STRATEGIES FOR PUBLIC ENGAGEMENT WITH OFFICIAL STATISTICS

Jim Ridgway, James Nicholson, Sinclair Sutherland and Spencer Hedger
Durham University
jim.ridgway@durham.ac.uk

High quality data are available on a very wide range of topics. However, there are barriers to be overcome before citizens engage with these data. First is access; data are often distributed across different websites. Second is exposure; data need to be presented in the most commonly used information sources. Third is engagement – creating a climate where citizens are interested in evidence. The House of Commons Library (Library) and Durham University are collaborating in order to promote better use of data. The Library is the first port of call for politicians; their blogs and tweets are a major information resource for journalists. A website created ahead of the UK general election provides a rich resource for data at constituency level. Quizzes have been created which run on smartphones, in an effort to engage a broader section of people in evidence-based discussions. We discuss strategies designed to encourage public engagement with official statistics.

INTRODUCTION

There is a need for more and better use of official statistics in both public and private life. An important development has been the 'open data' initiative taken by governments and Non-Governmental Organisations, such as *data.gov* in the USA and *data.gov.uk* in the UK that explicitly state political objectives, in order to promote the democratic process by giving citizens access to data which can stimulate debate and inform policy making. High quality data are available on a very wide range of topics. However, there are barriers to be overcome before citizens engage with these data. First is access; data are often distributed across different websites, and on different spreadsheets within websites. 'Open access' requires expertise and considerable persistence. Second is exposure. Everyone has access to essentially unlimited information, and major providers are expert at saturating potential users with their own messages. The web is not a level playing field - high quality data, displayed effectively, will not be explored if it does not appear in the most commonly used information sources. Third is interest and engagement – there is a challenge to create a climate where citizens are actually interested in evidence. These (although obvious with hindsight) were hard-learned lessons from a notable failure - the creation of a website (http://www.smartcensus.org.uk/) with rich, interesting resources and exciting data visualisations (population pyramids, interactive travel maps and displays that facilitate interaction with multivariate data sets). The website has gone largely unnoticed, registering a few hundred hits per month. There was no real sense of user audience(s) or how to engage with them, and no obvious user needs that the resource was designed to satisfy.

Our overarching goal - to support a climate of evidence informed decision making, where 'what is the current state of affairs' and 'what do we know about successful interventions' are the first questions asked by policy makers and citizens - remains unchanged. However, we now have a different strategic approach, based on collaborating with groups and individuals who have influence with a variety of target groups.

THE COLLABORATION

We are working in partnership with a key data provider, namely the House of Commons Library (the Library), which is the first port of call for information by UK politicians. The May 2015 UK elections provide a focal point for the first phase of the collaboration. At election time, a range of audiences want access to accurate, relevant data. In the UK, 650 areas ('constituencies') are each represented by a member of parliament (MP). The political party (or coalition of parties) with a majority of MPs can form the government. Constituencies, then, are the political battlegrounds at election time; data visualization tools that facilitate engagement with local, regional and UK-wide issues are likely to be used by a wide range of groups. The first resource we have created is the Constituency Explorer. This resource has its genesis in programs developed by the Data Visualisation Centre at the Office for National Statistics – a group with whom we have collaborated on earlier projects (see Ridgway and Smith, 2013).

THE CONSTITUENCY EXPLORER

The Constituency Explorer http://www.constituencyexplorer.org.uk/ has three interlinked components; the Explorer data visualisations (EDV), the Quizzes, and a set of pdfs which provide extensive data about each constituency.

