

How open data is redefining the roles of the journalist, audience and publisher

BBC Position paper for the workshop on USING OPEN DATA:
policy modelling, citizen empowerment, data journalism
May 2012

Andrew Leimdorfer, BBC News
Olivier Thereaux, BBC R&D

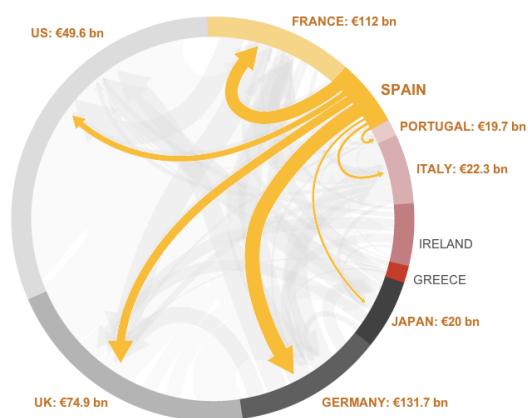
At first glance, data journalism could be mistaken for being nothing new at all. Journalism has always been, at its core, a group activities revolving around the collection of facts, the extraction of new narratives from those facts, and the presentation of the narratives in a way that the general public will find interesting, entertaining or personally relevant. Replace “facts” with “data” and you will have almost exactly one of the definitions of data journalism from the [Data Journalism Handbook](#) [1] published this year. The use and publication of Open Data, however, makes data journalism much more than mere business as usual: it blurs the boundaries between the roles of publisher and user of data.

Open data, by definition, is rather inaccessible and hardly meaningful to most people. Tables of numbers, rows of dates or lists of people will be appealing to the specialist looking for precise information, but less so to the general public.

As a public broadcasting organisation whose mission has been to “*inform, educate and entertain*” since its early days in the 1920s, the BBC cares about this challenge to get over information to a large audience in a way that is both accessible and meaningful.

Like others in the news and media industry, the BBC has been using interactive tools and visualisation on the web to help make complex information more appealing to a mass audience.

We have found formats such as our [budget calculator](#) [2] to succeed in making data more meaningful by anchoring its model on information which is personally relevant to its users.



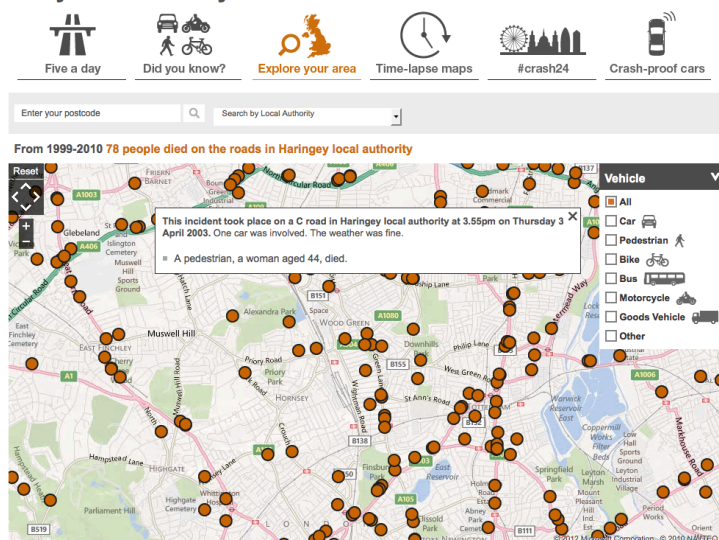
Eurozone Debt, Nov. 2011

Likewise, visualisation, can help the general public gain some understanding of complex mechanisms, such as [foreign debt](#) [3]. We believe, however, that simplified visualisation may not always be a panacea. Some visualisations, such as [graphs of European countries' government annual surplus or deficit](#) [4], are likely to appeal to a niche, while remaining opaque to most.

"I was no good at Maths at school so I don't want to be reminded of that in my free time"
— BBC News Specials User Research

Indeed, in some cases, it is best to keep a significant amount of complexity available for anyone to peruse. We increasingly endeavour to link to the research material that many of the science and health stories covered by the BBC are based on, even if this material may not be appealing to most of our audience, or even if they may only be partly available outside of academic pay-walls. Linking out in this way provides both a trust mechanism (provenance and access to source material) and an opportunity for our audience to participate in fact-checking.

Every death on every road in Great Britain 1999-2010



*Every death on every road in Great Britain 1999-2010.
Published Dec 2011*

The same principles apply when a news outlet chooses to publish data (e.g the [public sector pay exploration](#) which we offered in 2010 [5]) in a relatively raw format, or when they don't create stories from data but work on the equally difficult task of creating data from stories, as the BBC does with [automatic tagging of its archives](#) [6] or when the Guardian [turns a policy speech into a spreadsheet of measures and aspirations](#) [7]: it is an invitation for others to join the search for innovative ways of combining this data with other sources and present it in an engaging

manner. Our experience in publishing data views of our knowledge on programmes, music, food or the natural world has shown in the past that there are many ready to answer that invitation.

It may then be more useful and accurate to consider using and publishing open data not as discrete and separate activities, but as roles in a spectrum ranging from from investigating and gathering the data to the interpretation, publishing and consumption.

Between a government agency exposing data in a standard structured format to the reader of a news story interested in the impact of a policy on their lives, there lies a complex array of possible actors, from people debating the impact of policies on blogs, to journalists (and citizen journalists) interpreting or presenting the data in stories through a variety of media. We can all be part of a web of open data user-publishers. Using open data means republishing it in a different, consolidated or curated format, or in a way which makes it easier to explore and make sense of.

Much of the value of the data work done by News organisations is derived from the “sense-making”; creating content and applications that make a complex world easier for people to understand. Alongside this function there is growing value in making the data we use available and linking to other interpretations, representations and combinations of that data. The task at hand is to find organisational and technical solutions to manage links between datasets and their interpretations and create something meaningful from that web.

We look forward to the upcoming workshop on policy modeling, citizen empowerment and data journalism, and the opportunity to exchange with other participants on the best practices and challenges facing this new ecosystem of user-publishers of open data.

Table of links

- [1] http://datajournalismhandbook.org/1.0/en/in_the_newsroom_1.html
- [2] <http://www.bbc.co.uk/news/business-17442946>
- [3] <http://www.bbc.co.uk/news/business-15748696>
- [4] <http://www.bbc.co.uk/news/business-13366011>
- [5] <http://www.bbc.co.uk/news/uk-11319918>
- [6] <http://www.bbc.co.uk/blogs/researchanddevelopment/2012/03/automatically-tagging-the-world.shtml>
- [7] <http://www.guardian.co.uk/news/datablog/2012/may/09/full-list-queens-speech-measures>