Statistics Commission

Report No. 34
Data on Demand – Access to Official Statistics

June 2007

Statistics Commission

Report No. 34

Data on Demand – Access to Official Statistics

Report by the Statistics Commission

June 2007

Statistics Commission
Artillery House
11-19 Artillery Row
London SW1P 1RT
020 7273 8008
www.statscom.org.uk

Contents

Foreword

By the Chairman of the Statistics Commission

In our preceding 33 reports we have repeatedly made the point that good quality and trusted official statistics underpin the activities of the state and the effective functioning of democracy. Yet there is little point in having such statistics if they are hard to find. The revolution in access to information wrought since 1991 by the invention of the World Wide Web built upon the Internet has transformed the way in which hundreds of millions of individuals, families, organisations and governments across the world now function – the first port of call for information is to search the Web.

This development – the so-called 'democratisation of knowledge' – presents both great opportunities and challenges. In principle, anyone can find almost anything and act upon it. On the other hand, some information and some information sources are much more reliable than others. But all of this is secondary if access to the information is constrained by poor policies on the part of the information providers, poor structuring or inadequate descriptions of the information, ineffective search and assessment tools or poor presentation on the Web.

In this report we consider the improvement of access to official statistics in the UK. We conclude that there is still much improvement to be made and make recommendations in the form of principles which we urge the future Statistics Board to adopt. Behind it all however is a vision of government statisticians acting not only as experts in the collection and guardianship of statistical data but also as promoters of the use of these statistics to a wide community, not just central government (important as that sector is). To achieve this vision and maximise the public benefit from the costly investment in information, while continuing to keep statistics free of charge, will require some changes in attitudes and behaviour and possibly some reallocation of resources.

I would like to thank all those who contributed to this report, particularly Commission member Ian Beesley who chaired the project board, Richard Cracknell from the House of Commons Library, and Kevin McHale from the Office for National Statistics (project board external members), Abigail Armstrong and Allen Ritchie from the Statistics Commission Secretariat who managed the project, and the Ipsos MORI team who undertook much of the research.

Chairman, Statistics Commission
June 2007

Introduction and summary

1. In this report we look at the ease with which both experienced researchers and novices can find and use UK official statistics through the Web. User expectations are changing as the Internet develops and as citizens become more adept at using the vast range of information it offers. Search engines are enabling more people to have more access to more data more often and more efficiently.

2. The report presents the results from two research projects that we commissioned – "mystery shopper" research into finding statistics on topical issues and a survey of the policies of government departments for publishing of statistics on the Web. The full research report is attached at Annex 2.

3. Our recommendations are designed to help meet rising user expectations and to keep UK official statistics in the vanguard of public use of high quality statistics. They are addressed to the new Statistics Board that will be established by the Statistics and Registration Service Bill currently going through Parliament. They are in the form of a set of principles, which we believe should determine dissemination of statistics on the Internet.

4. To provide further context and detail about accessibility, relevant extracts from the existing National Statistics Code of Practice and a review of the literature on accessibility are attached as Annexes 1 and 3.

SECTION 1: RECOMMENDATIONS

5. We propose the following eight principles of statistical dissemination, to be followed by all producers of official statistics:

   Principle 1: **Statistics are collected to be used and as wide a use of them as is possible should be encouraged**, including the re-use of raw data for research outside government.

   Principle 2: Since the most satisfactory forms of data provision are still evolving, **UK government statisticians should adopt an exploratory and experimental approach to dissemination and access to statistical data through the Internet**.

   Principle 3: **Government departments that publish official statistics should seek the full involvement of other web professionals in the presentation of statistical data on their websites**.

Principle 4: **Government departments that publish official statistics should recognise that web design and web culture are still developing and should set up an appropriate mechanism to keep accessibility issues under review**. This might take the form of a technical advisory panel, with membership drawn from external user expertise such as resides in user-centred website design, communications experts, the Statistics User Forum (SUF) and the House of Commons Library.

Principle 5: **User needs, interests and capabilities should determine the design and operation of statistical dissemination over the Internet**. This necessitates interactive engagement with users and active pursuit of feedback plus better search engines (possibly in co-operation with commercial operators). Users of government statistical websites should not be expected to have a working knowledge of government departments or of 'who produces what'.

Principle 6: **Statistical products should be specifically designed for the Web**. The Internet brings a new dimension to the relationship between producers and users of statistics. It has transformed the way that people look for information and want to access information.

Principle 7: **Data should be presented in a layered or hierarchical way to allow users to drill down to the level of detail they desire**. Tables, charts, maps and online statistical programs for manipulating data should be increasingly the norm. Guidance on definitions, sources and methods, data quality and interpretation should be an integral part of dissemination (possibly in the form of Frequently Asked Questions).

Principle 8: **There should be one point of entry – a government statistics portal – giving access to official statistics across the UK government and those of the devolved authorities**. This portal would link to all government statistics sites, and include a search engine which operates seamlessly over all those sites.

SECTION 2: BACKGROUND TO THE REPORT

The UK has long had a healthy appetite for statistics

6. The UK appetite for statistical information is large. Many news stories are built around official statistics; examples from the headlines of the past few months include the numbers of immigrants, the inflation rate and the extent of obesity in the country. The Government's emphasis on evidence-based policy and performance targets increases both its own demand for statistical information and at the same that of the media and general public as they seek to use official data to assess government effectiveness and performance.

7. Historically there have been many examples of statistical evidence radically changing the way things are done. For example, Florence Nightingale's application of disease statistics to reduce mortality in those wounded during the Crimean war or Sir Richard Doll's research in the 1950's that made the link between smoking and lung cancer. Keynesian demand management of the economy would not have been possible without the development of National Accounts statistics. More recently the *Stern Review on the Economics of Climate Change* drew on statistical evidence.

8. Parliament and democratic accountability rely on good statistics. Many Members of Parliament have a keen appetite for statistics. Many of the Parliamentary Questions they put to ministers ask for statistical information; the House of Commons Library receives an average of 3,000 substantive statistical enquiries from MPs every year.

9. Official statistics are produced by central government departments, devolved administrations, local authorities, agencies or other organisations in the public sector. Most economic and many social statistics, including statistics on population, are produced by the Office for National Statistics (ONS). But statistics in other important areas are often produced by the government department responsible for policy in that area, eg most health statistics relating to England are produced by the Department of Health and education statistics for England by the Department for Education and Skills. Many statistics relating to Scotland, Wales and Northern Ireland are produced by the devolved administrations.

10. The UK has a tradition of good quality official statistics. A recent Statistics Commission[1] report concluded that, on the basis of research into public perception of statistics, "on the whole, opinion leaders believe that the quality of official statistics in the UK is up with the best in the world."

[1] *Official Statistics: Perceptions and Trust,* Report No 24, Statistics Commission (2005)

11. There is also a long-standing tradition of official statistics publications. The first Population Census was conducted in 1801. Labour market statistics have been published in the *Labour Department Gazette* and successor publications since1893, the National Accounts *Blue Book* first appeared in 1948, *Economic Trends* in 1953 and *Social Trends* in 1970. This long tradition of publication was reinforced by the National Statistics Code of Practice,[2] introduced in 2002, which included a commitment to free access to all official statistics through the Internet.

12. Nevertheless publication is not in itself enough – statistical information needs to be published in a way that makes it readily accessible and understandable to those who want to use it or who could use it if barriers to its use were removed. The accessibility of official statistics is a key issue for many actual and potential users outside government – the general public, the media, business and academics. This is recognised in the National Statistics Code of Practice, which requires that "…data will be presented to a standard that clearly and accurately expresses the contents to the widest possible audience, with choice and flexibility in the format where possible...". The Protocol on Data Presentation, Dissemination and Pricing states that "the Web will be the primary means of providing general access". We interpret the current Code of Practice and Protocols as requiring producers to make official statistics available free of charge on the Internet to the extent that it is practicable to do so.

13. However the Internet is a rather different medium for the dissemination of statistics from the traditional paper publication. The issues around accessibility of statistics through the Internet are also different. It is to this that we now turn.

The Internet changes everything

14. Computing power, enhanced storage and telecommunications have transformed statistical processing and computation, making it possible for a laptop to carry out computations that in the past would have required a roomful of machinery. Computing power per unit cost increased by a factor of 10,000 between 1975 and 1995[3] and the growth continues. Simulation and other data exploration tools are now readily available to the domestic laptop owner.

[2] 'National Statistics' is a quality marker applied to certain of the United Kingdom's official statistics. The National Statistics Code of Practice sets out the key principles and standards which official statisticians are expected to follow and uphold. It is supported by twelve Protocols. See Annex 1 to this report for relevant extracts.

[3] Bond, J. *The Drivers of the Information Revolution – Cost, Computing Power and Convergence*, Note No. 118 *Public Policy for the Private Sector*, The World Bank Group. July 1997.

15. The development of the Internet has brought changes in the way that people access and use information – not the least being an increasing familiarity with searching methods. The rate of households acquiring Internet access has been rapid – in 2006, 57 per cent had an Internet connection, up from 36 per cent in 2001.[4] Currently four out of five Internet connections are via broadband.[5] The Web is now the first port of call for information in many homes, businesses, schools, universities and government departments.

16. There has been an increase in information sharing websites. The Cabinet Office recently hailed this development as "democratising information and driving citizens' appetite for sharing advice and opinions in new ways"[6] and it is looking into how best to use them to improve information from government.[7]

17. Web users have the opportunity to select Internet content for their own needs, tailoring the format and content or sharing information with others about individual websites through the use of social bookmarking websites (ie sites that enable users to share online their favourite sites and to develop virtual information networks). Some website owners, such as newspapers, offer readers the opportunity to propose the page they are reading for inclusion in a hosted social bookmarking site, thus increasing readership of that newspaper's site. News feeds can be selected by topic so that the user in effect gets his or her own newspaper built only of matters of interest to them. All of this has raised expectations about the ability of websites to make content relevant to individual users and so facilitate finding information.

18. Rosling[8] has exploited technological developments to present information and enable new comparisons of statistics. He has argued for using technology more efficiently to enable users to access data, for example by adding a 'data' option to a search engine toolbar; Google has now bought his project.[9]

[4] The 2001 percentage of households with Internet connection covers the period Jan-March 2001 – source: *Historical Internet Access – Data (pre-April 2003),* ONS, (accessed April 2007); the 2006 figure covers to Jan-April 2006 – source: *Internet access – households and individuals*, Statistical First Release, ONS, (August 2006).

[5] The rate of broadband connections relates to March 2007 and is taken from *Internet connectivity March 2007*, Statistical First Release, ONS, (May 2007).

[6] *Power to the people – Information sharing hailed as a new force for social progress by Government*, Press release 8 February 2007, Cabinet Office.

[7] *The Power of Information Review: online advice sites could improve citizen empowerment,* Press release 5 April 2007, Cabinet Office.

[8] *Debunking third-world myths with the best stats you've ever seen*, Rosling H, conference presentation to Technology, Entertainment, Design (TED), (2006).

[9] *Business comment: Statistics website will bring openness, in all probability*, Conway E, The Telegraph, 6 April 2007.

19. The Internet also opens up many possibilities for interaction between the consumer and the provider. This has been highlighted by the Varney report on the opportunities for transforming the delivery of public services. It noted that "today's consumer is no longer the passive recipient of government services" and that technological changes "create an increasing expectation that access to services on the Web will be comprehensive, joined-up and capable of delivering a service almost instantly".[10]

20. In line with these trends, the Government recently announced its intention to make all online government services accessible via a central hub, such as DirectGov or Business Links[11]. Another recent proposal has been the development of a single statistical publication hub that would separate the release of statistics from political comment on the data.[12]

21. The Freedom of Information Act (2000) is also helping to generate a culture of making government information publicly available, and is raising expectations about the type of information that should be made available.

22. At the same time, the Internet raises difficulties about the authority of information. These are less acute when the information is from a trusted source but much more serious when it is produced by individuals or (sometimes) by a group of volunteers. Whilst blogs (web logs) encourage access to a wider source of ideas and comment, they also increase the risks of misinformation. Blogs are used both by the authoritative (for example, the ONS blog as part of its consultation on geographies[13]) and the merely opinionated. The online self-managing encyclopaedia Wikipedia is now much bigger than, for example, Encyclopaedia Britannica, although the quality of Wikipedia has been questioned. The Internet offers unlimited access to information posted on it – much of it accurate, some of it dishonest, much misleading. Hence the need for trusted sources, tools for 'triangulation' of information to check accuracy, good codes of practice for information suppliers and explicit descriptions of information quality have seldom, if ever, been stronger.

How we search for and scan information

23. The Internet has also transformed how we search for information; Google and the other leading search engines have democratised access to data and set a high standard for search software. Although users do not necessarily understand how the search engines work, they expect to find what they are

[10] *Service Transformation*, Sir David Varney, HM Treasury, Dec 2006.

[11] *Transformational Government Annual Report 2006*, Chief Information Officer Council, January 2007,

[12] The publication hub was announced during the second reading debate of the Statistics and Registration Service Bill in the House of Commons.

[13] The blog for ONS' small area geography consultation (now closed) can be found at: http://www.onsgeography.net/

looking for on the first page of search results. If their first attempt is unsuccessful, they typically refine the search terms in an iterative process rather than switching to another search engine[14] or using advanced search tools.[15]

24. The wide availability of information on the Web and the ease of accessing it via search engines have led to what has been called "information snacking".[16] This refers to the short time that people are prepared to spend scanning a website or web page. It is increasingly rare for people to scan an entire document other than by using search facilities. The recommended solution is to provide information in easy to read, bite-sized chunks, with the facility to drill down for more detail or explanation as appropriate.

25. Arguably, sophisticated search engines have made finding information seem almost too easy. Getting hold of statistical data is not the same thing as understanding what they mean. There is an "entry fee" to pay to understand and interpret what the data represent and to avoid being inadvertently misled by headings or complex descriptions. Historically, subject matter experts have often invested considerable time in understanding data – but it is unrealistic to expect most Internet users to do so. Other means of guidance must be used about definitions, sources and methods, the meaning of the figures and their limitations.

26. A 2005 study of the US Government website FedStats[17] (the American portal for official statistics) argued in support of the provision of information about statistics and a statistical commentary. Its authors argued that there should be "no naked data" – the story behind the numbers is always needed. They also recommended presenting such information in small amounts – with "just-in-time, just-enough" help that is presented when needed. They suggested this might be provided by an interactive statistical glossary with context-specific explanation.

What constitutes good 'data accessibility'? The importance of website design

27. The same study of the FedStats site identified website design as a possible barrier to accessibility and recommended that website designers aim to "minimize scrolling and clicking, provide alternative ways to slice and dice datasets, [and] closely couple search, browse, and examine functions...".

[14] *Search Engine User Behaviour Study*, iProspect and Jupiter Research, iProspect, (2006).

[15] McGovern, G in *The search lurch*, Buchholz GA, (2005), Prentice Hall PTR

[16] Nielsen, J in *The search lurch*, Buchholz GA, (2005), Prentice Hall PTR

[17] *Finding and understanding government statistical information*, Marchionini G, Haas SW, Zhang J & Elsas J (2005). Computer, 38(12) p52-61, Dec 2005.

As other research[18] has established, behaviour in looking for statistics is no different than when searching for other information – so the problems that emerge, such as with word matching or with documents in portable document format (pdf), are generally similar.

28. An earlier study[19] of website accessibility, also based on FedStats, had recommended designing a statistical website for novice users. Statistics producers and website managers should not assume that all users regularly accessed the site and, as a consequence, had learned how to use it. Use of "scientific" words in presentation of data was best avoided; easy navigation of the site should be provided; users should not be presumed to know the structure of organisations or agencies; there should be a facility for users to perform a 'comparative search', enabling, for example, comparison of statistics for two cities; advanced search facilities should be available; tools for analysing the data online should be provided; and users should have the option to choose the granularity of geography and time series.

29. All the studies so far mentioned have had a US focus. A rather different perspective is provided by 'Surfing with Ed',[20] a monthly series of reviews by a UK official statistician that has looked at a different country's statistical website each month for a number of years. Websites that have been reported as standing out include Iceland and Australia for their good navigation; Japan and Norway for their user orientation; Italy and Australia for their features; and Holland for its good practice in displaying metadata (data about data).

How are the providers of UK official statistics responding?

30. The combination of vastly increased data availability, easier to use technology, rising education and ability to use statistical information plus reducing deference to authority throws down a gauntlet to producers of official statistics. The official response of those responsible for UK government statistics to the advent of the Internet has been the current policy of making the Web the focus of dissemination of official statistics, with all statistics available freely online. Thus in theory it should be much easier than before for the general public and the non-expert user to access statistical information. There is no longer a need to go to a library to consult a paper publication (or to purchase one), or to subscribe to (paper) press releases.

[18] *What is the future of statistical compendia in the 21st century?* Zawitz MW, Statistical Journal of the UN Economic Commission for Europe; 2005, Vol. 22 Issue 2, p163-171.

[19] *Finding Governmental* Statistical *Data On The Web: A Case Study Of FedStats*, Ceaparu I, IT&Society, Volume 1, Issue 3, Winter 2003, Pp. 1-17.

[20] *Surfing with Ed on the Internet*, Swires-Hennessy E, monthly articles 1998-2006 and Edition 100, December 2006.

31. This is an important step. But to maximise the public benefits from the changes brought about by the Internet it is necessary to go further. In some cases, producers of official statistics have done so, and provided users with the opportunity to tailor official statistics or to produce customised tables online (for example, with benefits data on the Department for Work and Pensions website or the personalised mapping tools on the Environment Agency website[21]).

