Visualization for the Masses
Making public data more accessible

This paper will look at the role visualization and user interface can play in making public data more accessible, and highlight some of the problems associated with data production and presentation. Drawing on two recent real-world examples based on government data, issues of design and methodology will be explored, with observations on curation and deployment along the way. One of the projects under consideration is still in production, so more details will be available on presentation.

Introduction

In the age of the Internet, visualizing public data is of great importance to governments and the general public alike. The Web offers unprecedented opportunities for civic engagement, but better access is needed to promote dialogue. As more and more data becomes available in electronic format, we need new tools which improve access and facilitate analysis in order to make the most of these resources.

Nowadays, huge volumes of data are produced using specialized proprietary applications, which frame information in a way that can be easily understood, in support of formalized activities. But there are many exceptions where information is produced more freely, by data owners creating custom systems based on generic tools, and this is often the case with public data. In these instances, two very common output formats are spreadsheets, and declarative mark-up, possibly in the form of hypertext. Of the two, spreadsheets offer significant advantages in terms of usability and analysis: they are very easy to understand, and are familiar to a large cross-section of users. By contrast, data management online is usually done using web-based forms, with many associated usability issues.

Public data produced in these formats is almost always non-standard. Curating and redeploying material presents problems, and is frequently an iterative process. Production is prone to human error, so reliability is routinely called into question. Combining data from different sources is also a challenge. Visualization of data is not a panacea for these ills, but it can be extremely valuable in helping data owners understand, manage and maintain their archives.

Finding meaning

In creating visual platforms, the designer should seek to answer two very important questions, namely:
• what is the story behind the data, and
• who will be using the platform, and what are their requirements?
In answering these questions, we can build up narratives from which we can derive the appropriate graphical model and iconography, and devise suitable methods of interaction for audiences to explore the information.

Case Study I: A visual browser for the London Gazette

The London Gazette, the UK Government's Official Journal and Newspaper of Record, has been in circulation since 1665. Every working day, a new edition is issued, containing 300-500 public notices. A typical Gazette notice comprises an announcement which the applicant is required by law to make, such as urban planning notices, permit applications or insolvency reports. Alongside the Edinburgh and Belfast Gazettes, the journal is published by The Stationery Office (TSO) under contract to the Office of Public Sector Information (OPSI). The archive therefore forms an authoritative record of public activity, locally and nationally. [1]

In 2007, following Steinberg and Mayo's Power of Information Review, the London Gazette archive began a process of technological transformation. Notices in the Gazette, couched in strict legal language, can be translated very naturally into a declarative format. Using natural language processing, semantic mark up was added in line with OPSI's detailed "ontology of legislation", creating an unprecedented database resource of legislative history, which can now be reused in other contexts. [2]

Visualizing the narrative
When presenting documents, the convention is to display search results as lists, and individual items as web pages. Occasionally devices such as tag clouds are employed to help illuminate the contents. But metadata in the form of mark-up makes a different approach possible. By extracting events, locations and keywords, we can arrange notices into a visual form which communicates volume over time, identifies locations, and allows filtering according to keyword by type.

The first prototype was based on a simple premise. Legislation in early 2007 required agricultural installations raising livestock to reapply for pollution control permits. A visual browser makes it possible to see a high density of permit applications in this sector immediately after the change in the law, which tails off during the ensuing period.

Gazette notices are tightly worded. They follow a strict legal format, which allows dates, locations and legal entities to be clearly identified. Hence, mark-up follows a predefined pattern, in which keywords are classified according to the London Gazette ontology, and can be identified as businesses, installation types, government agencies and so forth.

In the visualization, notices are organized into a timeline, which users may browse, search and filter. Compound filters using Boolean logic offer targeted drill down, framing and visualizing relationships between complex queries. A concurrent map display shows the location of installations and government agencies, which are identified using postal codes. The resulting experience gives a unique perspective on the contents of the archive.

