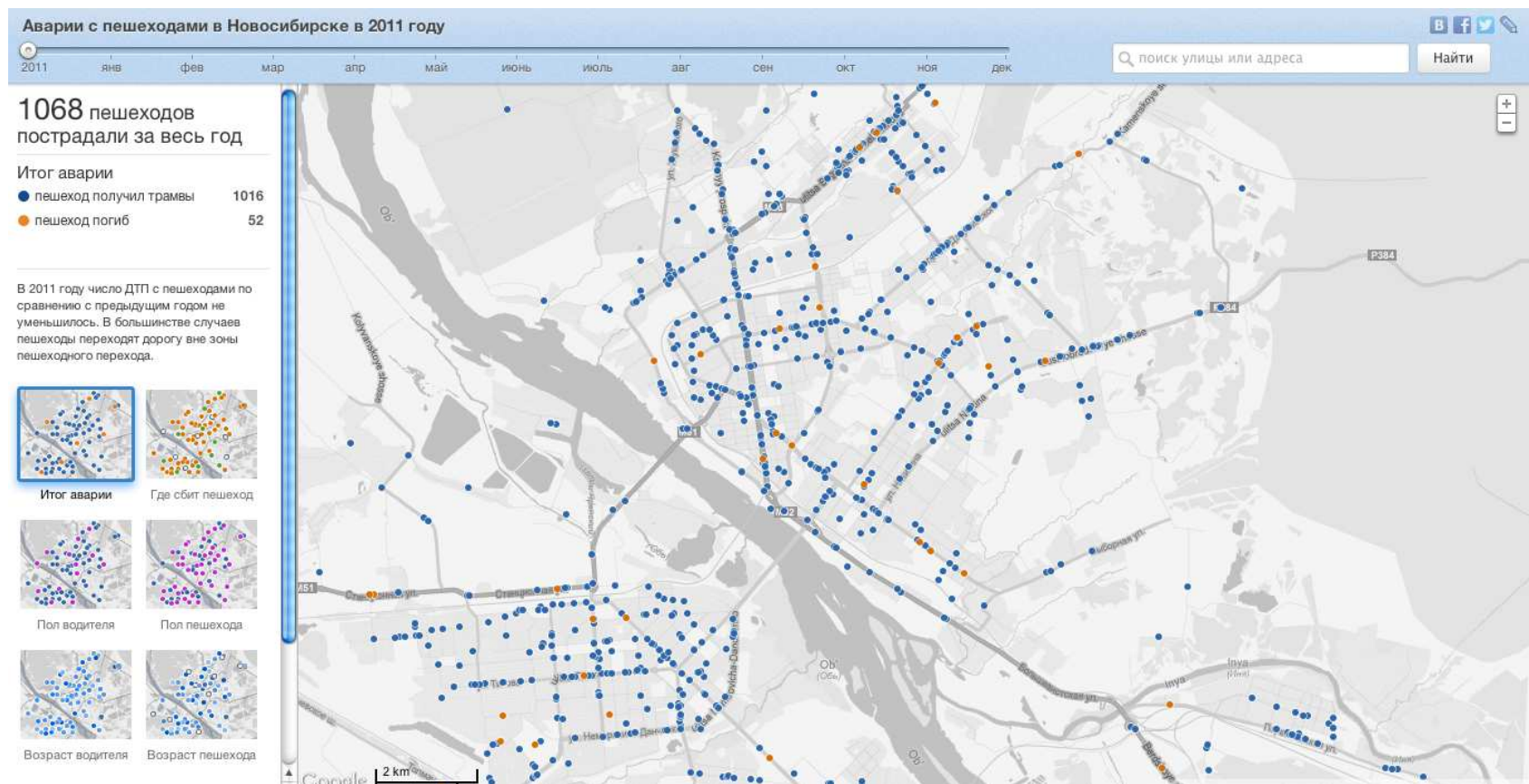
DATA VISUALIZATION AND STORYTELLING

National/international (*Twitter as a source*)



Data Journalism Awards (IV)

Local/regional (Russia - source & data - video)



Data Journalism Awards (V)

