

Annex 2

Accessibility of official statistics online

Report to the Statistics
Commission by Ipsos MORI

Contents

Introduction	35
Methodology	36
Interpretation of the data	39
The Policy Review	40
Who is the user?	40
The policies	41
Understanding the user	42
Ownership and control	43
A user-centred approach	45
The Primary Research	45
Finding the information	46
The starting points	50
The information journey	52
Understanding statistics	55
Trusting the information	58
Overall satisfaction	60
Summing Up	62
Appendices	65
Mystery Shopping Topics	67
Core Questionnaire	71

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 vocabularies⁵⁸

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 debriefing 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.

©/psos MORI/J28826

*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 effectively⁶³

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/](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.

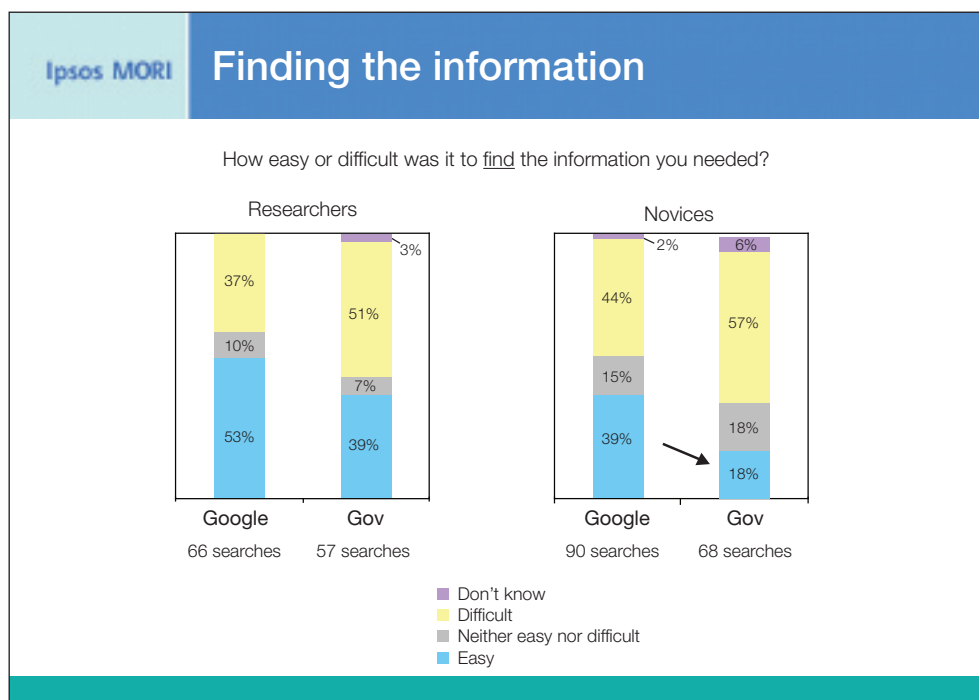
I got a calculator out and calculated it myself because the table never gave me exactly what I was looking for

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.