The Explorer Data Visualisations

The EDV comprise a suite of visualisations, many with a specific theme, such as Health, Demographics, Ethnicity, or the 2010 Election results, and some which are cross-themed. A 'cross-section' EDV has been created to illustrate the tool and the range of variables that can be explored. A screen shot is shown in Figure 1.
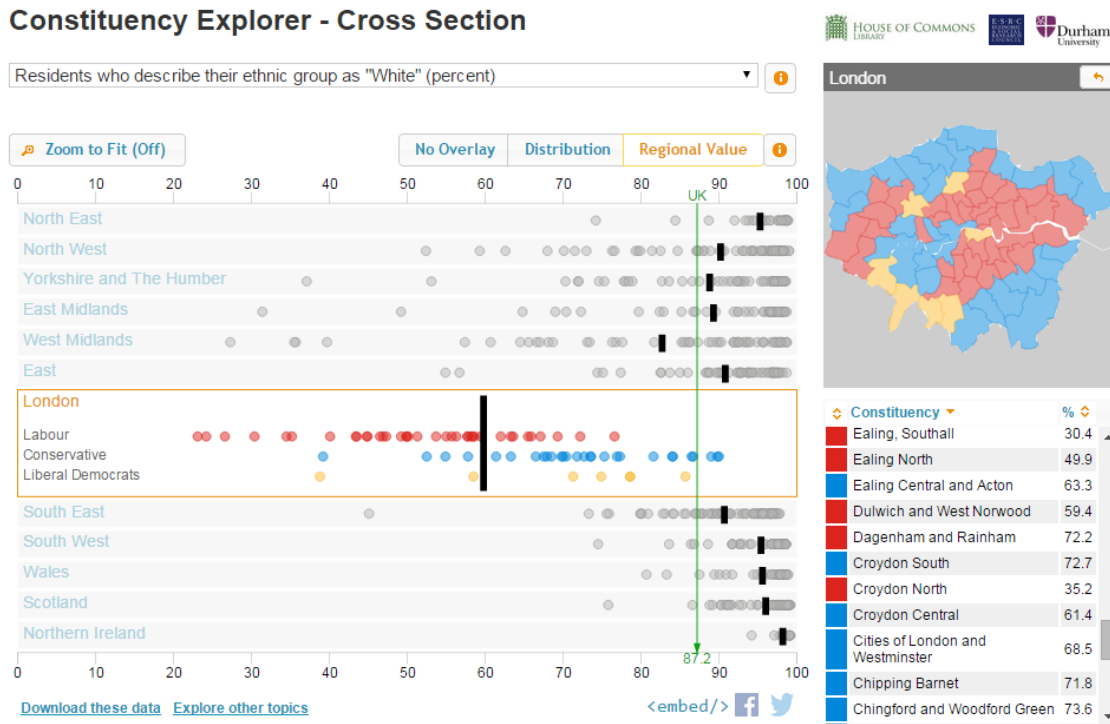

Figure 1: The Constituency Explorer – Country of Birth and Ethnicity

Each dot represents a constituency, and a scale is provided at the top of the display to allow users to compare constituencies and regions on a range of variables, chosen from a pull-down menu. Metadata are provided for each variable. If users click on a regional name (or on the map) the constituencies disaggregate by political party - in Figure 1, London has been selected. Users can examine regional differences (the black vertical lines show regional means), and can compare constituencies and regions with the UK average. It is easy to identify regional differences (Figure 1 shows large regional differences in the percentage of residents who described their ethnic group as "White" in the 2010 census; it is immediately obvious from the display that the ethnic composition of London is quite unlike the ethnic composition of the rest of the UK) and to identify those constituencies which are outliers on some variable. Individual constituencies can be highlighted via a pin, and constituencies can be compared with each other, on a number of variables. The same constituencies remain pinned, even though the variable of interest changes. This is true even if a different variable set is chosen within the EDV (for example, when switching from the Health EDV to the Employment EDV).

An important aspect of our strategy is to encourage key information providers to make use of this resource. There is a facility for any agency to embed the display in their own website by pasting a single line of code. The EDV itself is open source.

Data can be downloaded directly for all 650 constituencies as a csv file relevant to each group of variables, and for the entire data set. One of our principles is that data should never be allowed to escape from its own metadata, and so users download the metadata file at the same time as the data.

To encourage public engagement, there is a facility for users to post comments to social media sites (*twitter* and *facebook*). The header gives the name of the specific EDV, the variable name, a link to the page of interest, and space for commentary.