A note on performance
Over the two years in question, up to 900 notices are presented. All need to be accessible from the screen, but not all are visible over this time period, and displaying them all at once diminishes performance. To keep the interface running smoothly, only the notices required to indicate density and date stamp are shown, reducing the number of items on screen by up to 75%.
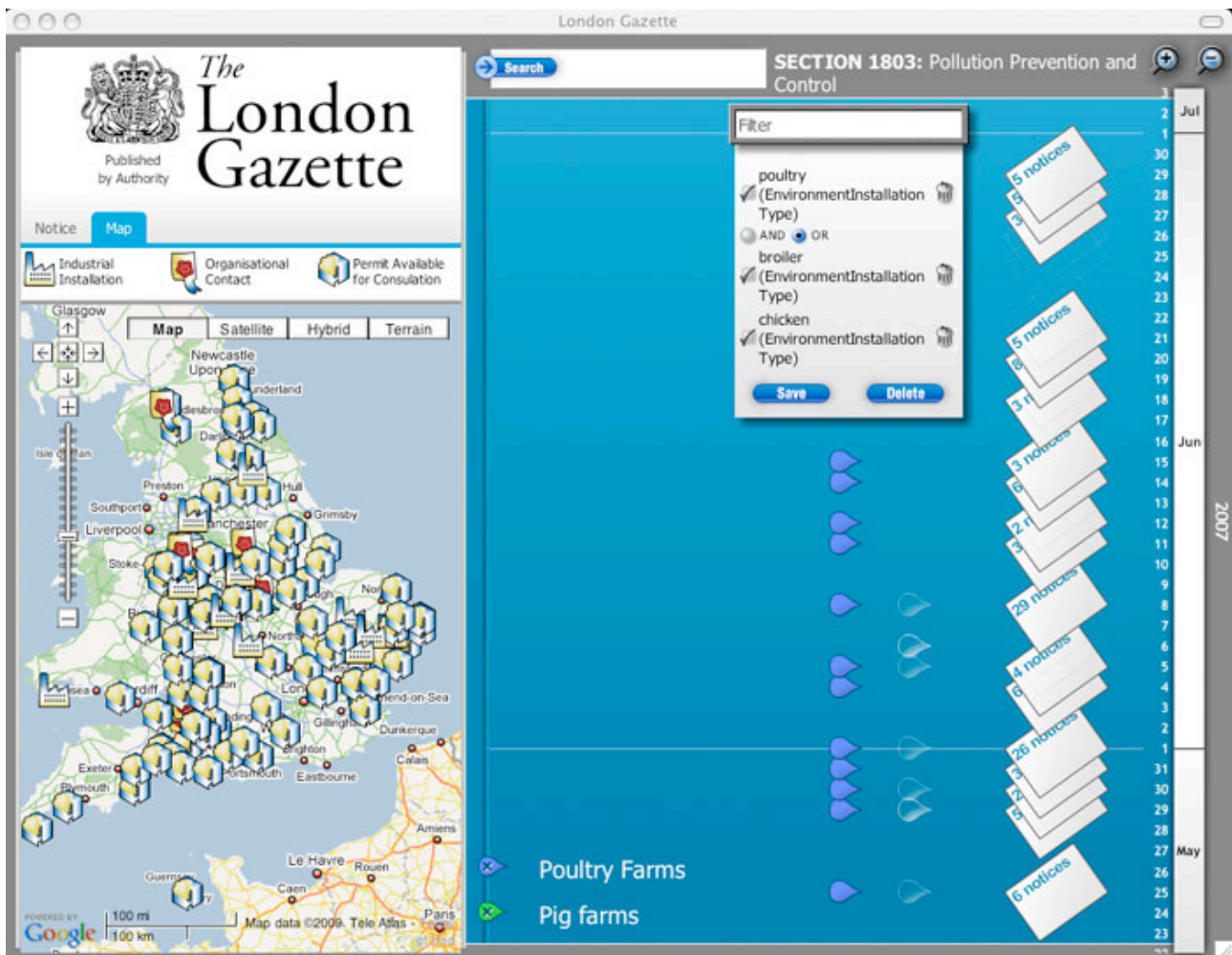
Improving the data

The first pass of visualization was very effective at highlighting problems with data. Incorrect or incomplete notices were quickly identified, as well as several undetected issues with language processing. Qualitative improvements were also made possible: one learning gleaned from the visualization was that further distinction was needed to distinguish between permit issuing agency and holding agency. This subtle but important difference added important detail for users wishing to locate permit applications, and contest them if necessary.

Understanding user behavior

Visual browsers carry a degree of unfamiliarity as well as novelty for users, so designers should take care to offer appropriate affordances, feedback and reward throughout the interface, by way of guidance. A good way to keep user's attention through the learning curve is to introduce an element of play into the process.

Since it is more or less impossible to anticipate user's responses and expectations, it is vital that designs are tested regularly. A small group of around three users can identify major problems in a single round: we test and and fix in cycles. [3]

Case Study II: Where Does My Money Go?
"A project to analyze and interactively represent UK public finance using maps, timelines, and best of breed visualization technologies."

Where Does My Money Go? is conceived by the Open Knowledge Foundation with the purpose of making UK budgetary data more accessible to the general public by means of visualization. A winner of the Show Us A Better Way Competition, the first prototype has been funded using public money. [4]

Source material
Data is drawn from the HM Treasury's Public Expenditure Statistical Analysis data, which forms the basis of the UK budget. For the initial prototype, figures are extracted from PESA spreadsheets [5] and used to create an overview of spending across departments by central and local government. As such, information is classified according to geographical regions, and by function and sub-function.

While this provides a useful entry point into public spending, it is very limited in scope, representing fiscal provision only at the highest levels. [6] Researchers are working to find alternative resources to build up a more complete picture: the challenge is to present information in a way that the main thrust of PESA data can be understood, while allowing affordances for further drill-down once detailed data becomes available.

Quantitative models
The visualization will show quantitative data shown by region. The model needs to express comparative spending across regions, allowing sufficient detail for data to be understood while avoiding confusing or unnecessary detail, as well as conflicting data models.

Where Does My Money Go? is currently under construction. A more complete overview of the project will be available in November 2009. Screenshots will be made available on request.

Links and References

London Gazette

[1] London Gazette
http://www.london-gazette.co.uk/
Visualization demo
http://www.london-gazette.co.uk/demo/

[2] Semwebbing the London Gazette
http://2008.xtech.org/public/schedule/detail/528

[3] Don't Make Me Think, Steve Krug
http://www.sensible.com/

Where Does My Money Go?

[4] Show Us A Better Way, submission
http://www.showusabetterway.co.uk/call/2008/07/where-does-my-2.html

[5] PESA spreadsheets, 2009
http://www.hm-treasury.gov.uk/pespub_pesa09.htm

[6] Discovering Where My Money Goes I: PESA
http://blog.okfn.org/2009/09/24/discovering-where-my-money-goes-i-pesa/

About the Author

Liz Turner is the co-owner and founder of iconomical, a consultancy specializing in the design and development of visual browsers for complex archives. As well as the visualization projects described herein, the company is involved in the development of multitouch table technology and mobile applications for data owners.

Liz began her career with a degree in Mechanical Engineering from the University of Bradford. She worked for a number of years in the oil industry, before entering the field of design and development on the Internet. As an independent designer, she has established a wide-raging portfolio of clients, including the RSA, Hewlett Packard and Joost.

In 2005 she completed a European Media Master of Arts degree at the Utrecht School of the Arts Faculty of Art, Media and Technology, where her continued work on the visualization of Semantic Web resources led to the foundation of iconomical.