32. A slightly different example is provided by the recent introduction by ONS of a personal inflation calculator[22] that enables users to enter their own expenditure to calculate their own inflation rate, based on movements of the retail prices of their personal 'basket' of goods and services. By so doing, ONS will help users better understand how the retail price index is compiled and how inflation rates can differ from person to person.

33. Meeting the needs of the Internet user is not just a case of putting the information onto a website as facsimiles of paper pages: this common practice often results in users having to type the numbers again into software for analysing the data. Information suppliers also need to think about the structure, and number, of websites, and about issues of website design. While not aimed specifically at statistical websites, the Transformational Government Strategy[23] takes a customer-centred approach to government websites and plans to rationalise their number, migrating content into DirectGov and Business Link websites. Research into DirectGov found that a substantial proportion of respondents (83 per cent) thought it a "good place to start" and felt it to be "trustworthy".[24]

34. ONS is undertaking its own usability research,[25] developing user *personas* in order to ensure that key needs are met. The research findings will inform the ONS website redesign due to be launched in stages beginning in 2008.

[21] *DWP Tabulation Tool,* Department for Work and Pensions, http://www.dwp.gov.uk/asd/tabtool.asp and *What's in Your Backyard*?, Environment Agency http://www.environment-agency.gov.uk/maps/?lang=_e

[22] *Personal inflation calculator*, ONS, January 2007 http://www.statistics.gov.uk/pic/

[23] *Transformational Government Strategy*, Cabinet Office, 2005.

[24] Research showed 83 per cent thought DirectGov a "good place to start"; 79 per cent feel it is trustworthy. *Transformational Government Annual Report 2006*, Chief Information Officer Council, January 2007.

[25] *Office for National Statistics i-dessmination project*, Simonon S, presentation at Statistics User Forum meeting "*Improving Access to Government Data*" (18 January 2007).

SECTION 3: THE ACCESSIBILITY OF UK OFFICIAL STATISTICS

35. Working with Ipsos MORI, the Statistics Commission has tested the ease of accessing and using statistics from official websites. Improvements in the accessibility and ease of use of official statistics are a clear priority for the user community, some of whom have been in formal correspondence with the Statistics Commission regarding accessibility issues – for example, around the Expenditure and Food Survey, or as manifested by the priorities for the Statistics User Forum. In a consultation organised by the Commission in 2004,[26] Statistics User Groups listed issues around accessibility amongst their top priorities. A particular concern has been the need for the less experienced user to have easy access to information about what is available.

36. Accessibility of statistical data has been a recurrent theme in Statistics Commission reports. The Commission's 'user perspectives' reviews of health[27] and schools education statistics[28] indicated that accessibility of data was an important issue for users of the information. As long ago as 2002, the Commission carried out a specific review of the accessibility of transport statistics,[29] which raised a number of similar issues. In a more recent (2005) report,[30] the Commission expressed concerns about whether official statistics were keeping up with the changing needs of users from different sectors, about the frankness and fullness of the commentary that accompanies the figures, and about communication with users of statistics and the public.

37. This report focuses on issues around data accessibility – how easy it is to find relevant and reliable statistical information. It also encompasses issues around the provision of adequate metadata and presentation format. In so doing, we draw a distinction between 'data accessibility' and 'Web accessibility', the latter refers to the usability of websites – whether they are easy to navigate, or whether they have been appropriately adapted for disabled users, etc. Web accessibility is not the focus of this report.

[26] *Initial Analysis of User Group Priorities*, Statistics Commission Meeting 8 July 2004.

[27] *Enhancing the Value of Health Statistics: User Perspectives*, Report No 21, Statistics Commission (2004)

[28] *Schools Education Statistics: Users Perspectives,* Report No 26, Statistics Commission (2005)

[29] *Access to National Statistics on Transport via the Web,* Report No 6, Statistics Commission (2002)

[30] *Official Statistics: Perceptions and Trust,* Report No 24, Statistics Commission (2005)

Research we commissioned

38. We wanted to find out how easy it was for non-subject experts to find specific statistical information on the Web, and to review how well the producers of official statistics are responding to the opportunities of the web format. Following an internal pilot study to test the feasibility and usefulness of the approach, we commissioned mystery shopping research to simulate the experience of people ranging from novices to experienced researchers looking for statistics about topical issues.

39. The topical issues were based on anonymised questions asked of statistical researchers in the House of Commons Library over the three months or so preceding September 2006.[31] A number of the questions asked for statistics for England, Wales, Scotland and Northern Ireland which entailed looking at the multiple websites of devolved administrations.

40. The mystery shopping research was conducted by Ipsos MORI across the UK in October-November 2006. Research was web-based, using a tracker program to capture the facts about the searches and focus groups to explore the experience.

41. In tandem with the mystery shopping, we commissioned (also from Ipsos MORI) a review of departmental policies in respect of the dissemination of statistics via the Web. This review looked at the procedures for data dissemination in the main government departments (defined in terms of numbers of statistics published) as well as the policies underpinning these procedures; departmental perceptions as to who their key users are; and the extent to which user feedback is sought and used.

Research findings

42. Ipsos MORI's report, attached at Annex 2, covers both the mystery shopper exercise and the review of departmental dissemination policies. In the following section, we highlight some of Ipsos MORI's key findings and draw some conclusions, taking account both of Ipsos MORI's key findings and of the contextual information and issues raised earlier in this report.

[31] See Annex 2, page 71 for the questions researched in the mystery shopper exercise

User experience

43. The mystery shopper exercise highlighted the difficulties of searching for statistical information through Government websites. Participants found Google much the easiest route to information, whether or not they were experienced researchers. Overall, fewer than half the researchers judged information to be fairly easy to find and amongst novices this fell to one third.[32] This was partly a problem of not knowing where, among government websites, to start searching for the statistics requested. (DirectGov – the government's preferred site for access to government services and information – does not itself offer statistics and is not specifically designed to link to statistical information. The National Statistics site only covers the 200-odd statistical series produced by ONS,[33] although there are some links to statistical information on other government departmental sites.) Part of the difficulty was also due to the limitations of the search engines on government websites.[34] The ease and speed of finding the data varied according to topic.[35] Web design showed little consistency and better site map or directions would have helped.[36]

44. Whilst the average time taken to find the target information was consistent between the groups at around 15 minutes, the range varied from 8 to 26 minutes. Research initiated from Google took, on average, 33 page views compared to 39 for research that started from the DirectGov site; the qualitative assessment of participants was that Google was more intuitive and easier to use.[37] However, Ipsos MORI also reported that researchers were more focused on finding information from reputable sources while novices trusted a wider range of sources.[38] Ipsos MORI noted that the overwhelming majority of mystery shoppers did not appear to learn search strategies during the course of the exercise. One participant commented, however, that when she began to think like a producer of statistics rather than as a user it was easier to find data.[39]

[32] Annex 2 – Ipsos MORI report, pages 46-47.

[33] Of some 1,000 UK statistical series designated as National Statistics (ie those which must adhere to the National Statistics Code of Practice), 240 are produced by the Office for National Statistics (ONS), headed by the National Statistician. Some 360 series are produced by other central government departments and agencies and nearly 400 by the devolved administrations.

[34] Annex 2 – Ipsos MORI report, page 47.

[35] Annex 2 – Ipsos MORI report, pages 48 and 53.

[36] Annex 2 – Ipsos MORI report, page 47.

[37] Annex 2 – Ipsos MORI report, page 53.

[38] Annex 2 – Ipsos MORI report, pages 59-60.

[39] Annex 2 – Ipsos MORI report, page 54.

45. One explanation for the consistency in time taken by novices and experienced researchers may be that the latter tended to check data against other information on a site and to look more closely at definitional and source explanations. As a result, 54 per cent of experienced researchers were confident in the data compared to 31 per cent of novice users.[40]

46. One website which attracted praise was that of the Scottish Executive.[41] But, in general, there was widespread dissatisfaction with aspects of the presentation of statistics on Government websites (including ONS). In particular the 'shoppers' did not like data to be embedded in large text files such as pdfs. They wanted 'easy to read' charts, bullet points summarising the key trends, graphics and small tables with headline figures. In other words, they wanted products tailored for use on screen with a facility for data manipulation, and not digitalised reproductions of paper documents.[42]

47. Statistics were not always calculated in the way that might be expected and the reasons for them being presented in a certain way were not always apparent – even to those who read the technical notes. Often, important definitions and qualifications were to be found buried in the pdfs of paper publications placed on websites.[43] This suggests that departments need to do more in the way of exploiting web technology to find ways of attaching key interpretative information to the actual statistics.

48. Novice users in particular would have preferred a hierarchical presentation of data that allowed users to "drill down" from headline figures and commentary. Unexplained scientific terms, or statistical or government jargon, often got in the way of comprehension.[44]

Departmental variations in policy and practice

49. The interviews by Ipsos MORI revealed that all eight departments producing the bulk of statistics aimed to meet the standards for availability of statistical data set out in the relevant protocols of the Code of Practice for National Statistics (attached as Annex 1). But that appeared to be the start and end of it – no department had a separate and distinct policy on statistical dissemination.[45] Yet, although departments all follow the same formal policies, there is in practice a good deal of variation in departments' approaches to the

[40] Annex 2 – Ipsos MORI report, page 60.
[41] Annex 2 – Ipsos MORI report, page 44.
[42] Annex 2 – Ipsos MORI report, pages 52 and 60.
[43] Annex 2 – Ipsos MORI report, pages 46, 63 and 64.
[44] Annex 2 – Ipsos MORI report, page 59.
[45] Annex 2 – Ipsos MORI report, page 41.

dissemination of statistics on the Web. This in part reflects different degrees of involvement of other experts (Web, communications, etc) and different perceptions about the "ownership" of official statistics.[46]

50. Hence, the protocols have to be regarded as setting some minimum standards for the publication and presentation of statistics in general – but no more. The specific opportunities and challenges of the Internet are not covered by the Code of Practice. By and large, departments appear to have shown limited ambition in responding to the medium.[47] As Ipsos MORI observe, "the people who control that [statistical] data are still applying a paper-based vision to the Internet".[48]

51. Government made a bold start in 2002 when it announced that the principal vehicle for releasing statistics would be the Internet, and that they would be free of charge. This contrasts with the situation where a number of government departments and agencies operating as Trading Funds charge for their services. Perhaps, however, and perversely, the very existence of the Code of Practice may have hampered further developments in adapting presentation to the new publication medium, as responsibility for the design and practices of Internet statistical release remained with statisticians whose priorities were often internal. The message is clear: to maximise the public use of and benefit from the investment in official statistics, government statisticians need to do better in presenting data, making them easier to find and use, and employing new ways to make official statistics accessible to the expert and non-expert user alike.

Becoming more alert to external users

52. The Chair of the government's Advisory Panel on Public Sector Information (Professor Richard Susskind) has commented in relation to the Transformation Government programme that "…many, but not all, knowledge management initiatives within the public sector are almost exclusively inward-facing, that is, devoted to improved performance and efficiency internally".[49] Government statisticians are perhaps more outward-looking than some other parts of government – but they still tend to see the key users of their statistics as those specialists and experts, more often than not elsewhere in government, with whom they have regular contact.

[46] Annex 2 – Ipsos MORI report, page 43-44.

[47] Annex 2 – Ipsos MORI report, page 45.

[48] Annex 2 – Ipsos MORI report, page 64.

[49] *APPSI Response to Transformational Government*, letter from R Susskind, 1 February 2007.

53. This is borne out by the Ipsos MORI discussions with departments which indicated that key users were usually seen as internal to government or drawn from bodies that had frequent interaction with the department concerned such as local government, academia and large businesses. Moreover, information providers often seemed to expect those using the data to be specialist analysts rather than those without technical training. Some interviewees expressed a keen interest in serving the needs of business and of the media acting as an intermediary and disseminator of official statistics. However, budgetary constraints, organisational pressures and a lack of political sponsorship for statistical initiatives towards these groups were often cited as barriers to progress.[50]

54. Looking at the international picture, our impression is that countries are increasingly accepting the responsibility to make statistics publicly available free of charge. But there is often an implicit assumption that the public is being given access to the information which is needed by central government, rather than access to a statistical service designed to meet a wider set of needs in the public interest. In this respect, the clause in the Statistics and Registration Service Bill which emphasises the importance of public good is much to be welcomed.

55. In contrast to this inward-focused approach, a recent Statistics Commission report on the uses made of official statistics[51] describes the wide range of uses of the statistics made by businesses and organisations beyond government and concludes that (despite the problems of access identified above), "the range of statistics on which decision-makers rely is as wide among users outside government as it is among those inside".

56. Ipsos MORI commented that, whilst specialist user feedback is in some cases assessed and actioned in detail by departments, the handling of more general feedback seems to be patchy and concentrated on the impact of new or *ad hoc* data releases.[52]

57. Not only has the Internet widened the range of users of statistics, it also has the potential to change the nature of the transaction between user and provider in a fundamental way. For example, the growing expectation of Internet users that they can interact with authority in real time – witness the recent development of television viewer participation through comment and photographs of news events or the volume of petitions on the No. 10 website. The scope for statisticians to interact with users of their data and to facilitate the extended use of their data sets has greatly expanded, as have user expectations.

[50] Annex 2 – Ipsos MORI report, page 40.

[51] *Use Made of Official Statistics*, Report No 33, Statistics Commission (2007)

[52] Annex 2 – Ipsos MORI report, page 43.

Website design

58. Departments differed in their approaches to web publication of statistics. Ipsos MORI found that accessibility of statistics was improved where communications staff were involved in the dissemination of data, or where outsourced web publishers were used. However, they reported a sense that statisticians were reluctant to involve others in the process of web publication of statistics.[53]

59. This desire to remain totally in control of data and their interpretation illustrates a dilemma for professional statisticians. On the one hand they want their data to be used – unused data are worse than yesterday's news. On the other hand they are only too aware of the qualifications and caveats that attach to what are usually estimates of a largely uncertain quantity. Wrongly interpreted data could have serious consequences. Ipsos MORI investigated the scope for a 'one-stop shop' for statistics but found that statisticians were not generally in favour of the idea.[54]

60. The Commission believes this desire to prevent misuse should not justify a monopoly on interpretation of data, still less on involvement with the design of how data are disseminated. There would appear to be substantial potential gain from a more open collaboration between statisticians and those with other skills such as marketing, communications and web design. The work of Surowiecki[55] and the worldwide success of Wikipedia may portend a change towards wider democratic participation in the interpretation and influencing of statistical priorities.

Section 4: Conclusions and recommendations

61. The governance of UK official statistics is presently in flux. We look forward to a new Statistics Board, reporting to Parliament, as proposed in the Statistics and Registration Service Bill currently going through Parliament, and accordingly, we address these recommendations to the new Board. In so doing, we recognise that the Board will decide for itself how it will operate and that, as we finalise this report, Parliament has not decided how the Board will be held accountable.

[53] Annex 2 – Ipsos MORI report, pages 43-44.

[54] Annex 2 – Ipsos MORI report, page 44.

[55] *The Wisdom of Crowds*, Surowiecki J, Little Brown, 2004.

62. We therefore present our recommendations in the form of eight principles, which we believe should determine future Internet dissemination policy for UK official statistics.

63. Statistics are collected, compiled and published in order that people may use them. We believe that the publishers of official statistics should be doing whatever they can to encourage their use. So our first principle of statistical dissemination is:

- Principle 1: **Statistics are collected to be used and as wide a use of them as is possible should be encouraged**, including the re-use of raw data for research outside government.

64. This report has described the major changes brought about by the Internet. UK government statisticians have readily embraced the Internet as a vehicle for publication of statistics, but have been slower to adapt their presentation and publication policies to the new medium. We believe that they need to be more ambitious. Our second principle is:

- Principle 2: Since the most satisfactory forms of data provision are still evolving, **UK government statisticians should adopt an exploratory and experimental approach to dissemination and access to statistical data through the Internet.**

65. A key finding of Ipsos MORI's review of departmental dissemination policies was that, where departmental statisticians had involved web professionals in the design and presentation of their statistical websites, those sites were noticeably more user-friendly. Our third principle recognises this:

- Principle 3: **Government departments that publish official statistics should seek the full involvement of other web professionals in the presentation of statistical data on their websites.**

66. Our fourth principle looks forward to the future. The Internet has transformed the ways in which information, including statistical information, is disseminated and accessed – but that transformation may be as yet far from complete. The publishers of statistics need to keep accessibility issues under continual review:

- Principle 4: **Government departments that publish official statistics should recognise that web design and web culture are still developing and should set up an appropriate mechanism to keep accessibility issues under review.** This might take the form of a technical advisory panel, with membership drawn from external user expertise such as resides in user-centred website design, communications experts, the Statistics User Forum (SUF) and the House of Commons Library.

67. As our first principle emphasises, statistics are for their users. In designing a statistical website, the first concern should be meeting the needs of users and potential users. This is our fifth principle:

- Principle 5: **User needs, interests and capabilities should determine the design and operation of statistical dissemination over the Internet.** This necessitates interactive engagement with users and active pursuit of feedback plus better search engines (possibly in co-operation with commercial operators). Users of government statistical websites should not be expected to have a working knowledge of government departments or of 'who produces what'.

68. Website design is important. It is not sufficient to simply download a set of statistical publications and releases designed to be read as 'paper' products onto the department's website. Statistical dissemination through the Web should be specifically designed for the Web – this is the message of our sixth principle:

- Principle 6: **Statistical products should be specifically designed for the Web.** The Internet brings a new dimension to the relationship between producers and users of statistics. It has transformed the way that people look for information and want to access information.

