THE NEXT GREAT LEAP – FROM OFFICIAL DATA TO PUBLIC KNOWLEDGE

Jim Ridgway, James Nicholson and Sean McCusker
Durham University, School of Education, Leazes Road, Durham, UK
Jim.Ridgway@Durham.ac.uk

ABSTRACT
Many statistical agencies are attempting to inform citizens about the state of society, and current social changes. The literature on statistical literacy suggests that this might be an uphill struggle. Here, we use a press release from Ireland's Central Statistical Office to show how important social trends that are masked in the press release can be made easier to understand via the use of new technologies. We advocate the use of mash-ups as the default form for press releases, and provide an example on the use of alcohol by young people.

INTRODUCTION

Many governments have embraced the concept of evidence informed policy, and many have expressed a desire to invigorate democracy by engaging citizens in debates around key issues of social policy, such as health. Organisations with a responsibility for gathering data relevant to policy increasingly are being asked by their funders to do more to ensure that the evidence they collect is put into the public domain, and is understood and used by appropriate communities.

A number of international agencies – notably The United Nations (UN) and OECD – are committed to programmes which set out to engage citizens across the world in: a radical reconceptualisation of the meaning of 'social progress'; in ways to measure social progress (e.g. The UN Millenium Development Goals); and in ways to hold governments accountable for promoting social progress in ways that do not have negative consequences for the environment, for long term sustainability, or for citizens of other countries. OECD's flagship project in this domain is their Global Project (GP). The goals are set out in the Istanbul Declaration (2007) – the scale of the movement can be judged by the impressive range of signatories. Full Partners in the Global Project (GP) include organizations such as the UN and The World Bank; Associate Partners (APs) include NGOs and some university research groups; the SMART Centre at Durham is one of these APs.

The challenges faced by organizations which create official statistics are clear. Most obvious to IASE members will be the poor state of statistical literacy in the adult population (e.g. Batanero *et. al* (1994)). There are other pernicious problems, such as the widespread mistrust of national governments and of national statistical offices (Giovannini, 2008), and the practical and technical issues associated with providing data in ways that make public access straightforward. These latter problems are important, but are beyond the scope of this paper. A number of statistics offices are actively engaged in promoting statistical literacy (e.g. Statistics Canada, Statistics Portugal, The Australian Bureau of Statistics, Statistics New Zealand, and the UK Office of National Statistics); a forthcoming special edition of the Statistical Journal of the International Association of Official Statistics (SJIAOS) will provide further examples. The International Statistical Literacy Project (e.g. Sanchez, 2008) is also doing excellent work promoting statistical literacy.

Work on statistical literacy has often focused on the problems of understanding seemingly simple statistical ideas, and in dealing with straightforward tasks such as interpreting tables and graphs (e.g. Schield, 2008). The problems faced by providers of official data are that most social phenomena are multivariate, and involve non-linear relationships between variables. There is usually a number of potentially confounding variables, and each variable may have an effect on the phenomenon of interest over only a particular range of values. It is reasonable to suppose that such complexity will prove to be an insurmountable barrier for any potential user who has no formal training in statistics. Recent work (e.g. Ridgway *et al* (2007)) has shown that some of the barriers to understanding complex evidence can be reduced and removed if data are presented in interactive multivariate displays.

The work of the SMART Centre has a number of foci, all associated with the idea of helping non-statisticians to understand large scale, authoritative, multivariate data. Activities include:

- Interface design, and the promotion of novel interfaces;
- Defining and describing 'new statistical literacies';
- Engagement with a variety of communities in official statistics and schools;
- Tutoring on OECD's GP courses for statisticians and policy makers.

Here, we discuss ways in which official data can be made more accessible to a general audience. In order to make the discussion more concrete, we base the paper on an activity developed for the first OECD course for statisticians and policy makers that took place in Siena (2008), designed to provide professional development to underpin the GP. The task presented to the delegates was to rewrite a press release (PR) from a national statistics office, using an interactive display we had created using official data.