The Quiz

The Quiz is designed to create interest in information relevant to the election in the general public, and takes the form of a game that runs on smartphones and other mobile devices. It asks questions about individual constituencies, on topics such as ethnic composition, employment and health. A screen shot is shown in Figure 2.
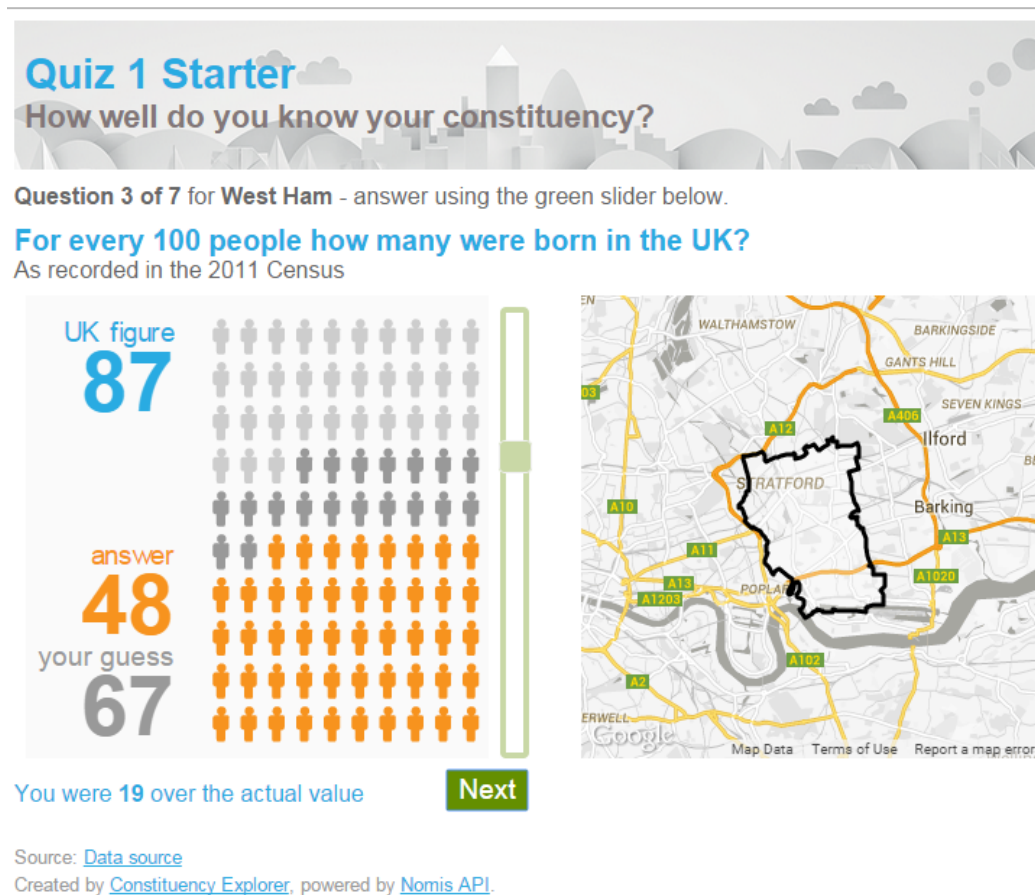

Figure 2: the Quiz

Links to social media are provided to encourage knowledge sharing by capitalizing on the pleasure people take in games, quizzes and competitions.

Users input their postcode (zip code) or the name of their constituency, or local town (part names are accepted) and are asked 7 questions. The correct answer is given (the colours of the icons change at a constant rate as the user's guess is changed to the correct answer – the worse the guess, the longer it takes), along with the figure for the whole of the UK. There are multiple versions of the Quiz, which address different themes.

A key feature is the link to the EDV. Once responses to the quiz have been scored, the quiz taker can link to the corresponding EDV where answers to each quiz question are available for *every* constituency, along with some associated information (for example, when the quiz asks about

the percentage of people in a constituency in good or very good health, the EDV also presents information about people in the constituency who reported bad or very bad health).

The constituency data pdfs
Users can download a pdf that presents a large collection of evidence about an individual constituency. More than 10,000 pages would be required to print the complete collection of pdfs.