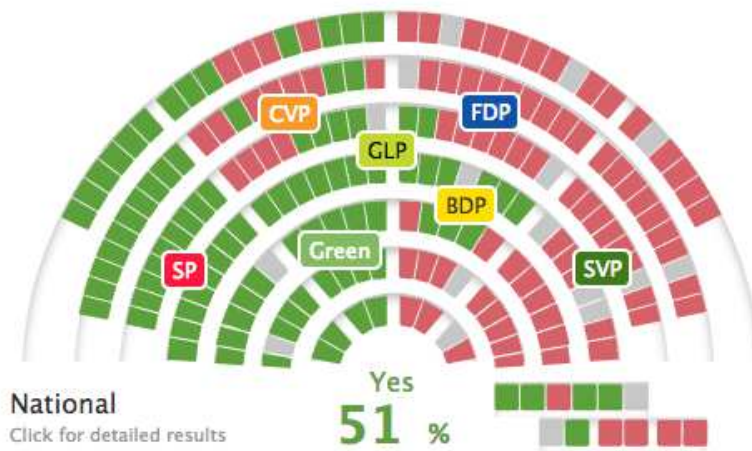
DATA-DRIVEN APPLICATIONS

National/international

["Session Overview"](#)

General tax amnesty

Vote in the National from Friday 28th September



Navigation

[«previous vote](#)

[vote next »](#)

How do the cantons?



A gender issue?



[Only Women Show](#)

[Show Only Men](#)

Data Journalism Awards (VI)

Local/regional

2011 Illinois school report cards

 Like  530 people like this. Sign Up to see what your friends like.

Test scores, class sizes, district finances and more

ex. Whittier Elementary, Peoria, or DuPage

ex. 5046 S. Greenwood Ave Chicago, IL 60615

Key findings in school report cards



How safe is your child's school? It's hard to find out

June 15, 2011

New Data: Crimes Reported at Illinois Schools

To investigate this story, Tribune reporters obtained detailed school crime information which was reported to the state through the SIRS system. Visit individual school pages on this site to see which crimes were reported at that school.

Where does
your online
education
lead you to?



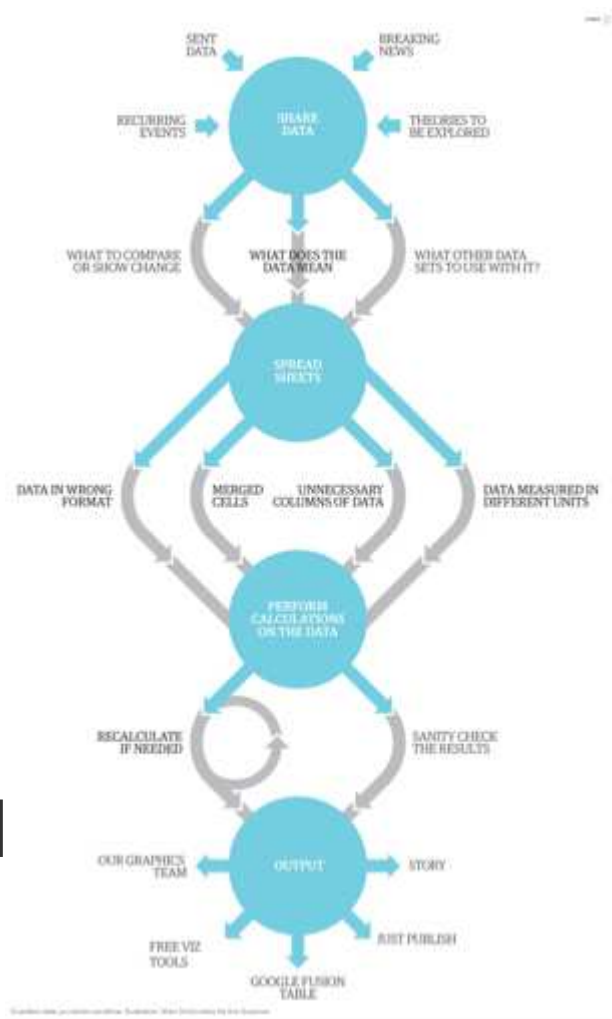
Things in common

- Analysis and interaction
- Multimedia not only as video
- Transparent about sources and method - TRUST
- Show data and documents
- Collaboration with computer engineers, "hackers" ([Hacks&Hackers](#), [CRJI](#), [OCCRP](#))

Hands to the dough!