A PRESSING NEED FOR BETTER PRESS RELEASES

Figure 1 shows a PR from the Irish Central Statistics Office (CSO). The PR relates to a full report on new statistical data relating to societal progress in Ireland, which is available at http://www.cso.ie/releasespublications/measuringirelandsprogress2007.htm. The obvious purpose of such a PR is to provide a succinct summary of key messages in the full report in order that journalists can write timely and accurate accounts that do not depend on a great deal of reading, or on extensive statistical knowledge.

Longer Life Expectancy for Men and Women

*The report **Measuring Ireland's Progress, 2007**, published by the CSO today, shows the progress made in Ireland in important economic, social and environmental areas. As well as showing developments over time, the report benchmarks the situation in Ireland against the other EU Member States.*

Key findings of the report include:

- *Life expectancy at birth was provisionally estimated at 81.5 years for Irish women and 76.7 years for Irish men in the period 2004-2006. In comparison with 2001-2003, men's life expectancy increased by 1.6 years and women's by 1.2 years, reducing the gap between men and women to 4.8 years in 2004-2006, the lowest it has been since the 1970-1972 period (Table 6.3).*
- *In 2007, 41.3% of the population aged 25-34 had completed 3rd level education. This was the second highest rate across the EU and well above the EU 27 average of 29.1% (Table 5.7). Irish students aged 15 years had the second highest levels of reading literacy in 2006 (Table 5.9).*
- *In 2006, Ireland had the second highest GDP per capita in the EU 27 at 45.4% above the EU average. However, based on GNI, Ireland was the fifth highest at 25.2% above the EU 27 average (Table 1.3).*
- *The average value of a new housing loan in Ireland rose from €62,000 in 1997 to €229,200 in 2006. Mortgage interest rates declined over this period from 7.22% to 4.2% while the number of loans taken out for housing increased from 57,901 to 111,253 (Table 8.3).*
- *The percentage of waste landfilled in Ireland decreased from 67.3% in 2004 to 63.9% in 2006. Glass and paper were the materials most likely to be recycled with 63.5% of glass waste and 55.3% of paper waste recycled in 2006 (Table 10.8).*

- *An average of €2,673 (at constant 2005 prices) per person was spent on non-capital public expenditure on health care in Ireland in 2006. This represented an increase of over 72% on the 1997 level (Table 6.1).*
- *Ireland's net official development assistance increased to 0.54% of GNI in 2006 from 0.42% in 2005. This was below the UN 2007 target of 0.7% of GNI (Table 4.12).*
- *Over the period 1999-2007, the euro increased in value against the dollar by almost 29% and by 3.8% against sterling (Table 1.15).*
- *The employment rate in Ireland rose from 59.7% in 1998 to 69% in 2007. The rate for women increased by over 12 percentage points over that period, while the rate for men rose by over 6 percentage points (Table 3.1). In 2006, Ireland had the eighth highest employment rate in the EU 27 (see Table 3.2).*
- *Productivity in Ireland, measured as GDP per person employed, was the second highest in the EU 27 in 2006 (Graph 3.4).*
- *Ireland had the sixth lowest unemployment rate in the EU in 2007 at less than two-thirds of the EU 27 average of 7.1% (Table 3.6).*
- *6.9% of persons in Ireland were in consistent poverty in 2006 (Table 4.6). 22.8% of unemployed persons were in consistent poverty (Graph 4.7).*
- *The proportion of Irish people at risk of poverty, after pensions and social transfer payments were taken into account, was 18% in 2006. This was above the EU 25 average of 16%. The effect of pension transfers on reducing the at-risk-of-poverty rate was low in Ireland compared with other EU 27 countries (Table 4.4).*
- *The pupil-teacher ratio at primary level in Ireland in the school year 2004/2005 was one of the highest in the EU 27 at 17.9. Eleven of the reporting EU member states had a pupil-teacher ratio of less than 13 at primary level (Table 5.4).*
- *Early school leavers represented 12.3% of the 18-24 age group in Ireland in 2006 (Table 5.12). The unemployment rate for early school leavers in this age group was 23.4% in 2007 compared with an unemployment rate of 8.4% for all persons aged 18-24 (Table 5.10).*
- *The population in Ireland increased by 17.2% to almost 4.34 million persons in the period 1998-2007 (Table 7.1). This was the highest rate of increase in the EU 27 (Graph 7.3). The rate of natural increase of the population in Ireland was 8.7 per 1,000 in 2006 compared to an EU 27 average of just 1.1 (Table 7.6).*
- *Ireland's greenhouse gas emissions were at 125.5% of 1990 levels in 2006. This was 12.5 percentage points higher than the Kyoto 2008-2012 target for Ireland of 113% of 1990 levels (Graph 10.1).*

Figure 1:Press release from the Irish Central Statistics Office.