69. Our seventh principle builds upon the fifth ('design for users') and the sixth ('design for the Web'). It would require the publishers of official statistics to design their web-based products with the aim of making them accessible to all users:

- Principle 7: **Data should be presented in a layered or hierarchical way to allow users to drill down to the level of detail they desire.** Tables, charts, maps and online statistical programs for manipulating data should be increasingly the norm. Guidance on definitions, sources and methods, data quality and interpretation should be an integral part of dissemination (possibly in the form of Frequently Asked Questions).

70. The mystery shopper research suggests that official statistics are not that easy to find. One reason is that people do not know where to start looking – the National Statistics site may be the place to go for ONS statistics, but only around 20 per cent of official statistics are produced by ONS. Our final principle looks to correct this:

- Principle 8: **There should be one point of entry – a government statistics portal – giving access to official statistics across the UK government and those of the devolved authorities.** This portal would link to all government statistics sites, and include a search engine which operates seamlessly over all those sites.

Annex 1

Extracts from National Statistics Code of Practice and Protocols

Introduction

71. The National Statistics Code of Practice and Protocols, launched in 2002, sets out the principles and standards that official statisticians are expected to follow and uphold.[56] This Annex presents extracts from the Code and Protocols that are relevant to the accessibility of statistics.

National Statistics Code of Practice

72. The Code of Practice states that *National Statistics will be*:

- *valued for relevance, integrity, quality and accessibility*
- *produced in the interests of all citizens by protecting confidentiality, and balancing the needs of users against the burden on providers*
- *enhanced through integration, accumulation and innovation; and by efficiency in costs, and fairness in prices.*

73. Three of these principles are most directly relevant to accessibility of statistics:

- *Relevance: National Statistics will inform significant decisions in government, business and the wider community, and in so doing contribute to the quality of national life.*
- *Quality: National Statistics will be fit for purpose and of high quality.*
- *Accessibility: Access to National Statistics will be fair and open.*

Code of Practice Protocols

74. There are twelve Code of Practice Protocols that set out how those involved in the production of National Statistics carry out their responsibilities. Three of these are of direct relevance to the accessibility of statistics:

- Protocol on Data Presentation, Dissemination and Pricing
- Protocol on Quality Management and
- Protocol on Customer Service and User Consultation.

75. We have selected relevant extracts from these protocols and these are presented in the sections that follow. Three dots (…) denotes where we have omitted text. The full Protocols are available on the National Statistics website.

[56] The National Statistics Code of Practice (2002) and protocols can be found at: http://www.statistics.gov.uk/about/national_statistics/cop/default.asp

EXTRACTS FROM THE PROTOCOL ON DATA PRESENTATION, DISSEMINATION AND PRICING

1. National Statistics will be made accessible to the widest possible community, and where appropriate with a choice of format, helping users to get what they want simply and quickly.

…

b) The web will be the primary means of providing general access to National Statistics.

i. When developing new products, and when reviewing existing products, producers of National Statistics will consider all ways and means of making statistics accessible, and where practicable, will give high priority to dissemination on the web.

ii. The National Statistician will work with Heads of Profession to provide user-friendly and coordinated entry-points to the entire range of outputs disseminated as National Statistics.

iii. Content will be organised as much as possible so that it can be easily understood and easily found.

iv. Producers of National Statistics will encourage comparative analysis and help to provide context to any particular output, through facilities such as metadata, links to related information, and cross-referencing to glossaries and background material.

…

2. Presentation of National Statistics will be integrated and will focus on users' needs.

a) Producers of National Statistics will endeavour to integrate and harmonise their publications and products in accordance with users' needs and give users easy access to related statistics through common gateways or interlinked websites.

b) Producers of National Statistics will respond to changing expectations about access to outputs. Formats, media, content and support materials will be regularly reviewed, and informed by an understanding of users' current and future needs and the standards set out in the *National Statistics Code of Practice.*

EXTRACTS FROM THE PROTOCOL ON DATA PRESENTATION, DISSEMINATION AND PRICING *(continued)*

c) Within the guidelines of departmental policy and the *National Statistics Code of Practice*, producers will exercise professional judgement to decide on the types of presentation – including commentary, analysis and interpretation – best suited to the range of users for a particular output.

3. Information will be presented objectively, in line with professional standards, and in ways that make the statistics clear and useful.

…

iv. Where useful and appropriate, presentation of data should include analysis and background information on methodology, quality, trends, etc to aid understanding of data.

…

vi. Statistical presentation and commentary will be adapted, wherever practical, to suit the different needs of the various communities and audiences likely to be interested in a particular output.

b) Consistent departmental formats will be applied to released outputs, including representations such as graphs, tables and maps.

i. Producers of National Statistics will follow departmental house-style guides and refer to appropriate best practice when presenting numerical and graphical information.

…

6. Pricing will comply with legislation and wider government policy, and will provide fair value.

…

i. Headline National Statistics outputs and other information of broad and general interest will be free on the internet.

EXTRACTS FROM THE PROTOCOL ON QUALITY MANAGEMENT

…

3. National Statistics will meet the needs of government, business and the community, within available resources.

Producers of National Statistics will have effective mechanisms in place to help them understand the key uses of their outputs, and the key areas of emerging demand and unmet need. In particular, producers of National Statistics will:

a) identify and consult key users to gain an understanding of their needs and to help determine priorities;

…

d) develop concepts, and design outputs, that are relevant to users' needs and intended uses, but that also take account of the relationship between quality and cost;

e) ensure that outputs are accessible and clearly presented to users;

…

4. Knowledge of National Statistics will be sustained by quality measures, regularly published.

…

c) quality measures will help users to understand better the strengths and limitations, interpretation and appropriate use of the data, and will enable producers to demonstrate changes in the quality of outputs;

d) where practicable, the presentation of information on quality will be tailored to meet the needs of different types of users, with more comprehensive quality reports prepared for 'expert users';

e) guidance will be provided to users on the interpretation of quality measures.

EXTRACTS FROM THE PROTOCOL ON QUALITY MANAGEMENT *(continued)*

…

8. Processes and methods used to produce National Statistics will be fully documented.

a) National Statistics will be readily accessible through the web. They will be well described and supported by published documentation on statistical methods and processes which is easy to access and use;

b) documentation will be sufficiently detailed to allow users to assess fitness for particular purposes;

…

EXTRACTS FROM THE PROTOCOL ON CUSTOMER SERVICE AND USER CONSULTATION

Customer service

…

2. Access to National Statistics will be made as easy as possible.

a) Producers of National Statistics will provide facilities which strive to ensure easy, user-friendly access to National Statistics for everyone, including new and non-specialist customers.

b) In addition, the National Statistician will:

…

ii. Provide online services, continuously available, where information will be set out to meet the needs of regular, professional users as well as casual users and the interested public.

…

3. Organisations producing customer feedback.

…

c) Producers of National Statistics will make regular assessments of customer satisfaction focusing on key products and activities.

…

iii. Users will be consulted to assess the relevance of products.

EXTRACTS FROM THE PROTOCOL ON CUSTOMER SERVICE AND USER CONSULTATION *(continued)*

User Consultation

1. User consultation will be an integral part of the statistical process.

An open culture involving genuine and effective user consultation is fundamental to strengthening public confidence in National Statistics. Producers should develop a clear view of who their users are, what are their interests, and how their needs can be met.

…

3. Engagement with users will aim to be focused, scaled and inclusive.

a) Producers of National Statistics will recognise the wide diversity in the user community and the fact that customer groups have differing needs and resources and differing degree of access to the Internet.

Annex 2

Accessibility of official statistics online

Report to the Statistics Commission by Ipsos MORI

Contents

Publication of the data

As with all our studies, these findings are subject to Ipsos MORI's standard Terms and Conditions of Contract. Any press release or publication of the findings of this research requires the advance approval of Ipsos MORI. Such approval will only be refused on the grounds of inaccuracy or misinterpretation of the findings.

Acknowledgements

Ipsos MORI would like to thank the following for their assistance:

- House of Commons Library
- Department for Education and Skills
- Department for Environment Food and Rural Affairs
- Department of Health
- Department for Transport
- Department for Work and Pensions
- Home Office
- Office for National Statistics
- Scottish Executive

INTRODUCTION

Web accessibility has been defined in various interlocking ways from access for disabled users and the usability of websites, to general access by citizens as a measure of the robustness of the country's ICT economy.

Data accessibility, on the other hand, examines the ability of users to find the information they are looking for, to understand what they have found and recognise that they've found what they were looking for, and to manipulate the information to ensure it satisfies their needs.

Increasingly, web-savvy users have placed growing demands on this level of data accessibility, combined with timeliness and accuracy. Because of their very nature, however, official statistical data often vary according to their conceptual bases, their content, scope and timeliness. Web access, however, can open up such complex data resources to many potential users in an unmediated and uncontrolled manner, as new users juggle information from multiple government agencies in varied formats.

The Transformational Government reforms – which are geared towards harnessing technology to develop a new relationship between citizen and State – aim to reduce this information overload by limiting the number of governmental websites. However, the nature of the Internet will mean that routes into information sources are likely to proliferate, creating dilemmas for data controllers. However, as Eric Schmidt, Google's chief executive, recently wrote:

> *Don't bet against the Internet … The past few years have taught us that business models based on controlling consumers or content don't work.*[57]

The functionality of current access tools also varies considerably. For example, search engines like Google are designed to allow the computers to do all the work in deciding which sites users should access from whatever data is entered by the user.

By contrast, official sites tend to have intrasite search tools, and their statistics sections often assume a working knowledge of the output and the tools required to get there. Each intrasite search tool makes its own assumptions, and while some of these differences undoubtedly reflect the real differences in the nature of various organisations' data, from the user's perspective, the task of specifying and identifying data is essentially the same irrespective of these variations and search engine functionality. This lack of a common language and of shared Internet conventions for UK official statistics is likely to hamper data access.

[57] Schmidt, E. "Don't bet against the Internet". *The World in 2007*. The Economist, 2007.

The importance of a shared language has long been recognised by the Internet's pioneers as the key to what has been termed the 'semantic web':

> *Progress towards better data integration will happen using the same basic technology that has made the world wide web so successful: the link … The key to this integration is to use common data formats that link the information with identifiable vocabularies*[58]

This apparent lack of 'read-across' was one of the reasons why the Statistics Commission invited Ipsos MORI to carry out this research – with the following objectives:

The inquiry will consider how accessible various data are, how easy to retrieve from the Internet, including the provision of any relevant metadata, assessing data quality, the format of the data, and the needs of users ... It will also include an assessment of individual departments' policy on data accessibility and dissemination

Methodology

There were two main strands to this research programme. Firstly a review of departmental data accessibility policies, and secondly, primary research into the ease with which official statistics can be found and understood online.

The policy review

A variety of departments were chosen to be illustrative of the breadth of likely public data needs. Firstly, relevant departmental policies dealing with the issue of data accessibility were gathered and examined to ascertain coverage and likely gaps. Then, the research team spoke to the people responsible for data dissemination within each department (in the main, this tended to be statistical Heads of Profession, although communications and web professionals were also involved).

The aim was – in a confidential setting – to discover the extent to which stated policies have been achieved. We also sought to ascertain:

- the procedures for data dissemination that are followed and the rationale behind them;
- the extent to which these procedures are kept to;
- the lifespan, ownership, sponsorship and review timings of relevant policies;
- the display formats, functionality and metadata of statistics published;
- any confidentiality or freedom of information constraints that may hinder full data access;

[58] Berners-Lee, T. "Welcome to the semantic web". *The World in 2007*. The Economist, 2007.

- departmental perceptions of their key user groups, and
- the extent to which user feedback is captured.

The primary research

In order to assess the accessibility of official statistics online, 80 people were recruited to participate in an online 'mystery shopping' exercise. These participants were recruited based on their level of expertise in terms of conducting research online, and as a result, 40 'novices' and 40 'researchers' were recruited (the difference between the two groups being experience searching the Internet for statistical information).

Each participant was assigned a series of five questions and was asked to find, to the best of their ability, the answers on the web.[59] The full list of 20 questions (from which five were randomly assigned to each participant) were derived from questions asked recently by parliamentarians and their staff of the House of Commons' Library. The complexity of each question also varied, with some requiring information to be sourced from more than one website.

The 20 questions were produced by the Statistics Commission and were grouped according to national statistics theme. Topics covered included:

- agriculture, fishing and forestry
- commerce, energy and industry
- crime and justice
- economy
- education and training
- health and care
- labour market
- natural and built environment
- population and migration
- social and welfare
- transport, travel and tourism
- other national statistics

[59] Those designated as 'researchers' were specifically not asked to find information in their own fields of expertise.

Each topic also included questions that required participants to find either tracking data (eg how does this compare to five years ago?), comparative data across Wales, Northern Ireland and Scotland, or increasingly detailed data (e.g. how many farms are there currently in England? How many farmers are there? How many farm workers are there?).

Participants were also given one of two points from which to start their search – either Google or direct.gov.uk.[60] These starting points were randomly assigned with each participant receiving a combination of the two. While each participant was given a starting point, they were not instructed to continue to use Google or Directgov throughout or to move from one to the other if their initial search yielded poor results. In order to ensure that the search was conducted in as natural a way as possible, participants were welcome to use whatever means they felt necessary to find the information requested as long as they began at their assigned starting point.

While this project embodied many of the underlying principles of mystery shopping, it did not require the shoppers to make any form of human contact while carrying out their tasks. It did, however, require them to make some form of subjective assessment of their experience, rating it in terms of how easy it was to find the information, whether they understood it, and how satisfied they were with the experience immediately after they completed each 'shop'.

Alongside this user assessment, we were able to generate more objective information. Keynote Systems' WebEffective™ monitoring tool provided us with a reliable way of recording and analysing the mystery shopping process. This loaded a program (with participants' permission) onto people's home PCs which tracked their search profiles and built up a map of the route taken by each shopper to find the data, any search terms used and the time they took to complete the 'shop'.

Following the completion of their shops, participants were invited to a series of de-briefing discussion groups, where they shared their experiences and drew up recommendations for a more effective presentation of official statistics online.

[60] direct.gov.uk was chosen as a starting point for 'governmental searchers' as it aims to provide a one-stop shop for government information with links to all the major departments.

Interpretation of the data

It should be kept in mind when reading this report that these findings are drawn exclusively from a qualitative methodology. While these interviews and groups featured a good cross-section of individuals, they (and therefore the findings drawn from them) may not be said to be statistically representative of the larger target population. While focus groups generally indicate appropriate directionality, they do not serve as a proxy for a fully representative quantitative methodology. For the reader's ease, these findings are depicted to some extent as definitive and universal. This is, however, true only for the universe represented by these participants.

Rather than paraphrase responses from the policy review and mystery shopping exercise, we have generally left the participants to speak for themselves in the form of verbatims quoted wherever possible. These should not be interpreted as defining the views of all concerned but have been selected to provide an insight into a particular body of opinion.

Oswin Baker
Lisa Valade-DeMelo
Seirian Hanner
David Henderson

THE POLICY REVIEW

The aim of the policy review was to link the user experience of accessing official statistics online to the departmental context within which the statistical data was produced. This exercise helped us assess the extent to which accessibility policies are linked to the actual user experience, aiming to determine what helps – and what may hinder – the effective dissemination of clear and easy to access statistics.

Who is the user?

Before turning to those policies, it is worthwhile to define what is meant by a 'user' of official statistics. At the international level, the United Nations defines key stakeholders for access to official statistics as:

- Government – encompassing the full range of departmental and topic needs currently covered by official statistics;
- Academia/International Institutions – key needs include rigorous, standardised and internationally comparable data;
- Business – both small and large businesses, each having clearly different needs;
- The Media – acting as an intermediary and disseminator for official statistics; and only then
- The general public.[61]

The policy review was partly designed to test this user typology, as it should have some significant implications for the presentation of official data.

For instance – it is highly likely that the key user of official statistics in volume terms is the public sector, encompassing central and local government as well as more 'generic' public sector specialists such as librarians. If this is the case, then it is also reasonable to assume that the structure and processes of statistical departments will have evolved to reflect this.

However, if policies were to be principally designed to serve this user group, departmental abilities to meet the data needs of the general public are also likely to be hindered, as the public will have different needs compared to specialist audiences. Furthermore, the degree to which departmental policies are structured to allow the other three key stakeholders – business, academia and the media – access to statistics will also impact on how statistics are presented in the online context.

[61] *Handbook of Statistical Organisation. Third Edition*. United Nations, 2003

Therefore, as well as assessing departments' views of their key audiences, the policy review also aimed to evaluate the extent to which current policies, and more importantly departmental practices, are coping adequately with these changing demands.

The policies

It should be stated at the outset that – while the departments under review all had Statements of Compliance with the ONS protocols as well as clear publication policies – we could not find any specific departmental policies which related *explicitly* to data accessibility.

Overall, the people we spoke to (generally, Heads of Profession) were helpful, transparent and clearly keen to help. It was also reassuring to find that they understood what was meant by 'data accessibility policy'.

However, each of the departments under review had no specific policy document or clause directly relating to the issue of data accessibility as defined in this study.

Instead, each department, including the ONS itself, set out any policies relating to accessibility in the form of a Statement of Compliance with ONS protocols. These are a list of detailed guidelines and practices covering a range of issues relevant to statistical dissemination – key protocols referred to in these Statements of Compliance were those on Release Practices, Data Presentation, Data Access and Confidentiality, and Quality Assurance.

The key tenets of ONS Protocols

Release Practices
- Equality of access
- Early release protocols

Data Access & Confidentiality
- Acceptable levels of analysis
- Range of confidentiality considerations

Data Presentation
- Wider community access
- Web as primary access point

Quality Assurance
- Meeting user needs

These Statements allow each department to set out links with other policy documents, any special confidentiality arrangements pertaining to market sensitive data and timescales for publication.