A "messy" process



TH

headline?

"Comment is free,
but facts are sacred"

The Guardian

*Spend time

*What's the

1. GETTING THE DATA (I)

- Data is everywhere!
- Structured v non structured ([life](#), sentences)
- Does it already exist?
 - * [proactive publication](#)
 - * open data - [Moldova](#), [Belarus blog](#) + [data](#) ([idea!](#))
 - * [official gazette](#)

1. GETTING THE DATA (II)

Use Freedom of Information!!

- In your country:

MOLDOVA

- Constitution art. 34 + 37 (environment, health and consumer)
- Law in force since August 2000
- 15 days

1. GETTING THE DATA (III)

UKRAINE

- FOI in January 2011
- Aarhus applies (environmental information)
- Difficult to get access to info on public spending, personal details about politicians, public procurement etc. ([more info](#))
- Recommendations:
 - *push the limits!**
 - *write about the process**

1. GETTING THE DATA (IV)

BELARUS

- No FOI
- But... journalists have access through Media Law

Don't worry, in Spain we have even worse access! :)

1. GETTING THE DATA (V)

- Think creatively - who may have information?
(eg. European Union - asktheeu.org)
- Globally: <http://www.rti-rating.org/>
- Importance of formats - .xls .csv
- *No news is good news*

1. GETTING THE DATA (VI)

- Ultimate goal - get it into a spreadsheet!
They're beautiful!
- Different formats - conversion
 - * cometdocs.com
 - * pdftoexcelonline.com
 - * zamzar.com
- Tables from web - [Table2Clipboard](#) - [TableCapture](#) ([example](#))

1. GETTING THE DATA (VII)

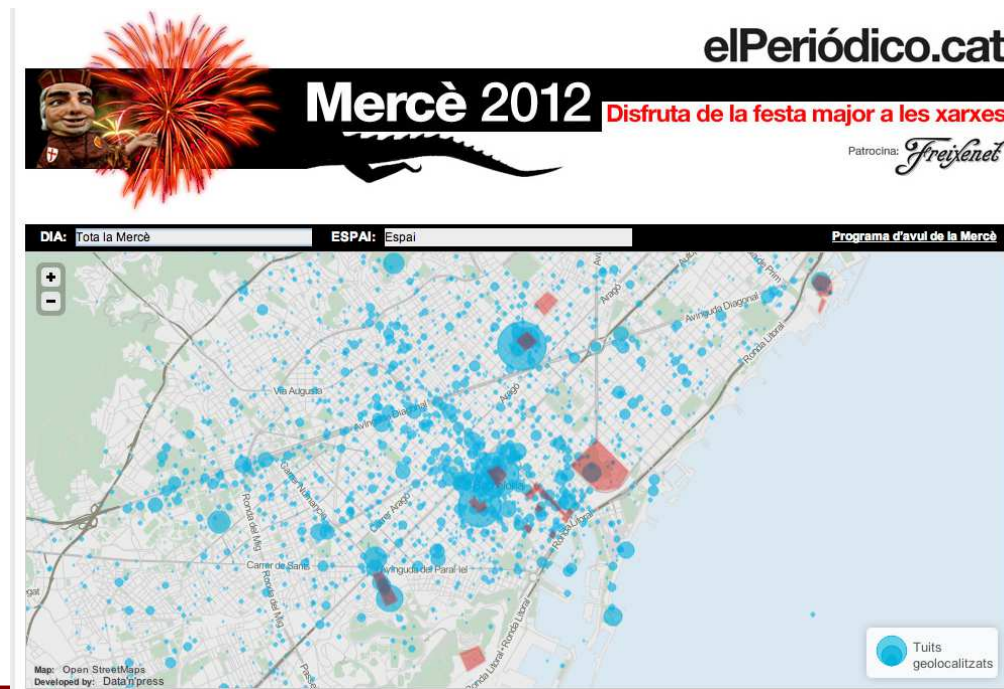
- Web scraping - [Spanish contracts](#)
 - * [Dapper](#)
 - * [ScraperWiki](#)
- Work with local hackers
 - * Open Data movement
 - * [Hacks and Hackers](#)