This PR has a number of obvious defects. The structure is poor. Information relevant to education appears in the second, fourteenth and fifteenth bullet points. Information relevant to the environment appears in bullet points 5 and 17. Bullet point 4 compares data from 1997 and 2006; bullet 5 compares data from 2004 and 2006; bullet 8 compares data from 1999 and 2007, and so on - making it very difficult for the reader to draw sensible conclusions.

However, the most significant flaw is the inclusion of findings of minor importance, and the failure to report findings of major importance. Here, we explore one data set, in order to draw some important conclusions.

> *Life expectancy at birth was provisionally estimated at 81.5 years for Irish women and 76.7 years for Irish men in the period 2004-2006. In comparison with 2001-2003, men's life expectancy increased by 1.6 years and women's by 1.2 years, reducing the gap between men and women to 4.8 years in 2004-2006, the lowest it has been since the 1970-1972 period (Table 6.3).*

Figure 2: statements in press release relating to life expectancy

Current estimates of the life expectancy of males and females provided here, in figure 2, are both interesting and important. Figure 3 shows the life expectancy at birth over a 26 year period for men and women. Figure 4 shows analogous data for men and women aged 65 years. These data have been downloaded from the CSO website and are presented in a SMART Centre interface (the website provides a facility for anyone to upload muiltivariate data sets into an interactive display).