However, we found that any such arrangements appeared to be shaped for the convenience of each department rather than relating to any centrally understood view of publication procedures.

Furthermore, for the most part, no scheduled reviews of these Statements were planned, and where there had been a review, this tended to be in reaction to wider departmental or statistical changes. As a concrete demonstration of this, we found some Compliance Statements which referred to dead or superseded Internet links, suggesting that little ongoing review is in progress.

In all cases, the Statements, protocols and relevant policy documents are generic rather than prescriptive in the guidance that is given. For instance, although timings are set out for the release of statistics, these tend to be in the form of an agreed benchmark to which departments sign up to. These practices ranged from a detailed list of publication days published a year in advance, to a looser provision of 'month of publication', with a proviso that precise dates would be given a fortnight in advance. Given that each of the departments defined their own release practices, it is hardly surprising to find that they met these benchmarks with little difficulty.

Understanding the user

The one area which most clearly highlights the different contexts within which each department operates relates to their understanding of their users. In all cases, it would be fair to categorise the key user as 'the specialist'.

A review of key users listed on the National Statistics website's statistics users' pages confirms this, with academic, government, business and international institutional users listed as active participants of various user groups. [62]

The general public, however, is conspicuous by its absence. That said, in all cases, we found a keen interest within departments to understand the information needs of the public more fully. There was a recognition that different users have different needs – with specialists and practitioners requiring data which can be intelligently interrogated while the general public need 'topline' and trend information – and that the presentation of statistics online does not generally meet the needs of the general public.

However, budgetary constraints, organisational pressures and a lack of political ownership tend to mean that putting the general public user at the centre of data dissemination is not a priority. Another blockage relates to the thorny issue of how statistics can be presented in a way which meets the needs of all users, with a natural suspicion among interviewees of a 'one-size-fits-all' approach.

[62] *http://www.statistics.gov.uk/about/statisticsusers/user_groups.asp*

ONS has taken the lead on finding solutions to this question – by developing user 'personas' which can help it target data dissemination more effectively. This is an approach which has been used successfully in the commercial world, and personas have a strong grounding in the research community:

> *Personas are archetypes or representations of real people. Personas effectively put a face on the data and help companies avoid the common perils of product design – the "elastic consumer" who is all things to all people and is never concretely defined, or watered down "one size fits all" solutions. When used as a communication tool, personas ensure that the organisation can not only internalise its customers' needs, but understand at a visceral level who they are and be able to articulate this effectively*[63]

Clearly, if departments are to benefit from the development of user 'personas' or even to find their own route towards an accessible presentation of statistics online, they need to understand and engage with all their users. However, in all cases, we did not encounter any specific targets with regard to the user experience – either specialist or general public – aside from customer complaint-style benchmarks (such as the ONS's ten working day response time for customer correspondence). While specialist user feedback and experience is in some cases assessed and actioned in detail, feedback from the general public tends to be handled in a more *ad hoc* fashion, although it should be noted that some of the larger departments have developed more proactive feedback channels to monitor the impact of new data releases among the general public.

Ownership and control

Statisticians from each contacted department tended to be satisfied with the processes they undertake when disseminating data online. These processes are currently designed and run by the statisticians themselves, and they therefore could justifiably be said to have a large degree of 'ownership' over them.

We were, however, able to identify instances where the level of ownership was shifting – and as a direct result of a new conception of data accessibility. The key difference underpinning differing accessibility practices was that of the involvement of non-statisticians in the dissemination process: marketing and communications professionals in the publication process and outsourced web publishers in the posting of statistics. According to one interviewee, when marketing people were involved, they provided a "brilliant service". Another explicitly mapped out why they wanted to work with these people: "We are trying to benefit from their experience in terms of improving accessibility and the understanding of material".

[63] Bertelsen, L. *Personas: Putting a Face on Market Research Data. Presentation at ESOMAR Conference.* Ipsos, 2005.

Where these other professionals become involved, data appears to be more accessible. There is, though, a tension inherent in such a process. As one interviewee noted, "Obviously, we control the content of the statistical element of it, and sometimes people forget that".

Finding the right level of detail

> *The Scottish Executive one, for example, is quite a good one because you can go into the statistics part and then you've got all the topics, and then you can follow through, you can always see the whole range of things they've got*
>
> Female Researcher, Edinburgh

Following an accessibility review and discussions with marketing and communications colleagues, extensive changes were made to the statistics sections of the Scottish Executive website (http://www.scotland.gov.uk/Topics/ Statistics/). In particular, a new level of information was developed: the High Level Summary of Statistics, which provides the headline figures which many members of the general public may require. In most cases, this summary – which is thematically grouped – is followed by links to further, more detailed data.

Another aspect of the 'control' question is whether there is scope for a 'one-stop shop' for statistics. The concept of a centrally integrated system could be achieved in a number of ways, from fully harmonised data systems and processes to a central 'portal' of links to relevant sites. While it is felt that the ONS site partially fulfills this model, there is a general view among the statisticians who were interviewed that the ideal of the 'one-stop-statistical shop' does not yet exist. There are also the questions as to whether it *can* exist – and whether it *should* exist.

Departments see the budgetary implications involved in ensuring the many legacy systems and processes currently used across the country – and sometimes even within one department – as being too great a barrier to achieving a more comprehensively integrated model. Additionally, the speed of technological change would likely render any initiative obsolete before completion. Such considerations led to a general view among statisticians that a fully harmonised central hub of statistical data is not achievable in the short- to medium-term.

It should also be noted that a number of interviewees felt that such a centralised approach to official statistics would not necessarily be a wholly positive development. Again, the question of control would appear to be a significant factor here, with professional pride at stake.

A user-centred approach

The peculiarities of process, the specialist nature of key users and the demands of the statistics themselves appear to take priority over any attempt to make statistics accessible to the general user. However, partly due to the clearly defined but highly limited standards of compliance that departments have set themselves, this approach gives the appearance of 'working', at least from the inside.

We anticipate that any major changes that challenge current practices are likely to cause internal tensions around the issue of statistical ownership. At present, the key focus of standards, processes and policies is to ensure control over the quality and dissemination of data. A shift towards a more user-centred, accessibility-focused set of policies and practices will require new skills to be explicitly recognised and incorporated into statistics units. These skills are strongly represented within allied professions – communications officers and web designers – which can help to bridge the gap between highly technical, detailed and accurate information, and the needs of the general public. This would seem to be a point of cultural change, however, that may prove difficult to resolve.

It should also be recognised that the Internet allows for a wholly new way of envisaging the presentation of information. As Eric Schmidt notes, "Simplicity is triumphing over complexity. Accessibility is beating exclusivity. Power is increasingly in the hands of the user".[64] But when conducting the policy review, we did not find evidence of a widespread understanding within departments of the challenges and opportunities afforded by the web.

THE PRIMARY RESEARCH

As will be seen, the mystery shopping confirmed many of the concerns raised by the policy review.

While the 'researchers' were more likely than 'novices' to worry about the context of the data and the criteria of search items (such as the definition of "a farmer"), all participants assumed that their tasks would be relatively easy. They also assumed that the answer would exist in the exact form they required – that a table or chart existed somewhere that would contain what they were looking for.

> *I think sometimes when I found things where I did have to add things together I thought, you shouldn't have to do this. The entire answer must be somewhere in the form that's required*
>
> Female Researcher, Edinburgh

[64] Schmidt, E. "Don't bet against the Internet". *The World in 2007*. The Economist, 2007.

A lot of the government websites have got lots and lots of writing and you don't get clear direction

Male Novice, London

Finding the information

However, when asked to rate how easy it was to find the information immediately after completing each task, fewer than half (42%) report that the information was 'very' or 'fairly' easy to find. While this differed slightly between researchers and novices, still fewer than half of researchers (49%) report having found the information easily. This is compared to only a third of novices (34%).

In fact, in order to come to what they felt was the correct answer, one technique highlighted in the groups was to derive what was felt to be the 'right' statistic from a number of information sources which they had found. Calculating the result themselves was considered a necessity by some participants, especially when the question asked for specific criteria that the individual could not find immediately. This approach, however, was not universally supported – particularly among researchers who had concerns for the final reliability of personally-calculated results.

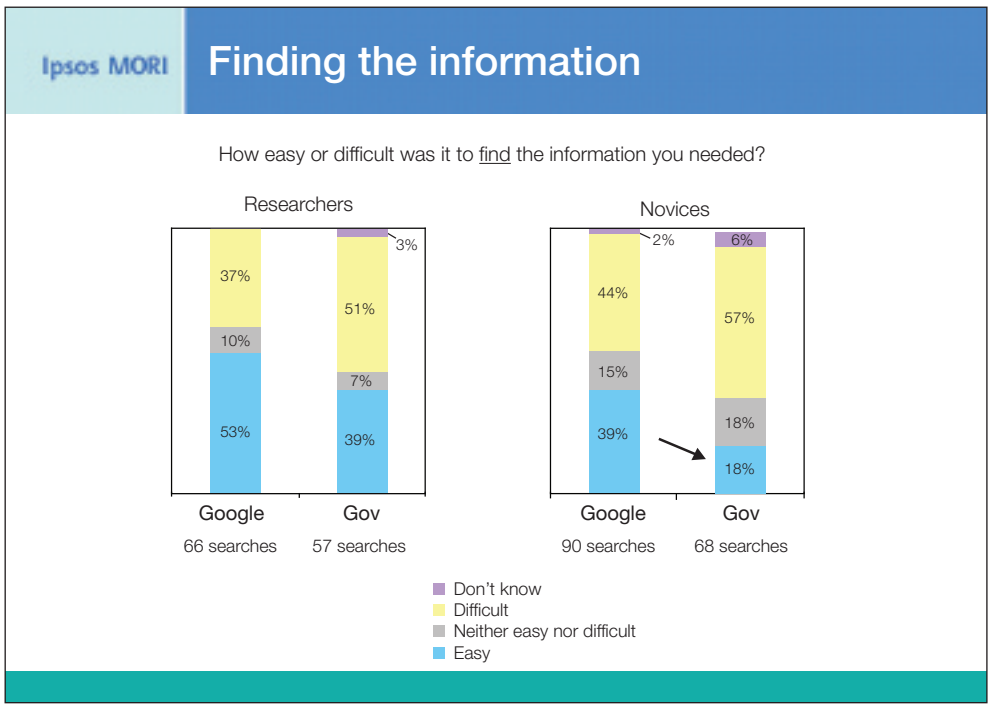
I got a calculator out and calculated it myself because the table never gave me exactly what I was looking for

Female Researcher, Bristol

I had a question on migration and I ended up having to do some calculations to get the numbers that I was being asked to get

Female Novice, Leeds

A stark difference emerges when comparing the starting point for each type of participant. Both researchers and novices found the information easier to find when using Google as a starting point – more than half of researchers (53%) found the information easy to find as did 39% of novices when starting from here. When using direct.gov.uk as a starting point, however, only 39% of researchers and 18% of novices found the information easily.



It's easier to find DWP publications in Google than it is on the DWP site

Male Researcher, London

The direct.gov website was not very useful as the links weren't often very clear ... I would use the search tool on the website itself and type in key terms. This often was not very helpful either. An easier way was to use Google and type in the statistics wanted. This would generally give quick links to statistics and government websites

Male Researcher, Leeds

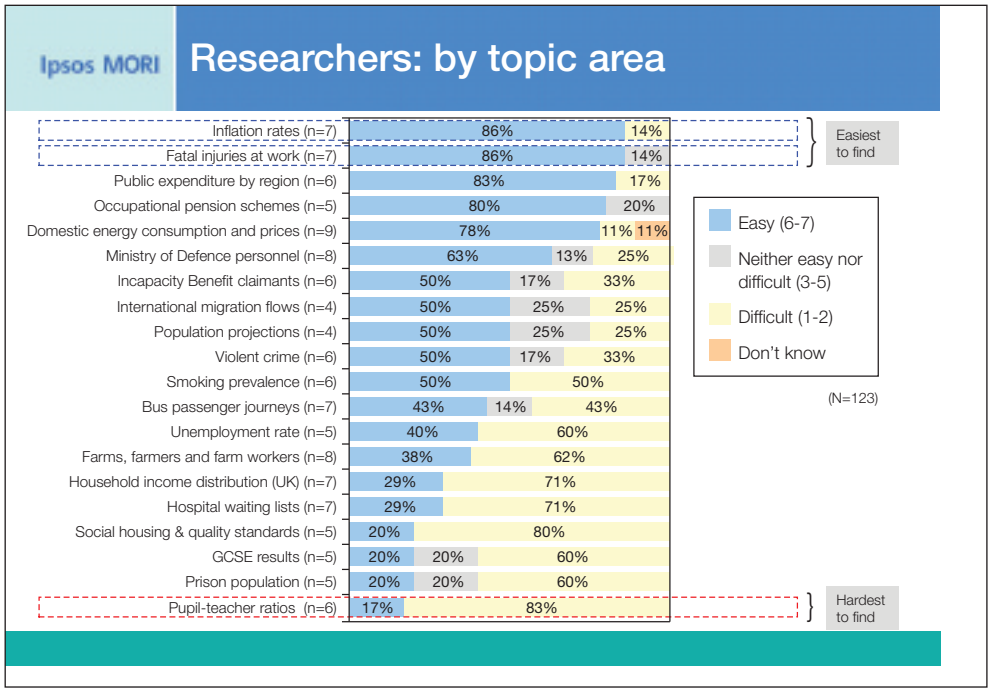
All the information could be found on the National Statistics site and government department websites. However, researchers mentioned a lack of consistency between sites and differing website structures as barriers to accessing key information efficiently.

It's quite difficult to get to the source data sometimes because, well, certainly on the National Statistics site, they've put this layer on top to steer you towards an overview of the latest figures, but then frequently that doesn't answer the precise point which you're trying to answer

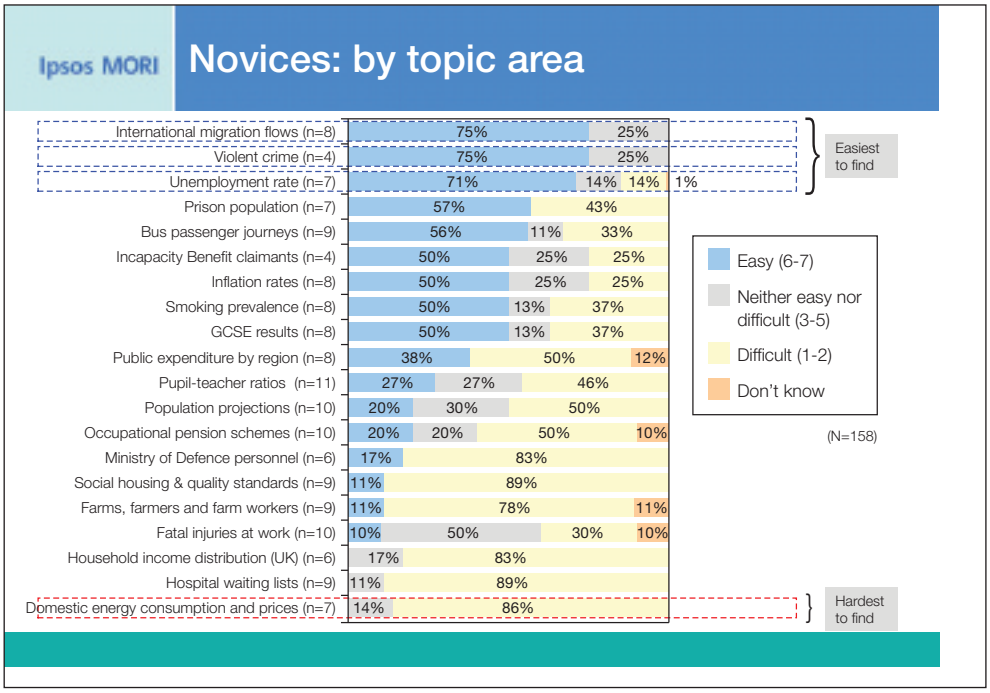
Male Researcher, Bristol

For both researchers and novices, how easy the information was to find depended heavily on which questions they had been assigned. For researchers, the easiest information to find involved inflation rates and fatal injuries at work (86% found both of this information very or fairly easily). On the other hand, 83% of researchers found pupil-teacher ratio data very or fairly difficult to find.

> *On the site they do have regional information, which Scotland tends to be, but I found that difficult to know if they were comparable to what I was looking at from one to the other* [this participant was trying to source information for each of the UK countries]
>
> Female Researcher, Edinburgh



On the other hand, novices were most likely to find international migration flows and violent crime data easily (75% found this information very or fairly easily) and domestic energy consumption and prices relatively difficult to locate (86% report that it very or fairly difficult). While we do not have information on why this is the case, we have surmised that this could be related to the topicality of 'media-friendly' subjects.



I couldn't find anything in the official sites [for MoD data]. I found a lot of information. I found PDF links to the MoD site, which gave really specific information. Although it was interesting it wasn't what I was looking for

Male Novice, Edinburgh

If you put prison population into the search on that website [direct.gov.uk] *it didn't come back with prison population. It comes back with other things, where prison or population are in. That was the trouble*

Male Novice, Bristol

The starting points

Both researchers and novices are very comfortable with Google as a means of searching for information on the Internet. All participants started by entering keyword search terms, accessed the resulting pages, refined their keywords if necessary, and eventually accessed a website they felt would garner the response they were searching for. Where they differed, however, was in the initial keywords used and the sites accessed.