1. GETTING THE DATA (VIII)

- Real time data

- *Spot patterns - and news

- *Who's tweeting what? - holiday in Barcelona



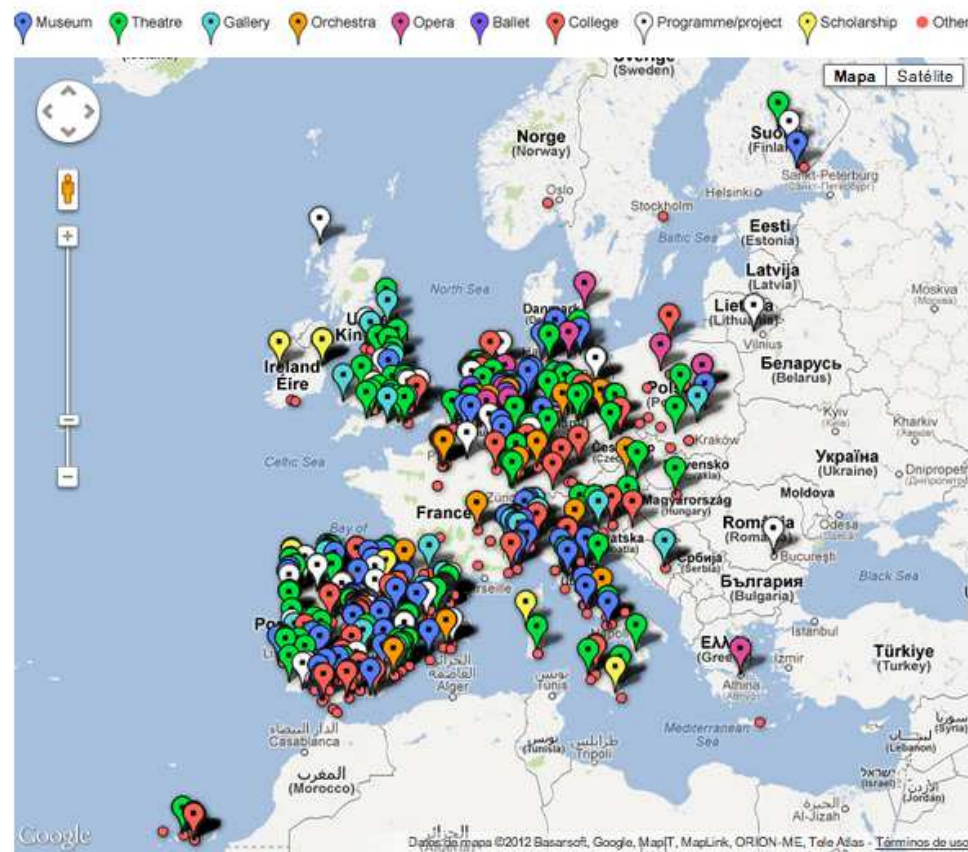
1. GETTING THE DATA (IX)

- You (media) have data - eg. [crime news](#)