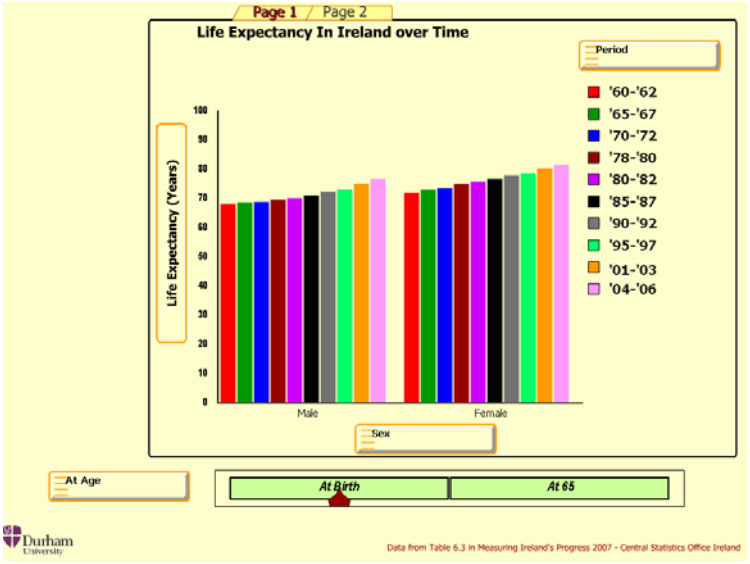

Figure 3: Life expectancy from birth for males and females over time in Ireland

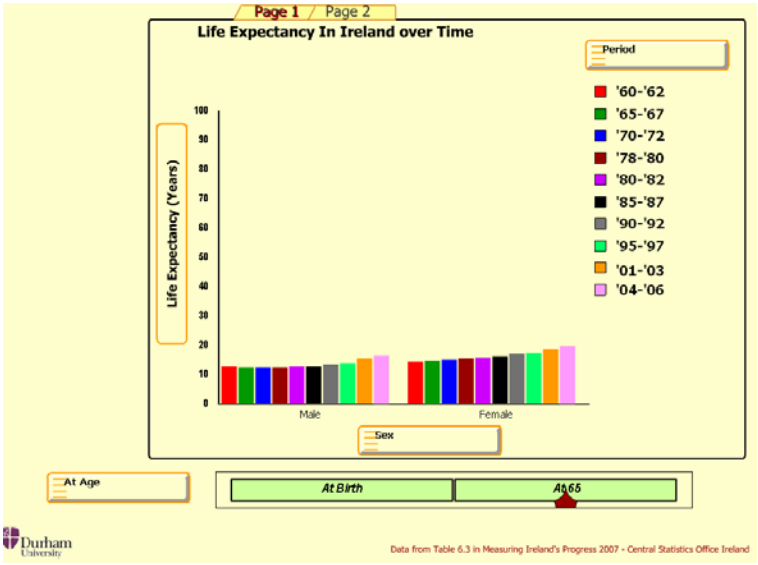

Figure 4: Life expectancy at age 65 years for males and females over time in Ireland

The choice in the PR to compare 2001-03 with 2004-06 is somewhat arbitrary. The most important feature of the data is the steady increase in life expectancy over a long period for both males and females. One might ask questions about the exact shape of the function (linear or other?), but the key message is a monotonic increase in life expectancy over time. This trend is important both as a summary of the data, and as a harbinger of demographic changes, with their associated impact on likely demands on health, leisure and social services.

Figure 5 shows the difference in life expectancy between men and women over the same period.

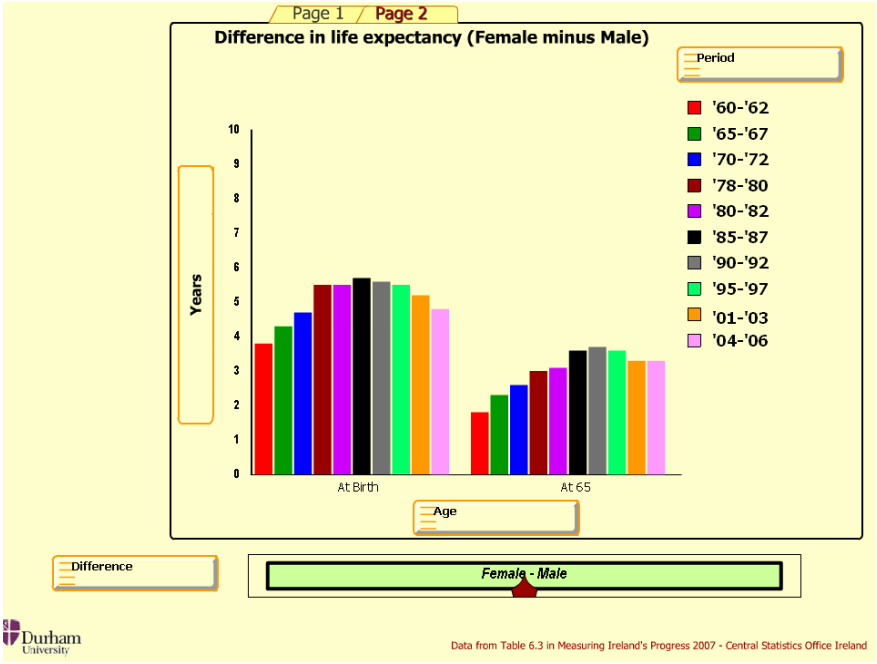

Figure 5: Differences in life expectancy for males and females at birth and at age 65 years over time in Ireland

Again, the most interesting feature of the data is the trend over time. As before, the decision to report a comparison with a single earlier reference point

> *…reducing the gap between men and women to 4.8 years in 2004-2006, the lowest it has been since the 1970-1972 period (Table 6.3).*

misses the most interesting features of the data. The inverted U-shape does invite conjectures and explanations.

In the absence of evidence, it is interesting to speculate about the relative difficulty of interpreting graphical displays of the complete set of raw data, and text based on numbers. The two key salient features of the data here are a linear (or approximately linear) increase in life expectancy over time, and an interesting inverted U-shaped function in the differences in life expectancy between men and women over time. These (we assert) are immediately evident in the graphical displays. A description in words that can be understood by naïve audiences is probably very difficult to write; there is literature (e.g. Schield, (2006)) on the problems that adults have working with tabular data, rates, and raw numbers.