> *I'm putting in terms, which I think will find me that information. And then I check the source. That looks good, right click on that. Right now where am I? What is this source? Is this an official source?*
>
> Male Researcher, London

> *Sometimes you do use quote marks. I also found that putting in words like statistics and figures helped because that tended to mean that statistical sites containing the figures were at the top*
>
> Male Researcher, London

It should be noted that, regardless of what keywords were used initially, for most shoppers, Google's first suggested site was usually a government department site, and in half the cases tested here, the first suggested site did in fact contain the correct answer (for instance, when searching for the number of farmers in the UK using the keywords detailed below, Google would return a hit from the DEFRA website, where the correct answer can be found).[65]

[65] The correct answers to each question and the websites where they could be found were provided by the Statistics Commission in the design phase of the project.

Question	Frequent keywords used	First Google suggestion	Where the answer(s) can be found
1	England farms total number 2006	www.defra.gov.uk	www.defra.gov.uk
2	Domestic consumption fuels	www.dti.gov.uk	www.dti.gov.uk
3	Fatal injuries construction	www.cdc.gov	www.hse.gov.uk
4	Crime figures	www.crimestatistics.org.uk	www.homeoffice.gov.uk www.scotland.gov.uk www.psni.police.uk
5	Prison statistics	www.homeoffice.gov.uk	www.homeoffice.gov.uk www.scotland.gov.uk www.nio.gov.uk
6	Public expenditure	www.hm-treasury.gov.uk	www.hm-treasury.gov.uk
7	Inflation rates	www.statistics.gov.uk	www.statistics.gov.uk
8	Percentage of children with five or more GCSE	www.dwp.gov.uk	www.dfes.gov.uk www.new.wales.gov.uk www.scotland.gov.uk
9	Education statistics	www.dfes.gov.uk	www.dfes.gov.uk www.scotland.gov.uk
10	Percentage of smokers	www.ash.org.uk	www.statistics.gov.uk www.ic.nhs.uk www.csu.nisra.gov.uk
11	Hospital waiting lists	www.performance.doh.uk	www.performance.doh.uk www.statswales.gov.uk www.new.wales.gov.uk www.isdscotland.org www.dhsspsni.gov.uk
12	UK unemployment rate	www.statistics.gov.uk	www.statistics.gov.uk
13	Decent home standard	www.odpm.gov.uk	www.communities.gov.uk www.communitiesscotland.gov.uk www.new.wales.gov.uk
14	Population 2011	www.homeoffice.gov.uk	www.gad.gov.uk
15	Immigration statistics	www.homeoffice.gov.uk	www.statistics.gov.uk
16	Household income	www.statistics.gov.uk	www.dwp.gov.uk
17	Incapacity benefit claimants	www.dwp.gov.uk	www.dwp.gov.uk www.statistics.gov.uk
18	Occupational pension schemes	www.dwp.gov.uk	www.statistics.gov.uk
19	Bus passenger statistics	www.scotland.gov.uk	www.dft.gov.uk
20	Number of armed forces	www.census.gov	www.dasa.mod.uk www.statistics.gov.uk

Researchers have personal, tried-and-tested research techniques and, therefore, tend to narrow their search before entering any keywords. They are also more willing than novices to access documents (PDFs or Word) highlighted by Google, although, according to comments made in the focus groups, this is not their preferred way to view data. Researchers also tend to use their knowledge of Internet resources to narrow down their search to sites they assume would contain the answer. For example, they would assume that information about farmers might be found on defra.gov.uk and look for any Google hits from that particular site.

> *The first time I went to the appropriate ministry, used search facility and looked for stats publications but found nothing. Then tried ONS and found easily with the search facility*
>
> Female Researcher, London

Novices, on the other hand, seem to cast their net a bit wider (usually entering the entire question verbatim into a search engine) and are more hesitant to access large PDF documents, looking instead for easy-to-read charts and summative bullet points. As evidenced in the focus groups, they are also less likely to narrow down their search before diving into sites and are happy to let Google do the bulk of the 'thinking'.

> *Put just prison population into Yahoo! with the UK, or Google or whatever, it comes back with much more specific searches. It almost seems to order them in best matches*
>
> Male Novice, Bristol

> *Yeah I'd try it. And then if not I'll just Google. Literally just end up typing the question into Google*
>
> Male Novice, London

When starting from direct.gov.uk, most participants attempted to navigate the site using various methods including simple browsing, the in-site search engine and directories. In most cases, however, both researchers and novices eventually moved to Google and used a search technique as described above. The time spent on direct.gov.uk before moving to a more familiar search engine varied considerably – some participants left almost immediately (within the first two minutes), while others only resorted to Google after 15 minutes or more.

The information journey

The segmentation of the shopper sample into researchers and novices and the allocation of starting point were both done for very specific reasons: we felt that researchers were perhaps better able to find information and that Google would be quicker.

However, while on average, participants spent 14.8 minutes browsing during each of their tasks before 'answering' the question, the difference between researchers and novices was minimal and measured in seconds rather than minutes. Researchers were slightly faster at 14.5 minutes, compared to 15.2 minutes for novices.

Starting points seem to make little difference as well. Those beginning at Google spent 14.3 minutes total browsing, while those beginning at direct.gov.uk spent on average 15.3 minutes. For researchers, their starting point made very little difference, while novices spent nearly two more minutes browsing after starting from direct.gov.uk.

Starting page	Average browsing time (in minutes)	Researcher	Novice	Number of page views	Page views/ minute	Seconds/ page view
Google	14.3	14.4	14.3	33.35	2.3	26
Direct.gov.uk	15.3	14.5	16.1	39.49	2.6	23

Examining each starting point in more detail unveils one significant difference in terms of the number of pages viewed (or clicks). Google clicks ranged from two to 80 pages, while direct.gov.uk clicks ranged from two to only 40.

On average, though, participants beginning at Google viewed 33 pages over the course of their search and spent 26 seconds examining each page. Those starting at direct.gov.uk viewed just under 40 pages (39.49 pages) on average and spent approximately 23 seconds examining each page.

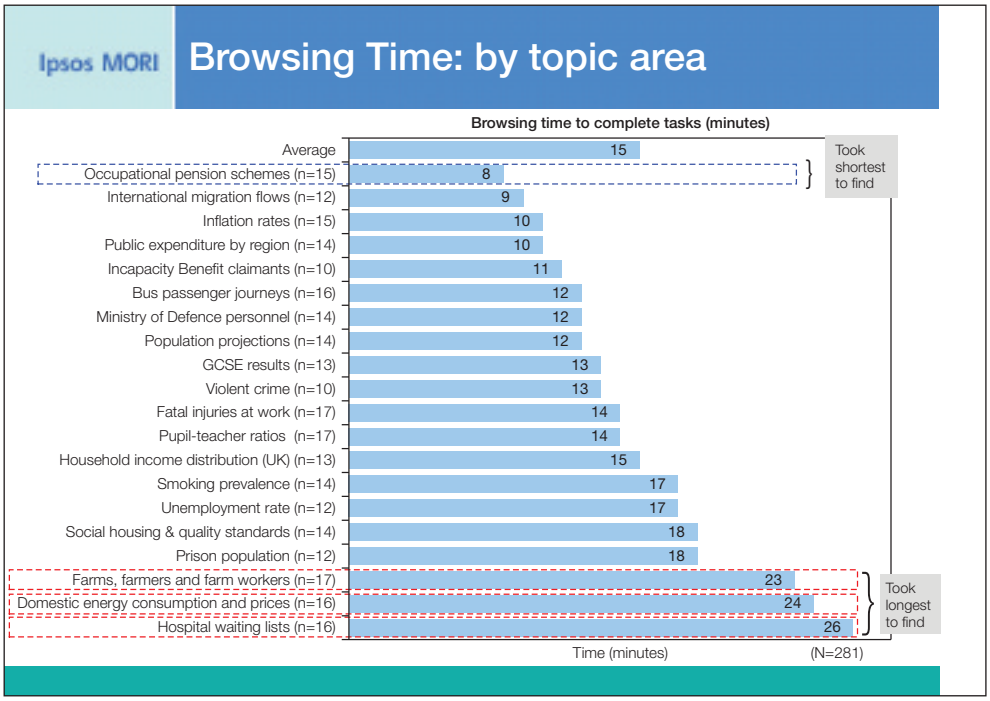
The average browsing time of approximately 15 minutes, however, disguises a wide range of search times. Depending on the topic area, participants took anywhere from eight to 26 minutes to find the data they were searching for. Those data taking the longest to locate (over 20 minutes each) include farms, farmers and farm workers; domestic energy consumption and prices; and hospital waiting lists. International migration flows and occupational pension schemes, on the other hand, each took less than ten minutes to locate.

> *I thought I'm going to give myself an hour to complete the task and I thought I would do it because I use Google every single day at the moment … But I found it really hard actually because the information I thought I could get on the website I couldn't*
>
> Female Novice, London

> *I found the answer really quickly but prior to that question, it [the site] was just taking you around in a loop and I was trying the advanced search engine and everything and it was just getting really, really frustrating*
>
> Male Novice, Edinburgh

I think one of my tasks was to find the number of farmers in the UK … I thought, oh God, this is going to take me two seconds… Everyone in the office was laughing at me because I couldn't find it and I just couldn't find the statistic. It took me about 25 minutes, half an hour

Male Researcher, Bristol



The questions were assigned in a randomised way in order to reduce any potential learning – the assumption being that participants would learn where reliable sources were and become quicker at finding the data. This was an unnecessary precaution as neither researchers nor novices seemed to evolve searching strategies throughout their five tasks, and returned instead to the methods described above for each of their five questions. Researchers remained process-driven, relying on their own techniques and favourite resources, and novices' enthusiasm varied depending on how interesting they found the topic.

Only a couple of participants seem to have adapted their search strategies. One novice reported that he found the exercise quite easy after discovering statistics.gov.uk during his first task and returning to it for each subsequent search, while another said that she only began to find information when she started putting herself in the shoes of the person who designed the site. This reinforces one of the messages picked up in the interviews conducted for the policy review – that the sites often appear to be designed for the convenience of the statistician. As one policy review interviewee noted, data dissemination is currently built upon statistical releases, which "almost drives them into thinking in those terms".

Understanding statistics

Accessibility involves a two-pronged approach to data dissemination. Once an individual finds a data source they believe to contain the information they are looking for, it is just as important that they understand what they are looking at. When asked immediately after completing each of their tasks to what extent they could understand the information they found, three in five participants (59%) felt the information was very or fairly easy to understand.

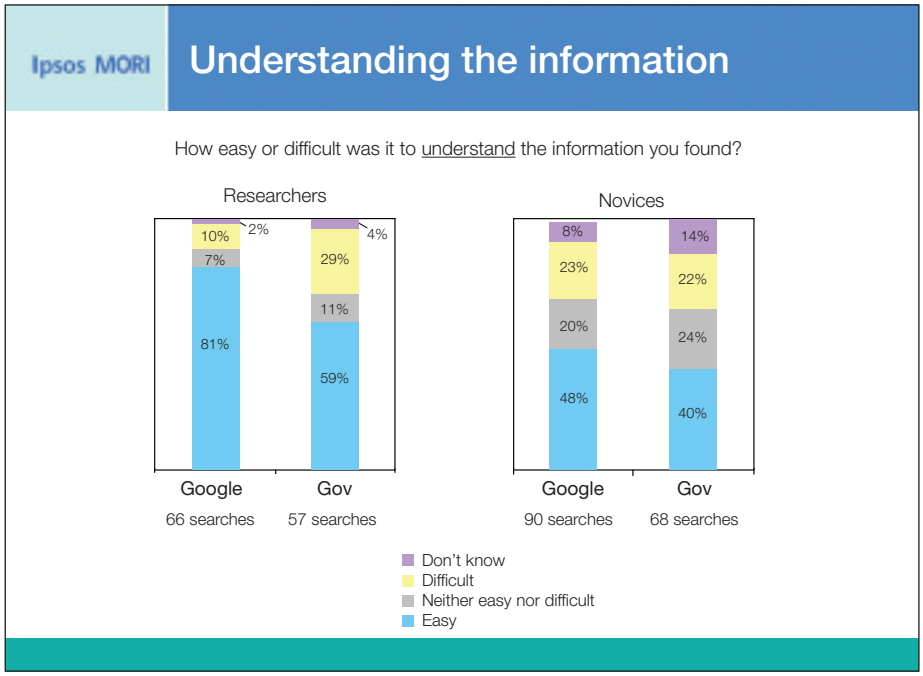
Unlike the similarity between researchers and novices when it comes to finding the information, researchers were far more likely to understand the information. Bred by their familiarity with government data formats, nearly three-quarters of researchers (71%) understood the information they found. Fewer than half of novices (47%) report that the information they found was very or fairly easy to understand.

That being said, researchers who began at Google were far more likely than those starting at direct.gov.uk to feel the information was easy to understand (81% compared to 59%). The starting point, though, made very little difference to novices (48% of those starting at Google found the information easy to understand compared to 40% of those starting at direct.gov.uk).

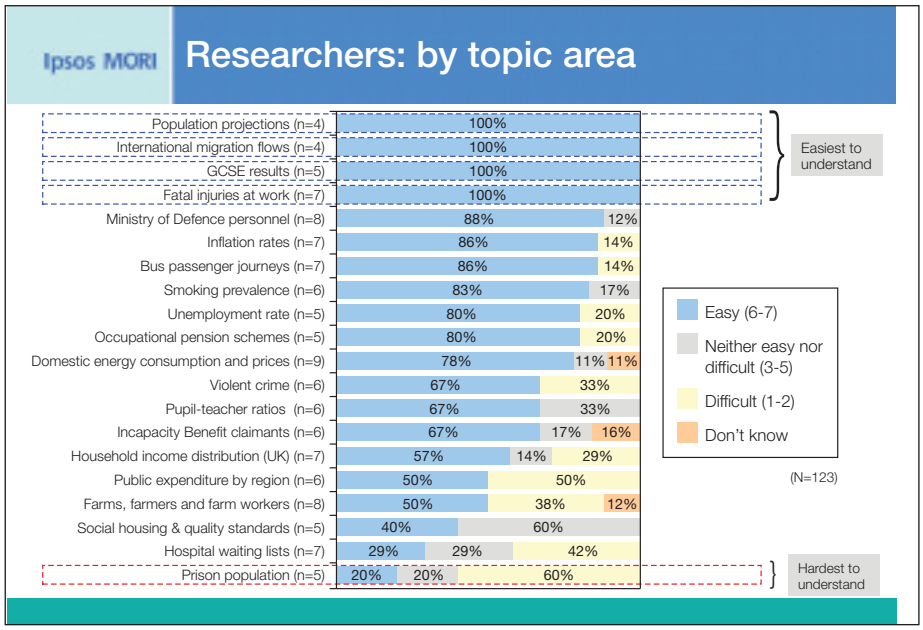
> *I thought there was a lot of jargon at times, and that put me off quite a lot actually. I think that my will to actually carry on was hampered by the jargon*
>
> Male Novice, Edinburgh

> *A lot of it did seem to be directed to people that actually understood it, like accountants, and the general public who've not got much experience of accounting or who really don't know, might not have understood some of the wording*
>
> Female Novice, Edinburgh

> *The main difficulty was the poor relevance of many suggested pages listed by in-house search engines – a lot were often way out of date and Google often performed better*
>
> Male Researcher, London



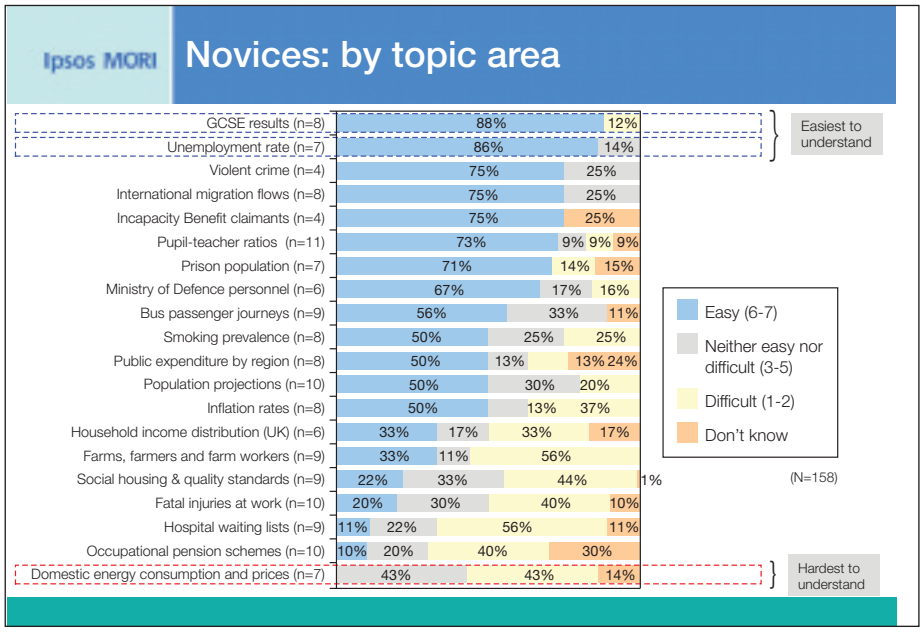
As with finding the information, the level of understanding varied depending on the question. Over half of researchers found the data easy to understand in 17 out of 20 questions. Every researcher looking for population projections, international migration flows, GCSE results and fatal injuries at work found the information very or fairly easy to understand. Statistics they found the most difficult to understand involved data sets on social housing quality standards, hospital waiting lists and prison population statistics.