USERS, USER NEEDS AND ENGAGEMENT
The Library itself is a key user group. They respond to tens of thousands of requests for information from politicians every year. People working in the Library will benefit from a single resource that provides ready access to extensive data sets, so that they can answer politicians' queries quickly, without having to find appropriate data sources, locate items of interest, and perform transformations and/or analyses. In the case of census data, to assemble UK-wide data from scratch, one must collate data from three different sources. The EDV also incorporates current data on labour market statistics. Comparing a specific constituency to another constituency, its region, or the UK as a whole, on a range of variables would be almost impossible without the EDV.

For politicians, the EDV provides key information, such as the results of the last election (and the size of the majority and electorate turn-out), and information relevant to voting, such as population demographics. It also has implications for policy (such as the large regional variations in health status, and the proportions of very young and very old people). Another use will be during election time; the EDV can provide a rapid tutorial about many aspects of an unfamiliar constituency for politicians visiting (or perhaps contesting) marginal seats.

Journalists are a third key audience. There are striking differences between constituencies and between regions (for example in ethnicity, identity, and median income), all of which can be the focus of articles. The EDV is well suited to TV because of its interactivity. For print media, screen shots are attractive, and the EDV can be embedded easily in websites.

An important group of users who are emerging are independent fact checking organisations such as Full Fact (fullfact.org). Such groups both critique statements in the media on an on-going basis, and provide specific services to media during (for example) live on-screen political debates. Large media groups (such as the BBC) are able do this for themselves but others cannot. The EDV will be a valuable resource to fact checkers.

Communicating with a range of disparate groups (politicians, TV, print media, bloggers, high-profile tweeters) can be difficult for academics; this is not the case for the Library. There are direct channels for communicating with politicians, via emails and internal video messaging.

Media people queue to attend media briefings from the House of Commons, and the Library can feed information and resources into weekly briefings. There is a House of Commons blog, and a twitter feed, with over 400,000 followers, and there are good personal contacts with key media figures.

The collaboration with the Library allows us to benefit from these communication channels. Somewhat paradoxically, these channels make it easier for us to communicate with academics in other fields, such as political science. We do have some experience of cross-disciplinary collaboration with teachers of geography, citizenship and social studies, and with political scientists (e.g. Nicholson, Ridgway and McCusker, 2012, 2013a, 2013b; Ridgway, Nicholson and McCusker, 2014). However, blogs from the Library are regularly reblogged on key websites which are major sources of information and communication for political scientists, notably *Election 2015* and *British Politics and Policy*, and so dwarf the impact that arises from personal communications.

The Constituency Explorer has been designed specifically to encourage user engagement and reuse. The quiz is designed to be adopted by popular newspapers. The EDV is designed so that it can be incorporated directly into others' websites.

The Constituency Explorer also has direct implications for statistics educators. It offers opportunities to engage with large scale, real data (over 150 variables synthesized from different sources) as advocated by the GAISE Report (ASA, 2005). It is designed to promote public engagement with evidence, and could be used to promote statistical literacy in class. Gal (2002, 2003) argued that statistically literate citizens should be able to evaluate statistical evidence

provided by government agencies and the media. The Explorer provides a rich resource, and students could be asked to deconstruct different ways of presenting data, and different media accounts in the light of Explorer data.

THE EXPLORER AS A RESEARCH TOOL

Google analytics provides information about levels of use (for example, on the first day 1500 different visitors spent more than 5 mins on average on the website). It also provides information about those topics and variables that attract most interest.

The Quiz provides information about public perceptions of politically sensitive data - such as the proportion of people born outside the UK, or the proportion of Muslims in a constituency. The samples will be biased, and not representative of the population as a whole, but are likely to be conservative - users are likely to be better informed than the populace as a whole.

Following up on media presentations, and related blog discussions is likely to be a rich source of ideas on statistical understanding.

IMPLICATIONS FOR PRACTICE

Ridgway and Smith (2013) argued the case for greater collaboration between academics and official statisticians, in order to improve statistical literacy in the population at large. The collaboration between the Data Visualisation Unit at the UK Office for National Statistics referred to in that paper continues - we share data visualisation concepts and code. Here, we see an extension to this idea - working with the Library (a key user group themselves) in order to co-create knowledge and resources with people who have a pivotal role in fostering statistical literacy. This approach offers a radical alternative to the passive creation and dissemination of resources.