RELEASING STATISTICS FROM STATICS

Our criticisms of the CSO PR can be categorised in two ways. The first set of problems relate to improvements that could be made to structure and content. If one accepts the constraints imposed by a format of short, paper-based PRs, then the resolutions of the problems identified here would be based on improving structure, and on reporting trends over time, rather than differences in point values. The second set of problems relates to user engagement. These problems are inherent in the medium, and are not specific to this PR. It would be quite

inappropriate to create a PR that contained graphs in sufficient quantity to provide adequate illustrations of all the data in the National Report. However, web access is ubiquitous in developed countries, and it is easy to give users (that include interested citizens as well as journalists) access to interactive displays.

Interactive displays should become a major feature both of the websites of data providers, and of PRs. We advocate the creation of PR in the form of mash-ups – a mixture of text and interactive display that allow users to explore data, test assertions, describe phenomena, and make conjectures of their own.

We have been exploring the use of mash-ups in helping people understand complex relationships in a variety of contexts, such as vulnerability to strokes, sexually transmitted diseases, and the use of alcohol by young people. Figure 6 shows a display we have used to engage students aged 13 -16 years in discussions about alcohol use by young people.

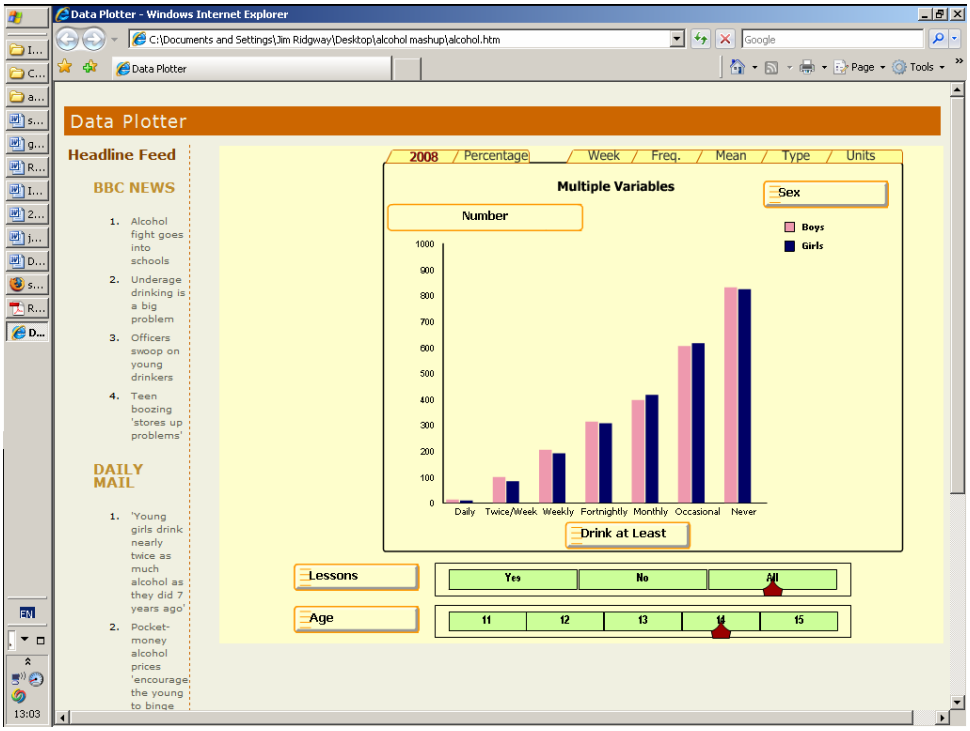

Figure 6: A Mash-up on Alcohol Consumption

The salient features of the mash-up for this discussion are the presentation of a large number of interactive displays that use data from a variety of sources, and the direct links to newspaper accounts. Each of the tabs presents an interactive display on a different aspect of the evidence relevant to the consumption of alcohol by young people, such as: the association between lessons on alcohol and actual reported alcohol consumption (shown here for students aged 14 years); and self reports of weekly consumption, frequency of drinking, type of alcohol consumed, and the number of units of alcohol consumed. The text is all taken from newspaper accounts.