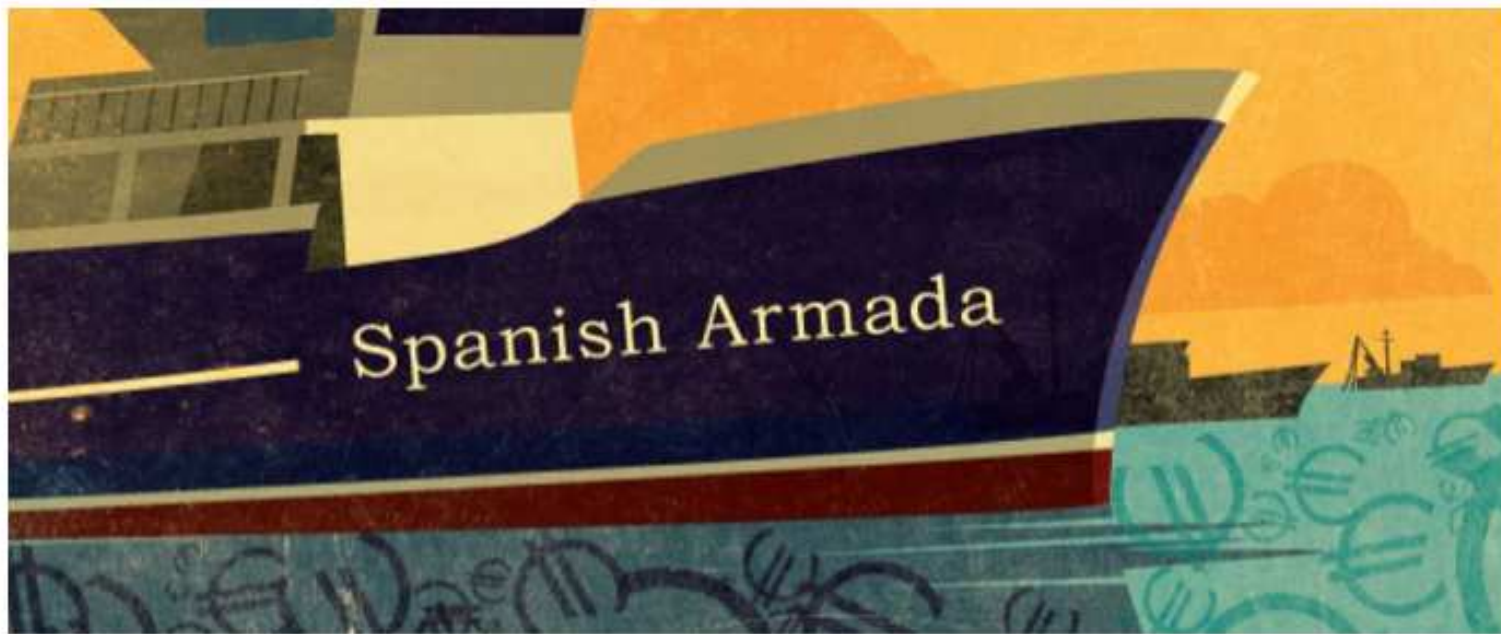
1. GETTING THE DATA (X)

- Users generate data - European culture cuts



The "millionaire" fish |

Nearly €6 billion in subsidies fuel Spain's ravenous fleet



Ajani Winston/Watch News

As stocks crash, 1 in 3 fish paid for with public money

By Kate Willson, Mar Cabra and Marcos Garcia Rey 12:01 am, October 2, 2011 Updated: 5:23 pm, January 24, 2012

The "millionaire" fish II - Sources

- Subsidies at a beneficiary level ([fishsubsidy](#) and [Ministry](#)).

But who gives what?

- [Regions](#)/BOE to fill in holes
- Licenses for foreign fishing
- Other reports

The "millionaire" fish III

KEY FINDINGS

- The Spanish fishing industry has received **more than €5.8 billion** (more than \$8 billion) in subsidies from the EU and Spain since 2000 – far more than the industry of any other EU country.
- Subsidies account for a third of the sector's value. Simply put, **nearly one-in-three** fish caught on a hook or raised in a farm is paid for with public money.
- More than **80 percent** of subsidized fishing companies that were fined in Spain for fishing infractions – and then lost subsequent court appeals – continued to receive subsidies.