We need to understand a great deal more about the dynamics of evidence dissemination and use. Better engagement (both in quantity and quality) is a prerequisite for higher levels of statistical literacy in the populace. The details and dynamics of engagement with different groups in a UK setting may well not generalise to other contexts; however, a number of principles *are* likely to transfer. Key principles to foster statistical literacy in the community at large include: collaborate with excellent, high-profile agencies who communicate directly with politicians; identify multiple possible user groups, and understand their needs and methods of obtaining evidence; develop resources which people know they need, in a form they can use; work with people who are passionate about valid uses of evidence (such as Full Fact); find channels to engage with opinion formers (such as journalists); create resources designed to be easy to use by target audiences, and engage potential partners from the earliest design stages; present the same data sets in different forms, to accommodate the needs of different groups; develop ways to engage a broad constituency of users (here, via a quiz) with direct links to richer data resources, in order to encourage more and deeper engagement.

CONCLUSION

Evidence-informed decision making depends on access to information. The paper reports an on-going collaboration between Durham University and the House of Commons Library designed to increase public engagement with official statistics. A number of principles are set out, based on this collaboration. Co-creating resources with providers of information who will use new tools in their own work is more likely to be successful than creating resources, then asking others to adopt them. Resources should be both powerful and easy to use, and should enable users to function more efficiently, and perhaps in new ways. User engagement is important throughout the whole design process. Key user groups need to be identified, and resources geared to their needs and interests. One should aim to exploit existing channels which engage opinion formers (such as journalists), and should work with people who are passionate about valid uses of evidence. Resources should be designed to make it easy for secondary users and information providers to embed resources into their own websites. Designing resources that work on smartphones and mobile devices, and using games to engage the general public, with links to richer data resources, will be far more successful that creating custom-built websites.

REFERENCES

Gal, I. (2002). Adults' statistical literacy: meanings, components, responsibilities. *International Statistical Review, 70*(1), 1-25.

Gal, I. (2003). Expanding conceptions of statistical literacy: an analysis of products from statistics agencies. *Statistics Education Research Journal, 2*(1), 3-21.

Nicholson, J., Ridgway, J., and McCusker, S. (2013a). Getting Real Statistics into all Curriculum Subject Areas: Can Technology Make this a Reality? *Innovations in Statistics Education*, 7(2). Pp12. http://escholarship.org/uc/item/7cz2w089

Nicholson, J., Ridgway, J., and McCusker, S. (2013b). Health, wealth and lifestyle choices - Provoking discussion on public spending. *Teaching Citizenship 36*, 23 – 27 https://www.dur.ac.uk/resources/smart.centre/Publications/Healthwealthandlifestylechoices-Provokingdiscussiononpublicspending.pdf

Nicholson, J., Ridgway, J., and McCusker, S. (2012). The summer riots 2011... lessons to be learnt? *Teaching Citizenship, 34*, 24 – 27. https://www.dur.ac.uk/resources/smart.centre/Publications/Thesummerriots2011_Lessonstobelearnt.pdf

Ridgway, J., Nicholson, J., and McCusker, S. (2014). Exploring 'white flight' via Open data and Big Data. In K. Makar, B. de Souse, and R. Gould (Eds.). Sustainability in statistics education. *Proceedings of the Ninth International Conference on Teaching Statistics* (ICOTS9, July, 2014), Flagsaff, Arizona, USA. Voorburg, The Netherlands: International Statistics Institute. Iase-web.org

Ridgway, J & Smith, A. (2013) Open Data, Official Statistics and Statistics Education – Threats, and Opportunities for Collaboration Keynote Talk: Proceedings of the first Joint International Association for Statistics Education and the International Association for Official Statistics, Statistics Education for Progress, Macau.

http://iase-web.org/documents/papers/sat2013/IASE_IAOS_2013_Paper_K3_Ridgway_Smith.pdf