We believe that statistics agencies should make extensive use of such displays in their PRs (and indeed, that they should regard this format as the default format, and 'paper-only' PRs as a decadent form with a rather restricted set of uses). In the context of the CSO PR, tabs could relate to different features of the National Report – life expectancy, education, GDP, health, environment and the like. The associated text can be pitched at a variety of 'grain sizes' – from descriptions that are easy to cut and paste into newspaper articles, through more discursive accounts on a few pages, to detailed statistical accounts that give full details of the metadata, as well as comprehensive analyses.

CONCLUDING REMARKS

If evidence is to reach a large public audience, data displays need to be created in a way that makes it easy for TV and internet media to incorporate them into their existing websites, and into a form suitable for direct TV broadcast. This approach to getting important messages in the public eye by paying attention to the channels of communication that actually work on a large scale, requires a sea change in the positioning of (some) statistics agencies. It represents a shift from activities geared to dissemination to activities geared at user engagement. This shift puts pressure on providers to understand a good deal about both the competencies of users, and to develop effective communication methods. Communication involves far more than dissemination, and requires active engagement with users.

We believe that these developments are essential if agencies concerned with official data are to move from a focus on 'statistics' (the analysis of data about the state) to Giovannini's (2008) conception of 'societistics' – the development of information systems to further progress of societies.

REFERENCES

Batanero, C., Godino, J.D., Vallecillos, A., Green, D., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematics, Education, Science and Technology*, 25(4), 527 – 547.

Giovannini, E. (2008). The Role of Communication in Transforming Statistics into Knowledge. Proceedings of the OECD Seminar on Innovative Approaches to turning Statistics into Knowledge. Stocholm, Sweden, 2008. http://www.oecd.org/dataoecd/30/32/40293021.pdf

International Statistical Literacy Project (ISLP). http://www.stat.auckland.ac.nz/~iase/islp/

Irish Central Statistics Office (2008) http://www.csoireland.eu/newsevents/pressrelease_measuringirelandsprogress2007.htm downloaded Sept 2008]

Irish Central Statistics Office (2008). Measuring Ireland's Progress. http://www.cso.ie/releasespublications/measuringirelandsprogress2007.htm.

Istanbul Declaration (2007). http://www.oecd.org/dataoecd/0/24/39070305.pdf

Millenium Development Goals. http://www.undp.org/mdg/

National Support for Statistical Literacy, Australian Bureau of Statistics http://www.abs.gov.au/websitedbs/cashome.nsf/Home/Entry%20Page.es

New Zealand's School corner, http://www.stats.govt.nz/schools-corner/default.htm

Statistics Canada http://www.statcan.gc.ca/edu/index-eng.htm

UK http://www.censusatschool.ntu.ac.uk/default.asp

Statistics Portugal http://alea.ine.pt/english/

Ridgway, J., Nicholson, J., and McCusker, S. (2007). Reasoning with Multivariate Evidence. *International Electronic Journal of Mathematics Education* 2(3), pp245-269.

Sanchez, J. (2008). Using Statistics to make Sense of Data and Summarized Information: the Role of NSOs. *Government Statistical Offices and Statistical Literacy*. ISLP http://www.stat.auckland.ac.nz/~iase/islp/stats-offices-book

Schield, M. (2006). Statistical literacy survey analysis: Reading graphs and tables of rates and percentages. In A. Rossman, & B. Chance (Eds.), Proceedings of the *Seventh International Conference on Teaching Statistics, Salvador, Brazil*: International Statistical Institute and International Association for Statistical Education. Online:http://www.stat.auckland.ac.nz/~iase.

Schield, M. (2008). Quantitative Literacy and School Mathematics: Percentages and Fractions, *Calculation vs. Context: Quantitative Literacy And Its Implications for Teacher Education*. Edited by Bernard L. Madison and Lynn Arthur Steen. 2008 Mathematical Association of America. P. 87-107. See www.StatLit.org/pdf/2008SchieldWingspread.pdf.

SMART Centre http://www.durham.ac.uk/smart.centre

Statistical Journal of the International Association of Official Statistics http://isi.cbs.nl/iaos/journal.htm