Conversely, novices were less likely to find the information they came across easy to understand. While nearly nine in ten found GCSE results (88%) and unemployment data (86%) easy to understand, no novices understood domestic energy consumption and prices data.

> *I got one about serious injury, fatal injury in the construction industry and there it gave you three categories and the last category just summarised the first two categories, but I didn't quite understand that until I'd worked out the percentages and things. Somebody looking at it as a statistician no doubt would've said, oh that's obvious, but to me it wasn't*
>
> Female Novice, Leeds

> *'The DEFRA site … I found it a bit confusing because there were quite a few different ways to find the information … You could go into the database and put in your own criteria and that would just bring you up a little table. Or you could go into publications and find something and pull out information from there and then I think they had either a PDF or Excel tables of the same information*
>
> Female Novice, Leeds



Novices also expressed support for graphical and visual data as it helped them understand the information without the need for reading large volumes of explanatory text.

> *It's much easier with a picture. Simple pie chart would have reduced that much text and just made easier to understand, easy to look at*
>
> Male Novice, Leeds

Trusting the information

The issue of trust in primary sources highlights a difference in approach between the two participant groups. Researchers saw the reliability of the primary source as a key aspect within their search, and efforts were focussed on identifying the key information from reputable sources. Novices also wanted to use reliable primary sources, but were more trusting in links provided by the Google search. The National Statistics 'brand' was not mentioned by any participants as a guarantor of trust.

> *National statistics and Government website would have been the only ones I would have trusted the data from as Google links give to many other sites whose reliability wasn't certain*
>
> Male Researcher, London

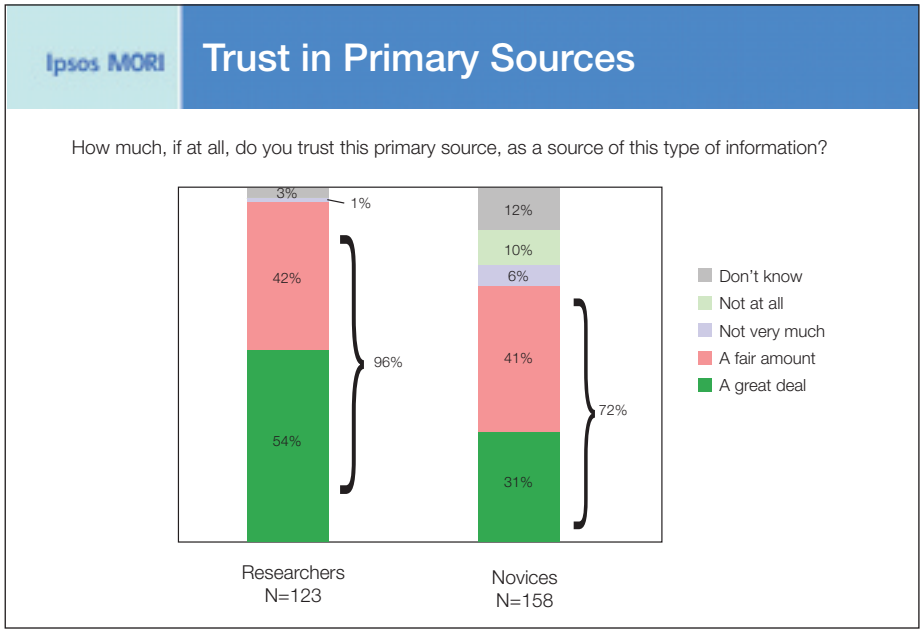
> *Any time I see .gov.uk then I automatically assume rightly or wrongly that I'm going to get something official*
>
> Female Researcher, Edinburgh

> *Google's great for finding the relevant sites and you scan and can still search for the buzzwords and then you look at the link and it looks like a fairly trusted site*
>
> Male Novice, Edinburgh

To enhance the quality of the information, researchers were also more likely than novices to 'double check' their statistics to confirm their validity. Many researchers would identify the same statistic from a number of sources before being confident that the answer was correct.

> *What I would tend to do would be if I find statistics on there I would then cross reference them with statistics that I could get from the National Health Service or from General Household Survey just to make sure that they're OK*
>
> Female Researcher, Edinburgh

The overall result is that researchers have a considerably greater degree of trust in the answers they found compared to novices (54% and 31%, respectively, trusted their primary source a great deal).



For researchers, this comes from a combination of familiarity with the types of sources they use on a daily basis as part of their job on and comfort with the process of finding official statistics online. They are also more likely to assume that government sources are trustworthy and, therefore, deliberately search them out.

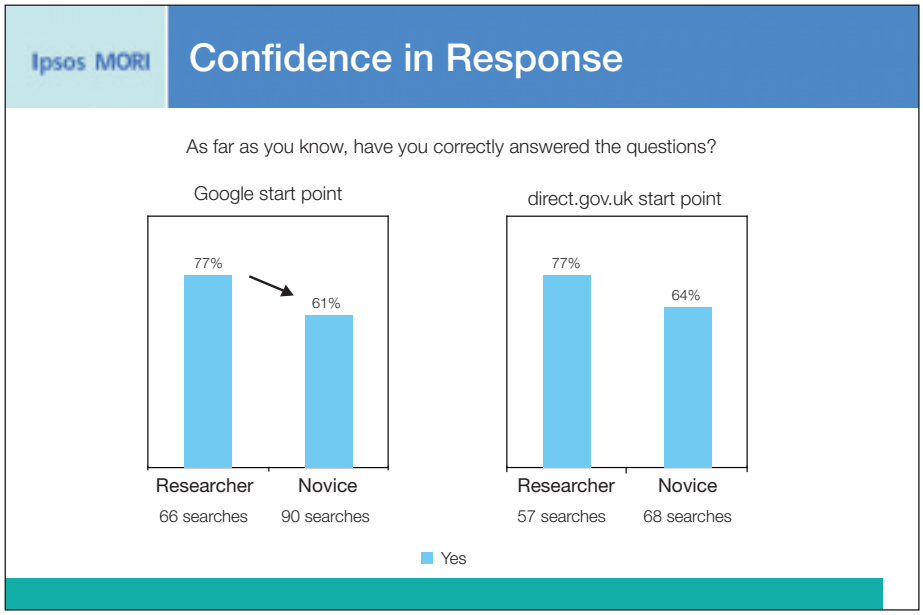
> *If I knew it was an ONS or Government a lot of the trust would go into it … So I suppose you'd look for something familiar or, and trustworthy. I would trust the Government sites*
>
> Female Researcher, Leeds

In fact, when asked to what extent they trusted their primary sources, researchers are most likely to trust their sources for inflation rates (86% trust a great deal) and public expenditures (83% trust a great deal). In both these cases, researchers listed government sites as their primary source (statistics.gov.uk and hm-treasury.gov.uk).

Novices, on the other hand, are more likely to trust a wider range of sources, including media sites (the BBC primarily), government sites and NGOs. When asked about trust in their primary sources, novices are most likely to trust sites where they found MoD personnel data (83% trust a great deal) and smoking prevalence data (75% trust a great deal).

> *The section on migration … was set up in my mind quite logically, in that it had things like pie charts, because not everybody necessarily likes to look down a table of statistics, sometimes people learn in a very, or understand information in a very visual sense … So that seemed logical to me*
>
> Female Novice, London

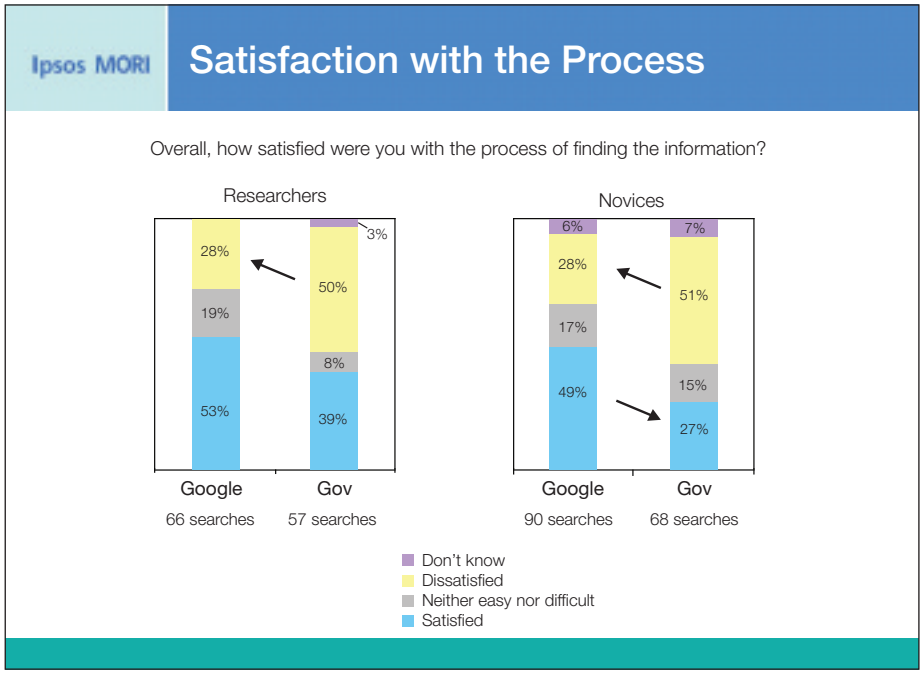
Overall, 70% of participants felt they had correctly answered the question. Not surprisingly, researchers are more confident in their response, with more than three-quarters (77%) reporting that they had found the right answer. Novices were less sure, with 62% reporting the same. Starting points had little impact on these levels of confidence.



Overall satisfaction

Although in general, there are few difference between starting points, time taken, pages viewed and even between researchers and novices (with the exception of trust), when we turn to overall satisfaction with the experience, starting points come through as highly important.

Overall, fewer than half of participants (46%) were satisfied with the process of finding the information. This was consistent between researchers and novices (48% and 43% were satisfied respectively). Both researchers and novices, however, were far more satisfied when starting from Google than when they started from direct.gov.uk.



Google search, type in question get answer, type in question get answer. I'm full in love with that Google because it is the way that I know how to use a computer

Female Novice, London

I just kept doing Google searches and refining my search terms until that put me onto a website which gave me more specific statistics that I needed. I found that if you search within something like the ONS site that the search is so pathetic

Male Researcher, Bristol

At first glance, it would seem that satisfaction is reliant on how easy the information was to find, including the time taken to complete each task, rather than a clear understanding the data or finding it on a trusted website.

This is confounded somewhat when examined by participant type. For researchers, it seems that if they found the data easily, understood it when they found it and found it on a trusted website, they are generally satisfied with the process – in other words, they tend to be process-and-results driven. This relationship is tempered, however, by their initial expectations: their familiarity with the general task (searching for official data online) raised expectations of finding it easily on an official site and understanding it.

For novices, on the other hand, trusting the source of the information does not seem to influence their overall satisfaction with the process. Novices tend to be more topic-driven (i.e. they are motivated to find information about topics they find interesting) but, equally, they can find the process of finding the information frustrating regardless of the topic. For example, violent crime data was relatively easy to find and most thought that it was easy to understand, but only a quarter of novices trusted the source of the information a great deal. And yet, three-quarters of them were satisfied with the process of finding that particular information.

SUMMING UP

Overall, the mystery shopping and focus groups serve as an active demonstration of many of the issues raised by the policy review. Namely, that official statistics are difficult to find online without prior knowledge and that official government websites do not display information in an intuitive, user-centred manner.

Both researchers and novices felt that the following would greatly improve their ability to find, understand and interpret official statistics online:

- Better in-site search engines

  I would trust the Government sites, but when that whole list of stuff came up, like I said when I put a word in and you get all these things that come up it doesn't look official, the list doesn't look official, it just looks like a random generated list of things that might mention the word

  Female Researcher, Leeds

- More intuitive and conventional labelling

  I think the problem, the thing with government web sites is the fact that they are now organised very thematically and they use language which sounds great on the stump or in a magazine or a sound bite. 'Building communities'. Sounds great. Now is that environment? What does 'building communities' mean?

  Male Researcher, London

- Data presented in tables and graphically rather than in text or large PDFs

  Some of them were impossible to understand as you had these PDFs and they were all attached to a bit of statistics and that was confusing

  Male Novice, London

- Detailed directions on where and how to find specific datasets (i.e. 'how to use this site')

> *I think that whole front end of your website is really important, the first thing that people see because like you say they're designed by professionals for professionals and a lot of people they look at it and they're scared of it*
>
> Female Researcher, Edinburgh

More generally, the areas identified both through the primary research and through the policy review broadly follow the 'Top Ten Mistakes in Web Design' highlighted by the web usability guru, Jakob Nielsen:[66]

> *1. Bad search*
>
> *2. PDF files for online reading*
>
> *3. Not changing the color of visited links*
>
> *4. Non-scannable text*
>
> *5. Fixed font size*
>
> *6. Page titles with low search engine visibility*
>
> *7. Anything that looks like an advertisement*
>
> *8. Violating design conventions*
>
> *9. Opening new browser windows*
>
> *10. Not answering users' questions*

If there is one 'quick fix' recommendation from this work, it is that general public users want a clearer and perhaps more hierarchical approach to the presentation of official statistics online. They want to be able to access 'topline' figures from which they can drill down into the detail if they so wish. At the moment, however, that 'frontend' rarely exists (the Scottish Executive's website being a notable exception, with High Level Summary Statistics providing that 'headline' with links to further information).

> *Have two sites for national statistics. One for folk who're used to sifting, somebody who's like you, who is doing a research paper or whatever and needs things to be that specific. And then one which is giving you ballpark figures on what's going on*
>
> Male Novice, Edinburgh

Control and ownership is also another key issue. Firstly, there is the question of statistics units working with marketing, communications and design professionals to improve the online user experience, without compromising the integrity of the statistics. This will, of necessity, mean that ownership of the *presentation* of statistics – though not of the *production* of statistics – will need to be shared.

The issue of control was also showcased by the primary research, though from a slightly different perspective. Many focus group participants complained that official statistics were presented in an extremely user-unfriendly way, often in text-based PDFs. They also felt that sites were designed more for the convenience of the people who controlled the data than for that of the end user.

> *When sites like the [statistics.gov.uk] site are developed they're developed by professionals for professionals and especially professionals who have a wide range of statistical knowledge and sometimes that's the difficult thing whereas the Council website it's really an idiot's guide to the website. You get in there and it's whatever words you put in the search engine will find the course through to whatever you need. Whereas I think with this one there's, because there's different interpretations on different words and different emphasis then it's slightly harder to do that*
>
> Female Researcher, Edinburgh

> *Some of the more technical data sets … do seem to be written by statisticians for statisticians. And they are quite technical and I think Joe Public would have real problems*
>
> Female Researcher, London

Equally relevant to the question of control, however, is the fact that the Internet allows for a wholly new way of marshalling, presenting and accessing data. But it would appear that the people who control that data are still applying a paper-based vision to the Internet. The challenge of the Internet is that it requires a loosening of that control – a loosening which may be perceived as being in conflict with a rigorous approach to statistics. However, we believe that the 'brave new world' of the Internet can still offer up opportunities to present statistics in an online environment which does not compromise data reliability.

Appendices

Mystery Shopping Topics

This is a list of topic questions for use in the "mystery shopper" research. Questions are grouped by National Statistics theme, with at least one question per theme.

Agriculture, Fishing and Forestry

1. ***Farms, farmers and farm workers (England).***

   How many farms (holdings) are there currently in England? How many farmers are there? How many farm workers are there?

Commerce, Energy and Industry

2. ***Domestic energy consumption and prices (UK).***

   How has domestic consumption of the main fuels (gas, electricity, oil and coal) changed over the last 5 years? How have domestic prices for the main fuels moved over that period?

3. ***Fatal injuries at work (GB).***

   How many fatal injuries have there been in the construction industry in Great Britain over the last 10 years? How has the number of injuries changed relative to the number of employees?

Crime and Justice

4. ***Violent crime (England and Wales, Scotland, Northern Ireland).***

   What has been the trend in violent crime in England and Wales over the past 5 years? Find comparable figures for Scotland and for Northern Ireland.

5. ***Prison population (England and Wales, Scotland, Northern Ireland).***

   What has happened to the prison population of England and Wales the past 5 years? Find comparable figures for Scotland and Northern Ireland.

Economy

6. ***Public expenditure by region.***

   How does the level of public expenditure in Scotland compare with the level in the South West region of England? What does the comparison look like in terms of public spending per head?

7. ***Inflation rates (UK).***

   What is the current rate of inflation (consumer prices)? How have prices moved over the past 12 months for (a) food; (b) clothing; (c) new cars ?

Education and Training

8. ***GCSE results (England, Wales, Scotland).***

   What percentage of children in England have gained five or more GCSEs by the end of 'Key Stage 4? (Figures for latest available year). What percentage have gained five or more GCSEs at grades A* to C? What are the equivalent percentages in Wales and in Scotland? [NB. Scottish equivalents: SCGF level 5 or better for GCSE A*-C]

9. ***Pupil-teacher ratios (England, Wales, Scotland, Northern Ireland).***

   How many pupils are there per qualified teacher (primary and secondary schools) in England, in Wales, in Scotland and in Northern Ireland?

Health and Care

10. ***Smoking prevalence (Great Britain, Northern Ireland).***

    How many smokers are there in Great Britain, as a percentage of the population (16 and over)? How many in Northern Ireland? How has the number of people smoking changed over the past 10 years?

11. ***Hospital waiting lists (England, Wales, Scotland, Northern Ireland).***

    Over the past year, what proportion of the people admitted to hospital as in-patients or treated as day surgery cases had been on a waiting list for treatment for three months or more? Find figures for each UK country.

