

By Ewan Sutherland

Opening and linking data



Introduction

In the not so distant past reports, newsletters and statistics for media and telecommunications were published on paper - indeed the IIC once maintained a library of such documents. Access to databases was quite expensive, with a requirement to purchase a subscription. Then material moved online, though typically in PDF formats, sometimes locked to limit re-use. Gradually, this model seems to be changing as governments adopt an open data or open government data (OGD) approach. There is a parallel and complex debate in the academic world about moving journals from behind pay-walls to open access.

In the previous issue of *Intermedia*, Mike Nelson set out the importance of “big data”.¹ Telcos and media corporations are obliged to think how data analytics can change their businesses, with a promise of superior performance if they can master the techniques. The importance of the analysis of data at the level of politics was illustrated by Nate Silver in his very accurate predictions of the US presidential elections, calling all fifty states correctly.²

Much of the information behind the pay-walls maintained by specialist providers is derived from public sources. The justification for charges will increasingly depend on adding real value, rather than just “scraping” public websites and press releases.

Linked open data

Everyone is familiar with the world wide web (WWW) and its uniform resource identifiers (URIs), operating at the level of documents, pointing to an EU Regulation or a white paper by a Japanese ministry. One challenge is to disaggregate the documents into data and to use meta-data to connect them as linked open data (LOD). A second is to convert government datasets, typically available “raw” or “as is”, into structures and formats with semantic meaning suitable for machine processing.

Giving third parties access to government data is intended to generate economic growth by harnessing innovative talent available to the private and voluntary sectors. Such roles have often proved difficult for governments.

The Obama Administration has been very keen to open up government data, in order to keep citizens better informed about its activities. For example, the FCC now provides open data access to information about 3 million licenses, though not yet to its Electronic Comment Filing System (ECFS) system of comments filed on regulatory matters. The Securities and Exchange Commission (SEC) not only publishes data in open formats, it also accepts data from corporations which can file their accounts using a version of the eXtensible Mark-up Language (XML). Data from the Federal Election Commission is being used

by third parties to track contributions to political parties and individual politicians, which can be linked to voting records, to understand influence within the US Congress.

Those familiar with the European Commission implementation reports on telecommunications markets, later the Digital Agenda Scoreboard, will know them as PDF files. For re-use the charts could be cut and pasted or the data retyped. Now, through an EC-funded research project known as LOD2, the underlying data are available in machine-readable form.³ Similarly, data from Eurostat is available allowing linkages to be created by third parties.

The United Nations has a portal to a considerable volume of data, including historical runs of statistics.⁴ Data can be extracted to build charts (see Figure 1 overleaf). Google provides tools to manipulate data, accessing them via APIs (e.g., Eurostat, US Census Bureau and the World Bank), allowing the creation of charts similar to Figure 1.

The EC is developing Europeana, a project that exposes data to allow search and discovery among more than 17 million items held by 1,500 cultural institutions. It incorporates the approaches of different groups (e.g., museums, and libraries) to ensure a rich metadata. The collection includes text, images and video, (e.g., many photos of television and studios, telephones and exchanges).

Broadband data

In the USA, information on the provision and quality of service of broadband Internet access is provided as the National Broadband Map (NBM), a multiagency effort, with input from the NTIA, FCC and state governments and utility commissions.⁶

The previous efforts at 5-digit ZIP code or census tract level for data collection were seen as misleading and so were reduced to the census block, a much smaller geographic unit yielding more and better data.

Additionally, there is data on the Broadband Technologies Opportunities Program (BTOP) and Broadband Initiatives Program (BIP) applications.⁷ By comparison, other countries appear to have very much less data and what they have is seldom available. Unusually, New Zealand has a broadband map.

Conclusion

There are ambiguities in the term "open government data", with open government suggesting releasing information previously held back, while open data is merely a technology that greatly eases access and enables third party innovations.

The concern for those making and seeking to influence policy-making and implementation is to make data more readily available and more easily processed.

To date, open data is much less common amongst regulators than ministries. Yet it offers interesting tools to reduce the cost of data collection from operators and to improve access to data that has been collected.

Potentially, vastly more data could be processed allowing for improvements in regulation by allowing operators and third parties access to data. Those overseeing regulators would also be better able to determine how well they perform their work.

Legal information should also be made available in open data forms, including legislation, regulations and court judgements.

For some players the opening of data represents the threat of disintermediation - displacement from the value chain. For the rest, it represents access that is easier, cheaper and more flexible, enhancing evidence-based policy formulation and implementation.

Ewan Sutherland is a research associate at the LINK Centre (University of the Witwatersrand) and at CRIDS (University of Namur). He is a Trustee of the IIC and Interim Editor for Intermedia. He can be followed on [Twitter](#) @sutherla

¹ Mike Nelson (2013) "The big picture on big data" *Intermedia* 41 (1) 12-16. See also Viktor Mayer-Schonberger & Kenneth Cukier (2013) *Big data: A revolution that will transform how we live, work and think*. Boston: Eamon Dolan/Houghton Mifflin Harcourt

² <http://fivethirtyeight.com>

³ <http://lod2.eu/>

⁴ <http://data.un.org/>

⁵ <http://www.google.com/publicdata/directory>

⁶ <http://www.broadbandmap.gov/developer>

⁷ <http://www.data.gov/details/1321>

Figure 1 Mobile telephone subscriptions per 100 population in Caribbean islands⁵