2. ANALYZING THE DATA I

CLEANING - [Google Refine](#)

- Look for: names in different order, accents/other letters, SL v S.L. v S.L

[Example](#)

2. ANALYZING THE DATA II

INTERVIEWING THE DATA

- The tool:
 - *Excel (PC 2010, Mac 2011)
 - *Open Office
 - *Google Docs
 - *[Manual](#) I recommend

2. ANALYZING THE DATA III

- Freeze panes
- Order, count
- Filter
- Formulas:
 - =SUM(first cell:last cell)
 - =SUBTOTAL(9,A1:A30)
- Percentages

2. ANALYZING THE DATA IV

- Average
=AVERAGE(A1:A30)
- Median
=MEDIAN(A1:A30)
- Pivot table

3. PUTTING DATA IN THE STORY (I)

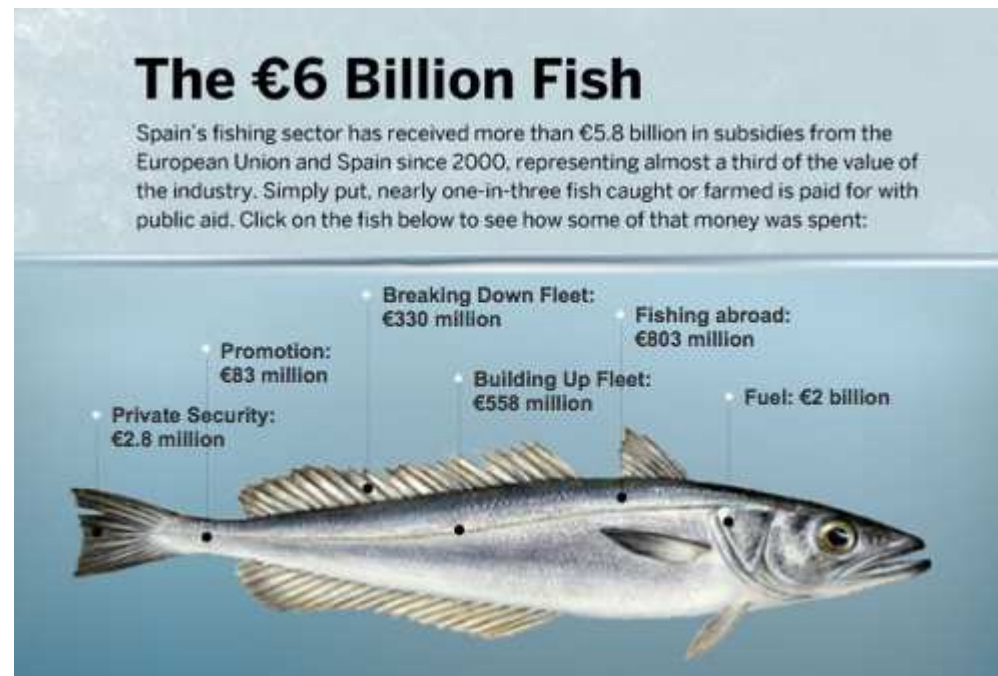
IN THE COPY

- Do not overload with facts
- Representative detail ([WSJ example](#))
- Best summary

3. PUTTING DATA IN THE STORY (II)

VISUALIZING DATA

- Graphic
([fish](#))



3. PUTTING DATA IN THE STORY (III)

VISUALIZING DATA

- Timelines

- *To analyze: [TimeFlow](#)

- *To see: [Dipity](#)

3. PUTTING DATA IN THE STORY (IV)

VISUALIZING DATA

- Maps and charts - [Google Fusion tables](#)

- *[Localize](#)

- *[Show areas](#)

- [Example](#)

3. PUTTING DATA IN THE STORY

- Show the documents ([Document Cloud](#), [Scribd](#)) - [example](#)

[illegible]

3. PUTTING DATA INTO THE STORY

VI

Other examples

- dondevanmisimpuestos.es
- Congressmen assets:
 - * [source](#)
 - * [crowd sourcing](#)
 - * lainformacion.com

3.PUTTING DATA INTO THE STORY

VII

-
- The importance of a [methodology](#)
- Just state the sources and how you got the facts

How to learn more?

- [The Guardian Data Blog](#)
- [Data driven journalism](#)
- [Center of Investigative Journalism](#) Summer School
- [Investigative Reporters and Editors](#)
- [Reporters Lab](#)

Thanks!

Mar Cabra

mar.cabra.valero@gmail.com

Twitter: [@cabralens](https://twitter.com/cabralens)

Skype: mar.cabra