Labour Market

12. ***Unemployment rate (UK and regions).***

    What is the current unemployment rate for the UK? What is the rate for Scotland? For Northern Ireland? For the South West region of England? How has unemployment changed over the past 5 years?

Natural and Built Environment

13. ***Social housing meeting housing quality standards (England, Scotland, Wales).***

    What proportion of social or council housing in England currently meets the "decent homes standard" (as defined by the Department for Communities and Local Government)? What are the equivalent proportions for Scotland and Wales? (based on Scottish and Welsh Housing Quality Standards) How do these figures compare with 10 years earlier?

Population and Migration

14. ***Population projections (UK).***

    What is the UK population projected to be in 2011? How many of that population will be children aged under 16? Can you find projections or forecasts for the number of children aged under 1 in the UK for each year up to 2015?

15. ***International migration flows (UK).***

    What are the latest figures for international migration into and out of the UK? Give separate figures for immigration, emigration and net migration. What have been the trends in migration over the past 5 years?

Social and Welfare

16. ***Household income distribution (UK).***

    What was average household income (£ per week) for the bottom 20% of households (ranked by income), (a) measured before housing costs, and (b) measured after housing costs? What were the equivalent figures for the top 20% of households? How has household income changed over the previous 10 years, for both groups and on both measures?

17. ***Incapacity Benefit claimants (England, Wales, Scotland, Northern Ireland).***

    How many Incapacity Benefit claimants are there in each UK country? How has the number of claimants changed over the past 5 years?

18. ***Occupational pension schemes (UK).***

    How many occupational pension schemes (both private and public sector) are there in the UK that are open to new members? How many active members do these schemes have?

Transport, Travel and Tourism

19. ***Bus passenger journeys (GB).***

    How many bus passenger journeys were there in Great Britain in the past year? How many of these were in London? What has been the trend in bus passenger journeys (GB) over the past five years? Has the trend differed much between different parts of the country?

Other National Statistics

20. ***Ministry of Defence personnel – armed forces and civilians (UK).***

    How many people are there currently serving in the UK armed forces? What is the strength or size of the Army? Of the Royal Navy? Of the RAF? How many civilians work for the Ministry of Defence and its agencies?

Core Questionnaire

Question #1 *required*

As far as you know, have you correctly answered the questions?

- Yes
- No
- Don't know

Question #2 *required* Question

Where did you find the answers? (Please list up to 3).

{Text Response}

Question #3 *required* Question

Which website was your primary source of information?

{Text Response}

Question #4 *required* Question

How much, if at all, do you trust this primary source, as a source of this type of information?

- A great deal
- A fair amount
- Not very much
- Not at all
- Don't know

Question #5 *required* Question

In general, how easy or difficult was it to find the information you needed?

- Very easy
- Fairly easy
- Neither easy nor difficult
- Fairly difficult

- Very difficult
- Don't know

Question #6 *required* Question

What if anything, would have made it easier to find?

{Text Response}

Question #7 *required* Question

How easy or difficult was it to understand the information you found?

- Very easy
- Fairly easy
- Neither easy nor difficult
- Fairly difficult
- Very difficult
- Don't know

Question #8 *required* Question

What, if anything, would have made it easier to understand?

{Text Response}

Question #9 *required* Question

Overall, how satisfied were you with the process of finding the information?

- Very satisfied
- Fairly satisfied
- Neither satisfied nor dissatisfied
- Fairly dissatisfied
- Very dissatisfied
- Don't know

Annex 3

Literature review for statistical accessibility

Report by the Statistics Commission

INTRODUCTION AND BACKGROUND

1. This Annex reports a literature review carried out by the Statistics Commission Secretariat to identify the main accessibility issues. The purpose of the review was to inform the development or our initial pilot research, to assist in the interpretation of our research results and to inform our recommendations. The scope was government websites, mainly statistical websites, and mainly relating to UK statistics but also including relevant research carried out in New Zealand, Canada and the US, within the past decade.[1]

2. The literature reviewed falls into four broad areas, with varying quantities on each area:

   - government policies (about statistics, government websites in general, and the publication of information);
   - research studies into people's online behaviour;
   - research into finding official statistical information; and
   - information about the technology and conducting user testing of websites in general.

3. This report sets out the main themes from the literature review. It starts by presenting what the published **standards** are for accessibility by looking at the policies for statistical dissemination, government websites and for publication of information. This is followed by a look at the **methods used for evaluating** these standards, both qualitative and quantitative. From the research studies and commentary we present the main **barriers to accessibility** that have been identified. We also review the literature about **users** and user behaviour to provide some context for the barriers to accessibility. The final section presents a review of the information relating to **technological developments** and the impact this has on our expectations of online statistics. We conclude by reviewing what we have learnt from the literature review and what future research might focus on.

4. This review was carried out during 2006-07 and presents the themes from the literature published mainly in the previous ten years. It should be borne in mind that government departments are constantly updating and developing their websites and many recommendations made in these studies will have already been implemented, or the recommendations already superseded by technological advances. The Office for National Statistics (ONS), for example, is planning to launch its redesigned website in 2008 as part of its modernisation strategy.[2]

5. There are distinctions made between web accessibility and statistical accessibility. The former refers to the usability of websites, such as whether they are easy to navigate, or whether they have been adapted for people with disabilities, while we use the term statistical accessibility to refer to the availability of official statistics online, how easy they are to find, understand and use.

WHAT ARE THE STANDARDS FOR ACCESSIBILITY?

6. Before reviewing the accessibility research, we look first at the criteria used for assessing accessibility. Standards for accessibility of statistics fall broadly into three types – those setting out statistical standards or codes of conduct; website standards and e-government strategies; and those setting out standards relating to official information generally.

Statistical standards and accessibility

7. Statistical standards are set out for both policy and practice in the National Statistics Code of Practice and protocols (see Annex 1) and are reviewed in Ipsos MORI's research report (see Annex 2). This literature review does not attempt to cover the same material. The literature search identified some guidance from the United Nations on how to present and disseminate information including on the Internet[3] and tips for writing statistical stories for publication on the Web and evaluating their impact[4] such as using simple language, using graphs, table formats, using maps appropriately and including explanatory information.

Website standards and accessibility

8. An important standard in the literature is for a user-centred approach to website design. There is a range of documentation from government – setting out website policy (the Transformational Government Strategy[5]), commenting on its progress (eg by APPSI), or standards for website design. For example, there are plans to rationalise the number of government websites, on the basis that users prefer to use a small number of "supersites" that bring a wide range of information together in one place.[6] In line with these trends, the Government recently announced its intention to make all online government services accessible via a central hub, such as DirectGov or Business Links[7]. Another recent proposal has been the development of a statistical publication hub that would separate the release of statistics from political comment on the data.[8]

9. Another standard is for transactional websites to cater for the shifting relationship between government and users. The Varney report on the opportunities for transforming the delivery of public services noted that "today's consumer is no longer the passive recipient of government services" and that technological changes "create an increasing expectation that access to services on the Web will be comprehensive, joined-up and capable of delivering a service almost instantly".[9]

10. Much of the website literature found relates to international and local technical standards for ensuring that websites are appropriate for all users including those with disabilities. Some of these standards are relevant to statistical accessibility, such as using the simplest language possible, or the need to identify row and column headers in data tables.[10]

11. In 2003, the Cabinet Office published its standards for good website design.[11] This included a definition of a good website, how to predict user satisfaction, and advice and guidance for designing websites (eg consistent navigation, a site map, and a good search facility) as well as advice to employ a professional agency to design the website. Aimed at government websites generally, the advice and guidance is also pertinent to statistical websites. The National Audit Office used these guidelines in carrying out an audit of government websites for their web accessibility.[12]

12. Website design guidance relating specifically to statistical websites includes advice to name the data appropriately, make sure data can be downloaded or copied into different formats, include metadata and have a good search facility. The author points out that "often the web presentation has not been developed by the paper publication team and has suffered from lost knowledge".[13] The UN guidance (mentioned above) includes similar advice.

Information standards and accessibility

13. The Freedom of Information Act (2000) is also helping to generate a culture of making government information publicly available, and is raising expectations about the type of information that should be made available. Government must be able to respond to these requests efficiently and has set up the Information Asset Register to hold information about what data is available. Users do not need to know which departments hold the information they seek, as a search is made across the whole of government.[14]

HOW IS ACCESSIBILITY EVALUATED OR TESTED AGAINST THESE STANDARDS?

14. The review of the literature revealed a number of different approaches to evaluating the accessibility or usability of statistical websites. These ranged from an audit approach of what the website covers; qualitative research into users' perceptions, expectations, and experience; to analysis of website "hits" or usage and analysis of the users' path using tracking software. Some studies used a combination of approaches.

Audit or inspection approach

15. Most of the website design standards have been evaluated using an audit approach, undertaken either manually, automated or a combination of the two.[15] Others have used an audit approach combined with user experience: for example 'Surfing with Ed',[16] a monthly series of reviews by a UK official statistician that has looked at a different country's statistical website each month for a number of years. In it, he looks at aspects such as navigation, user orientation, features or metadata.

Qualitative research methods

16. Qualitative research methods are used to assess the user's experience of the website. Much of the guidance relates to usability testing which tends to be small scale and unsystematic compared to social research methods. Also, its purpose tends to be formative rather than summative – the results of each test or set of tests contribute to the development of the website that is often then assessed by the next tester, and so on.

17. Guidance for usability testing, such as that written by Nielsen, covers advice about people doing the testing, the setting for the test, tasks, and measurement.[17,19] Some guidance is also given about other research methods than could be used to evaluate website design, such as observation, questionnaires and interviews, focus groups, computer logs, and user feedback.

18. Much guidance is based on or refers to Nielsen's work. The Cabinet Office's guidance on website design,[11] for example, contains advice on user testing including who to use as testers and how to treat them, how to measure the testing session, and the importance of knowing who the regular users are: their level of Internet experience, interest in subject matter, and demographic characteristics. The ONS is putting this into practice, and developing user

personas to ensure that the needs of a range of users are met.[2] Other researchers have recommended researching how the user achieves their goals and to include evaluation of the search engine as well as navigation through menus.[18]

Analysis of computer log files

19. Another way of measuring the user experience is to collect quantitative data while they carry out online tasks. Technology has developed, but basic methods to record the user's progress through the task have remained the same. Written or manual methods, such as self-reported logs or screen snapshots, have been replaced with web tracking software. However, similar items are measured, such as the number of keystrokes, the time taken and the path taken through the Internet to complete the task.[19]

20. Another measure is to analyse the links between websites using "webmetric" techniques. This counts the number of links between websites to find the most popular, and has been used in a study to compare online networks with social ones.[20]

WHAT ARE THE MAIN BARRIERS TO ACCESSIBILITY?

21. A good deal of the literature on accessibility research concludes that the websites they evaluated were not designed with the user in mind and identifies specific barriers and makes recommendations for improvements. The barriers were user guidance, website design and navigation, geography, search tools, data formats, and analysis tools, and these are described in more detail below.

Online help and user guidance

22. Much of the literature describes statistical websites that were insufficiently user-focussed, and did not provide sufficient explanatory information about the data or that it was difficult to find this information. It has been noted that online statistics are likely to present barriers to users already familiar with the paper versions,[21] though most of these barriers can be mitigated by a user-focussed web design.

23. A 2005 study of the US Government website FedStats (the American portal for official statistics) recommended presenting this explanatory information in small amounts – "just-in-time, just-enough" help that is presented when needed – which they suggested might be provided by an interactive statistical glossary with context specific explanation.[22]

24. Commentary about the statistics is an area of concern and the researchers in the study described above argued that there should be "no naked data"; the story behind the numbers is always needed. In addition, the Statistics Commission has previously expressed concerns about the frankness and fullness of the commentary that accompanies the figures.[23]

25. The importance of metadata is recognised – this refers to information about the statistical data that makes it easier to interpret and use, and techniques that help users avoid misusing data.[24] Recommendations for improvements include adding more context or metadata such as definitions, and explanation about the data.[25] This information should be easy to find, preferably located closer to the data.[2] It should also include information about the currency of the website and when each data item was last updated.[26]

26. Use of "scientific" words in presentation of data should be avoided;[27] one commentator has remarked that the data descriptions do not describe the information available very well and that there is a tendency towards using jargon and technical descriptions[21] and misuse of conventions such as alphabetic listings that list include items beginning with "the" under T.[28]

Website design and navigation

27. Reviewers of the literature for e-government[29] found that poor web design was a significant barrier to e-government in the UK; others have criticised the inconsistency of government websites.[21] Another difficulty observed of government websites generally is that websites have been structured around government organisation rather than user need.[28,30]

28. One study of website accessibility,[27] based on FedStats, recommended a general principle of designing a statistical website for novice users. Statistics producers and website managers should not assume that all users regularly accessed the site and as a consequence had learned how to use it, nor should they presume that users should know the structure of organisations or agencies and so easy navigation of the site should be provided.

29. Navigation refers to how easy it is for the user to find their way around the website, for example with menu options or a site map. Navigation is crucial to being able to use a website and some researchers have said that it is one of the key factors in determining e-government readiness and maturity. Website designers should aim to "minimize scrolling and clicking"[22] and avoid simply placing the digital versions of paper publications onto the website. Files in portable document format (pdf) cause particular problems for navigation compared with web formats because the information they contain cannot be indexed as effectively.[21]

30. A portal website is one that acts as an index or gateway to a number of other websites via a set of links to other websites, and/or a search engine. Some of the literature covers research carried out on the FedStats portal, while other studies cover the issue of portals in a theoretical way, as the possible solution to the problems caused by numerous or disparate websites. Some users have described the ONS website as a portal as it contains some links to statistics produced by other departments and administration, though they claimed it "inadequate for the task".[31]

31. The Neighbourhood Statistics website, described as an "online data resource", offers a possible model for a statistics portal. It has been cited as an example of good practice for drawing on different sources of data in various parts of government to provide users with a "one-stop shop".[32]

Geography

32. Of particular concern to commentators and users in the context of statistical accessibility is the decentralised system for UK official statistics, and in particular the impact of devolution. The system has been described as "fragmented", with users having to "cobble together" statistics for the UK as a whole from various departments and administrations.[33] The Statistics Commission previously recommended in relation to education statistics that the four UK administrations should recognise their users, and take a consistent approach to the publication of performance data for individual schools.[34]

33. Where available, statistics for smaller geographies, such as local authority or ward level, are made available on the Neighbourhood Statistics website. However, the lack of availability of local statistics on other websites remains a barrier.

Search function

34. Paragraph 58 below describes how search engines work. The commercial search engines have set high standards and the cry "couldn't it be more like Google" is frequently heard in usability studies when testing a website's own search function[18]. ONS's usability research,[2] for instance, found that Google-style search tools were preferred because they allowed for wider interpretation of search terms.

35. Other researchers have gone further in their recommendations, stating that there should be a facility for users to perform a 'comparative search', enabling, for example, comparison of statistics for two cities; and that advanced search facilities should be available.[27] Others have argued for using technology more efficiently to enable users to access data; this might include adding a *data* option to the Google toolbar.[35]

Data formats

36. A range of data formats is important for accessibility because users comprehend information in different ways – in text, tables, charts, or maps, and want to use it in different ways, such as downloading the data for further analysis. Users prefer statistics to be available in a variety of formats – the ONS presents information in "nuggets" or short statistical news stories on its website,[36] but has been criticised for its presentation of graphs without displaying the underlying data in a table.[21]

37. Documents in pdf format are unpopular with users and commentators, because they are difficult to navigate or transfer into a format for further analysis. From research with its users, ONS has found that these users primarily want the data to analyse themselves, and the commentary is of secondary importance.[2]

38. Statistics are increasingly presented in the form of maps – the literature highlights examples of good practice, eg by the Environment Agency and the Department for Environment, Food and Rural Affairs.[32] The literature review did not find any research into the accessibility of mapped statistical data.

Analysis tools and other advanced barriers to accessibility

39. The online maps mentioned above are examples of the effective use of technology in providing interactive ways of presenting the statistics. Other research has recommended that tools for analysing the data online should be provided and that users should have the option to choose the granularity of geography and time series[27] and web designers should "provide alternative ways to slice and dice datasets".[22] An early study found that sophisticated users want more access to raw data, or if this is not possible, tables that they can manipulate.[25]

40. Other barriers to accessibility are the lack of archived data,[2] restrictions on access to administrative data,[33] and charging for certain data.[37]

THE USERS

Who are the users?

41. Many researchers of the accessibility of official statistics have started from the list of users that the information is intended for; few have begun by finding out who are, or could be, the users. Various lists of categories of users have been produced. [3,38,39] Website usability studies and studies of statistical users have

tended to categorise users in terms of their expertise,[40] how processed is the data they use[41] or the purpose of their use of the data.[42] Others have categorised by topic or type of organisation.[43,44]

42. The development of the Internet has brought changes in the way that people access and use information – not the least being an increasing familiarity with searching methods. We describe in paragraph 54 below that the proportion of the population with access to the Internet is increasing and there is likely to be an increase in the number of non-specialist users accessing statistics. One commentator has described the low rate of statistical "literacy" amongst users as a 'gulf' between statisticians and their users and potential users[45].

43. The main users are not passive recipients of official statistics but can provide feedback to ensure that these statistics are relevant and of good quality. User groups, comprising professional users of official statistics, have asked that they be recognised as customers, with systems put in place to ensure that their needs are "identified, evaluated and implemented".[33]

44. In a review of the literature for e-government, reviewers noted that government is not using analyses of web hits strategically.[29] A forthcoming review of tax data by the Statistics Commission has found that HMRC appears to have little information about who accesses tax statistics on its two websites; the Statistics Commission believes that they need to find out more about their users.

