**Royal Statistical Society StatsLife**

Liberating the potential of open data: a Q&A with Heather Savory

Written by [Oz Flanagan - Features Editor](http://www.statslife.org.uk/component/contact/contact/22) on 15 January 2014. Posted in [Features](http://www.statslife.org.uk/features)



The open data agenda is one of the most radical ideas to take hold in governments across the world and the UK has aspirations to lead the pack in the initiative. Its main champions, Nigel Shadbolt and Tim Berners-Lee, have been joined by a commitment to open up government data that began with the last Labour government and continues through the current coalition.

The release of datasets is having an exciting impact on research and innovation in various sectors. From government spending figures to Land Registry and companies data and the potential of addressing and other geospatial data. It’s an exciting time for data users who previously had to work around the lack of publically unobtainable data in their endeavours.

The government at national and local level holds so much data that it set-up the [Open Data User Group](http://data.gov.uk/odug) in 2012 to gather views on what data should be released as a priority. Charing the group is **Heather Savory** who we talked to about the potential of the movement to transform research efforts everywhere from academia to business.

**As head of the Open Data User Group, what does your role entail?**

In my role as Chair, I work in conjunction with the Cabinet Office, to appoint a group of volunteers in order to reach out to the wider data community and pull together advice derived from the views of all data users.

So around the table I have people from academia, local government, small business and big business. Our role is to provide a voice at the heart of government for all these data users.

We have articulated to government that there is a real demand for specific datasets for specific purposes, so now there is also is a [Data Request facility](http://www.data.gov.uk/odug) in place which acts as a demand led request mechanism.

Since September 2012 there have been over 690 requests and we have worked with the Cabinet Office to prioritise them. We set out to see which ones are requests for new data and which ones are flagging up an issue with the quality and format of existing data. We then look at which requests will have the most benefit to society, develop business cases then make the case to government as to why these should be released.

**What difficulties are there is releasing data in the right format for easy analysis?**

There has been a significant cultural change, in that releasing a PDF file should not be recognised as ‘open data’. The type of file that data should be released in is anything from CSV upwards; [formats are defined](http://data.gov.uk/blog/guidance-very-basic-standard-file-format-data) on our website for each dataset. So the first big hurdle has been to persuade public sector officials that a PDF is not the same as data. To be useful to the user community, data has to be downloadable and include individual components in a dataset. We don’t prescribe certain file formats but there is a lot of work going on with linked data, and looking forward that would be the gold standard for datasets.

We have also put a proposal in to the Public Sector Transparency Board to put in place an incentive scheme for local government to publish datasets in standard formats. One of our priorities at the moment is to encourage local government to release their data in standard formats so that it can be pulled together by third parties to give a national view of local issues.

**How does the group address the privacy concerns around all the individual data held by the public sector?**

Open data cannot include private information and this is one of the main hurdles to releasing data as it needs to be anonymised in some way. At the moment there is a big debate going on about what is the best way to anonymise information. This was raised in the [Shakespeare Review](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198752/13-744-shakespeare-review-of-public-sector-information.pdf) and everybody in the open data agenda is aware that personal information cannot be released.

However, the onus also has to be put on the users of data so that if datasets can be triangulated to identify individuals, it is their responsibility to ensure individuals cannot be singled out by their putting multiple datasets together and publishing the results.

**What are your favourite practical uses of open data that have resulted from the initiative?**

The one that I find personally the most useful is the bus tracking information, its second nature now to look on your smartphone and check if your bus is either two or four minutes away.

There is also some very interesting analysis being done on spend data in health, a company called Mastodon C did an analysis on the prescription of statins by GPs based on open data. They discovered that if GPs were to prescribe generics rather than branded products they would save £200million a year from the public purse.

There are many other datasets we are looking to release, but the one I think has a lot of potential is the VAT register. Following a data request, HMRC are now undertaking a consultation on releasing the register as open data which would have a fundamental impact on the ability of small business to get improved credit ratings and future funding. We estimate that this would deliver a benefit of £50bn a year to the economy, so this is a very good example of the potential of open data to help stimulate innovation and growth in the UK.