A taxonomy of users' tasks (what users want to do)

45. Based on information supplied by the providers of US government statistics, researchers developed a general taxonomy of user tasks[38] which were later used to develop scenarios for other studies. Tasks were categorised by their purpose – to learn, verify, judge, explore, to plan and forecast, or to be referred to another information provider. The researchers note that the diversity of purposes means that there is no single right way for a website to handle all tasks or queries but a system designed as a referral system (or portal) should link to the relevant website rather than to specific databases within websites.

46. Other dimensions of the taxonomy include the constraints of time (information from a particular time period wanted), volume (a particular amount of information is wanted) and geography (information from a particular geographic area wanted). A semantic dimension described the topic, the level of abstraction and specificity and the number of facets. Comparison was an activity common to all user types as was the need for context for tables, including definitions, data quality such as variance and other metadata.

How we search for information

47. Researchers have found that behaviour in looking for statistics is no different than when searching for other information – so the problems that emerge, such as with word matching or with files in portable document format (pdf), are generally similar.[46] From this research it was recommended that web site developers enhance indexing structures to include users' terminology.

48. Users are strongly guided by the organisation of a website[38] and often turn to search engines to compensate for clumsy web design.[47] Although users do not necessarily understand how the search engines work, they expect to find what they are looking for on the first page of search results, and if their first attempt is unsuccessful, they refine the search terms in an iterative process rather than switching to another search engine[48] or using advanced search tools.[47]

49. Research by ONS[49] into how their users access the website found this was usually via a search engine, or through the site already being on the user's list of "favourites". This reiterates studies by Nielsen that indicate a search engine is most common way of arriving at a website;[11] however this could be partly due to ignorance about typing in URLs directly.[11] Once onto the relevant website, around fifteen to twenty per cent of web users go straight to a site's search facility rather than use the navigation.[11]

50. There is a distinction made between using search engines for "discovery" versus "recovery".[50] A study looking at how scientific researchers use search engines found that a discovery, or exploratory approach tends to be used in researching topics that are newer, or susceptible to rapid changes in domain names, while the recovery approach uses the search engine as an *aide memoire* for locating known sources for topics that the researchers are already familiar with.[20]

51. A commentator has noted that the way we expect to find, access and manipulate data has been influenced by the impact of Google, Amazon and statistical analysis software.[51] The same commentator voices a warning for accessibility that "as users become less and less technically specialist, and the interfaces become easier and easier to use, the risk of incorrect or methodologically flawed analyses increases dramatically."

How we scan information

52. The wide availability of information on the Web and the ease of accessing it via search engines have led to what has been called "information snacking".[47] This refers to the short time that people are prepared to spend scanning a website or web page. It is increasingly rare for people to scan an entire document other

than by using search facilities. The recommended solution is to provide information in easy to read, bite-sized chunks, with the facility to drill down for more detail or explanation as appropriate.

53. UK government guidance on website design relating to user behaviour is to ensure that text is "scannable" with highlighted keywords, beginning the item with the conclusion and keeping it concise and in plain language.[11]

TECHNOLOGICAL DEVELOPMENTS

The Internet

54. The rate of households acquiring Internet access has been rapid – in 2006, 57 per cent an Internet connection, up from 36 per cent in 2001.[52] Currently, nearly four out of five Internet connections are via broadband.[53] The principal reason for households to go online was to find information,[54] and following the increasing use of online shopping websites is the opportunity to expand the use of public service transactional websites.

55. There has been an increase in information sharing websites – a development that the Cabinet Office recently hailed for its effect in "democratising information and driving citizens' appetite for sharing advice and opinions in new ways".[55]

56. Web users have the opportunity to tailor Internet content for their own needs, customising web page content or sharing information about individual websites through the use of social bookmarking websites (ie sites that enable users to share online their favourite sites and to develop virtual information networks). Some website owners, such as newspapers, offer readers the opportunity to propose the page they are reading for inclusion in a hosted social bookmarking site, thus increasing readership of that newspaper's site. This has raised expectations about the ability of websites to make content relevant to individual users and so facilitate finding information, or for websites to offer users the options to personalise the content they view.

The authority of information

57. The Internet raises difficulties about the authority of information. Whilst blogs (web logs) encourage access to a wider source of ideas and comment, they also increase the risks of misinformation. Blogs are used by the authoritative (for example, the ONS blog as part of its consultation on geographies[56]) and the merely opinionated. The online self-managing encyclopaedia Wikipedia is now much bigger than, for example, Encyclopaedia Britannica. The Internet

offers unlimited access to information posted on it – much of it accurate, some of it dishonest, much misleading. Hence the need for trusted sources has seldom, if ever, been stronger.

Search engines

58. Users tend to only look at the first page of search results[48] and so a high ranking in search engine results is an important aspect for accessibility. Search engines work by indexing links with a combination of keyword search and an algorithm that ranks most commonly linked or rated websites – eg Google ranks by the volume of links to the page, Ask ranks by experts' usage of those pages.[57] Much has been written about the quality of the Google algorithm and the effectiveness of its intuitive approach.[58]

59. A related study[20] looked at how easy it was to find science websites and found that search engines varied in quality but all play a major gatekeeping role. The recall capacity of search engines varies and there were differences in the quality and content of results. The researchers concluded that the Web has a particular structure that determines access and the visibility or "presence" or popularity of a topic is important to being able to find it.

CONCLUSIONS

What has the literature review told us?

60. From the literature review we have gleaned sufficient information firstly to inform the design, and secondly in order to contextualise our research findings. It guided us in deciding what variables to control for and influenced us in deciding to split users according to their expertise in searching for statistics, and according to whether they started their search from a search engine or from a government website. It highlighted the most common barriers to accessibility and so informed the development of the questions we used to test all aspects of accessibility. The literature about technological developments and about user behaviour enabled us to set our findings into context and inform our recommendations.

61. The literature informed us that the type of accessibility barriers we found were not unique to UK government statistics. This means that solutions to these barriers might be found elsewhere – Swires-Henessey for example, notes a number of good practice examples in his final review of statistical websites.[16] Maps are very recently being used to display statistical information online, and the lack of literature about the accessibility of mapped statistical data is indicative of the lag between technological developments and evaluative research.

Future research

62. Human-centred design, including usability testing, is at the heart of UK government web design and testing with users will most likely to inform future developments. However, the UN advises keeping abreast of new technology: "technology is rapidly loosening the constraints that used to affect its (statistical offices') activities. Tracking technological advances has become almost as important as conducting studies of user needs and satisfaction."[3]

NOTES AND REFERENCES

[1] Searches were made in the House of Commons and Treasury libraries, publications and current projects by the Statistics Commission, and online searches of government and other organisations' websites, online searches using Google and GoogleScholar, and media websites. The search terms were: *statistics government website access user uk*. The searches were carried out between May 2006 and April 2007.

[2] *Office for National Statistics i-dessmination project*, Simonon S, presentation at Statistics User Forum meeting "*Improving Access to Government Data*" (18 January 2007). http://www.esds.ac.uk/news/eventsdocs/SUFsummary.doc

[3] *Handbook of Statistical Organization, Third Edition*, United Nations, New York (2003). http://unstats.un.org/unsd/publication/SeriesF/SeriesF_88E.pdf

[4] *Making data meaningful*, United Nations Economic Commission for Europe, 2005. http://www.unece.org/stats/documents/writing/guide.pdf

[5] *Transformational Government Strategy*, Cabinet Office, 2005. http://www.cabinetoffice.gov.uk/e-government/strategy/

[6] *Death of the surfer … birth of the supersite,* Cabinet Office press release, 6 March 2007. http://archive.cabinetoffice.gov.uk/e-government/directgov/

[7] *Transformational Government Annual Report 2006*, Chief Information Officer Council, January 2007, http://www.cio.gov.uk/transformational_government/annual_report2006/index.asp

[8] The publication hub was announced during the second reading debate of the Statistics and Registration Service Bill in the House of Commons http://www.publications.parliament.uk/pa/cm200607/cmhansrd/cm070108/debtext/70108-0006.htm

[9] *Service Transformation*, Sir David Varney, HM Treasury, Dec 2006. http://www.hm-treasury.gov.uk/media/53D/F2/pbr06_varney_review.pdf

[10] *WAI Guidelines and Techniques*, Web Accessibility Initiative, accessed 13/4/2007 http://www.w3.org/WAI/

[11] *Quality Framework for UK Government Website Design*, Cabinet Office: Office of the e-Envoy, (2003) http://www.cabinetoffice.gov.uk/e-government/docs/qualityframework/pdf/quality.pdf (Includes citations of work by Consumers' Association and Nielsen).

[12] *Progress in Making eService Accessible to All – Encouraging Use by Older People,* National Audit Office (2003)

[13] *Website Design Guide,* Swires-Hennessy E, *Surfing with Ed* website, accessed April 2007. http://www.surfingwithed.org.uk/web_site_design.htm

[14] *Information Asset Register*, Office of Public Sector Information, website last updated 19/10/2006. www.opsi.gov.uk/iar/index.htm

[15] *eAccessibility of public sector service in the European Union*, Cabinet Office (2005). http://www.cabinetoffice.gov.uk/e-government/docs/eu_accessibility/pdf/eaccessibility(eu)_report.pdf

[16] *Surfing with Ed on the Internet*, Swires-Hennessy E, monthly articles 1998-2006 and Edition 100, December 2006. http://www.surfingwithed.org.uk

[17] *Usability Engineering,* Nielsen, J. (1993), Academic Press, Boston.

[18] *Introduction to Usability for the Information Professional*, Agarwal SJ, conference paper (2003). http://www.ilrt.bris.ac.uk/publications/conf/Online%202003/Online2003_paper.pdf

[19] *The Usability Methods Toolbox,* Hom, J. (1998). http://jthom.best.vwh.net/usability/

[20] *The World Wide Web of Science: Emerging Global Sources of Expertise,* Schroeder R, Caldas A, Virkar S, Dutton WH, Oxford Internet Institute, University of Oxford, March 2007. http://www.oii.ox.ac.uk/research/WWWScience_Finalreport.pdf

[21] *British official statistics on the web: a consideration of the web and the British user of official government statistics*, Allan A, University of Sheffield Library, 2005.

[22] *Finding and understanding government statistical information*, Marchionini G, Haas SW, Zhang J & Elsas J (2005). Computer, 38(12) p52-61, Dec 2005. http://ils.unc.edu/~march/computer_submission.pdf

[23] *Official Statistics: Perceptions and Trust,* Report No 24, Statistics Commission, 2005 http://www.statscom.org.uk/uploads/files/reports/024%20-%20Perceptions%20Trust.pdf

[24] *Public access and use of government statistical information,* Marchionini G and Greene S (1997). http://ils.unc.edu/~march/govinfo_paper.html

[25] *Extending Understanding of Federal Statistics in Tables,* Marchionini G, Hert C, Liddy L, and Shneiderman B (2000). http://ils.unc.edu/~march/CUU/tables.pdf

[26] *Improving Access to Government Datasets,* Allan A, at Statistics User Forum meeting "*Improving Access to Government Data*" 18 Jan 2007, http://www.esds.ac.uk/news/eventsdocs/SUFsummary.doc

[27] *Finding Governmental Statistical Data On The Web: A Case Study Of FedStats*, Ceaparu I, IT&Society, Volume 1, Issue 3, Winter 2003, Pp. 1-17. http://www.stanford.edu/group/siqss/itandsociety/v01i03/v01i03a01.pdf

[28] *e-Government Information: the same old problem … newly digitized!,* Allan A, Intute: Social Sciences Online (2005) http://www.sosig.ac.uk/socsciweek/blog/?p=24

[29] *e-citizen research – literature survey*, MORI on behalf of Norwich City Council (2004). http://www.e-citizen.gov.uk/NP/rr/pdf/Literature_Review_survey.pdf

[30] *Digital divide and quality of electronic service delivery in local government,* Kuk G and Gow I, (2002) http://www.sisr.net/events/docs/0208kuk.pdf

[31] *The Market Research Society (MRS) Submission*, HM Treasury Consultation document: Independence for statistics, HM Treasury (2006). http://www.hm-treasury.gov.uk/media/24B/71/MarketResearchSociety.pdf

[32] *Realising the Value of Public Sector Information, Annual Report 2006*, Advisory Panel on Public Sector Information (APPSI) (2006) http://www.appsi.gov.uk/reports/annual-report-2006.pdf

[33] *Independence for statistics*, Tenth Report of Session 2005–06, House of Commons Treasury Committee (2006), http://www.publications.parliament.uk/pa/cm200506/cmselect/cmtreasy/1111/1111.pdf

[34] *Schools Education Statistics: Users Perspectives,* Report No 26, Statistics Commission (2005) http://www.statscom.org.uk/uploads/files/reports/026_School_Education_Report_User_Perspectives.pdf

[35] *Debunking third-world myths with the best stats you've ever seen*, Rosling H, conference presentation to Technology, Entertainment, Design (TED), (2006). http://www.ted.com/index.php/talks/view/id/92

[36] *Spring 2006 Departmental report,* Office for National Statistics, (May 2006), p17. http://www.statistics.gov.uk/downloads/theme_other/SpringDeptReport2006_v1.pdf

[37] The Guardian newspaper has run a "*Free our data*" campaign since March 2006 for the government to release more data free of charge. http://www.freeourdata.org.uk

[38] *Seeking Statistical Information in Federal Websites: Users, Tasks, Strategies, and Design Recommendations*, Hert CA and Marchionini G (1997) http://ils.unc.edu/~march/blsreport/mainbls.html

[39] *National Statistics Code of Practice Protocol on Customer Service and User Consultation*, ONS (2004). http://www.statistics.gov.uk/about/national_statistics/cop/downloads/customerservice.pdf

[40] *Access to National Statistics on Transport via the Web,* Report No 6, Statistics Commission (2002) http://www.statscom.org.uk/uploads/files/reports/transport%20stats%20via%20web.pdf

[41] *Use Made of Official Statistics*, Report No 33, Statistics Commission (2007) http://www.statscom.org.uk/uploads/files/reports/Report%2033%20Use%20Made%20March%2007.pdf

[42] *Enhancing the Value of Health Statistics: User Perspectives*, Report No 21, Statistics Commission, (2004) http://www.statscom.org.uk/uploads/files/reports/Enhancing_value_health_statistics.pdf

[43] *Statistics User Forum User Groups*, Royal Statistical Society (accessed April 2007) http://www.rss.org.uk/main.asp?page=1612

[44] *Users of official European statistical data: investigating information* needs, Blakemore M and McKeever L, Journal of Librarianship and Information Science (2001), Volume 33, No.2, pp 59-67

[45] *Taking a Horse to Water – and Giving it a Bottle,* Marder D, Presentation to the UN Statistics Commission and Economic Commission for Europe – Conference of European Statisticians (2006). http://www.unece.org/stats/documents/ece/ces/ge.45/2006/wp.17.e.pdf

[46] *What is the future of statistical compendia in the 21st century?* Zawitz MW, Statistical Journal of the UN Economic Commission for Europe; 2005, Vol. 22 Issue 2, p163-171, http://iospress.metapress.com/(gf4eslaxaspabd324pgh0t55)/app/home/contribution.asp?referrer=parent&backto=issue,6,8;journal,3,24;linkingpublicationresults,1:103187,1

[47] McGovern, G and Nielsen, J in *The search lurch*, Buchholz GA, (2005), Prentice hall PTR http://www.phptr.com/articles/article.asp?p=418857&rl=1

[48] *Search Engine User Behaviour Study*, iProspect and Jupiter Research, iProspect, (2006). http://www.iprospect.com/premiumPDFs/WhitePaper_2006_SearchEngineUserBehavior.pdf

[49] *National Statistics Website*, Simonon S, Thorhley T and Wood N, ONS, presentation at the National Statistics Open Day, 6 Sept 2005, http://www.statistics.gov.uk/events/nsopenday2005/agenda.asp

[50] Battelle J, speaking on *Analysis – Searching Questions*, BBC Radio 4, 27 June 2006. http://news.bbc.co.uk/1/hi/programmes/analysis/5222356.stm

[51] *How has the Internet changed the way we access data?,* Wright M, Intute: Social Sciences Online (2005). http://www.intute.ac.uk/socialsciences/archive/esrc_socsciweek/2005/blog.html?p=27

[52] The 2001 percentage of households with Internet connection covers the period Jan-March 2001 – source: *Historical Internet Access – Data (pre-April 2003),* ONS, (accessed April 2007). http://www.statistics.gov.uk/statbase/Product.asp?vlnk=5672; the 2006 figure covers to Jan-April 2006 – source: *Internet access – households and individuals*, Statistical First Release, ONS, (August 2006). http://www.statistics.gov.uk/pdfdir/inta0806.pdf

[53] The rate of broadband connections relates to March 2007 and is taken from *Internet connectivity March 2007*, Statistical First Release, ONS, (May 2007). http://www.statistics.gov.uk/pdfdir/intc0507.pdf

[54] Figures relate to Jan-April 2006 – source: *Use of the Internet,* Omnibus Survey, Office for National Statistics, (2007) http://www.statistics.gov.uk/CCI/nugget.asp?ID=1711&Pos=1&ColRank=2&Rank=224

[55] *Power to the people – Information sharing hailed as a new force for social progress by Government*, Press release 8 February 2007, Cabinet Office. http://www.egovmonitor.com/node/9340

[56] The small area geography consultation blog can be found at: http://www.onsgeography.net/

[57] *The un-Google*, The Economist, (17 June 2006) p73.

[58] See for example *New Tools to Vie with Google*, Financial Times, (30 March 2006) p9.