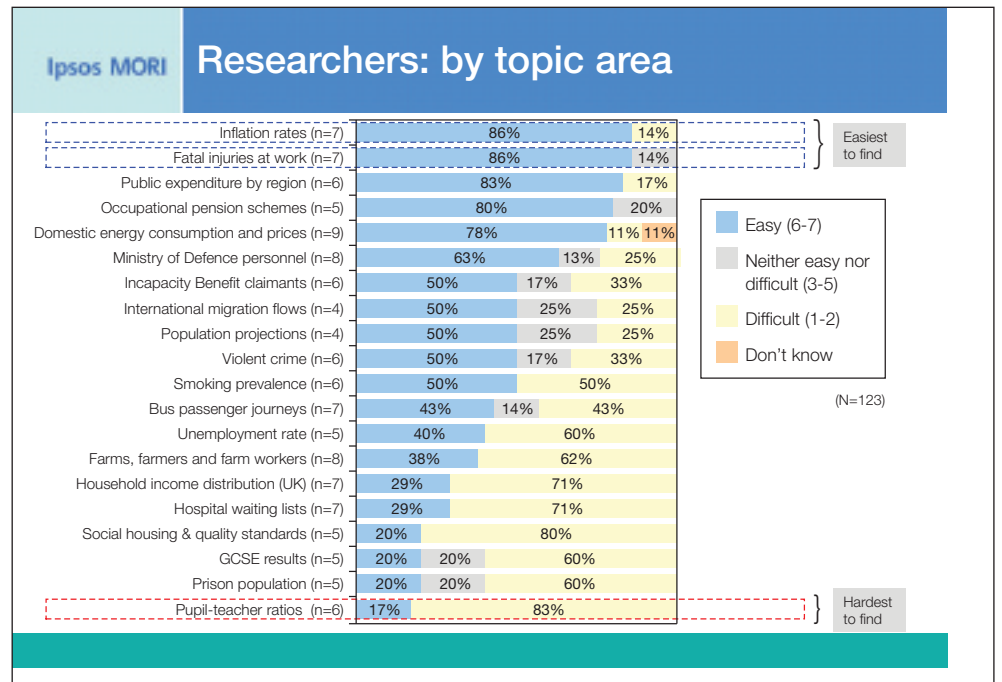
It's quite difficult to get to the source data sometimes because, well, certainly on the National Statistics site, they've put this layer on top to steer you towards an overview of the latest figures, but then frequently that doesn't answer the precise point which you're trying to answer

Male Researcher, Bristol

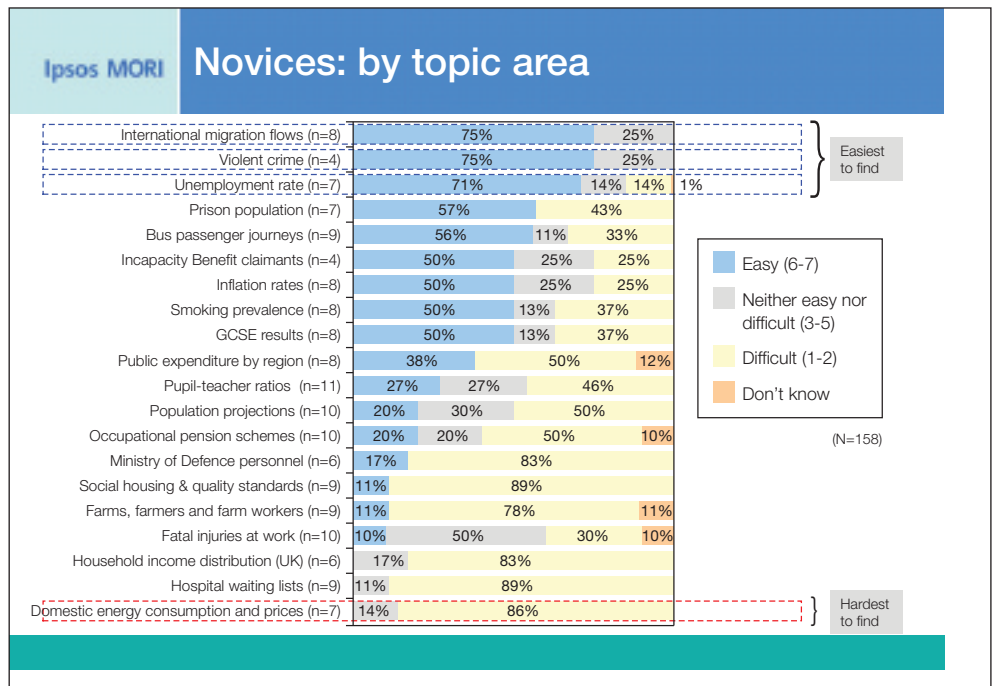
For both researchers and novices, how easy the information was to find depended heavily on which questions they had been assigned. For researchers, the easiest information to find involved inflation rates and fatal injuries at work (86% found both of this information very or fairly easily). On the other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

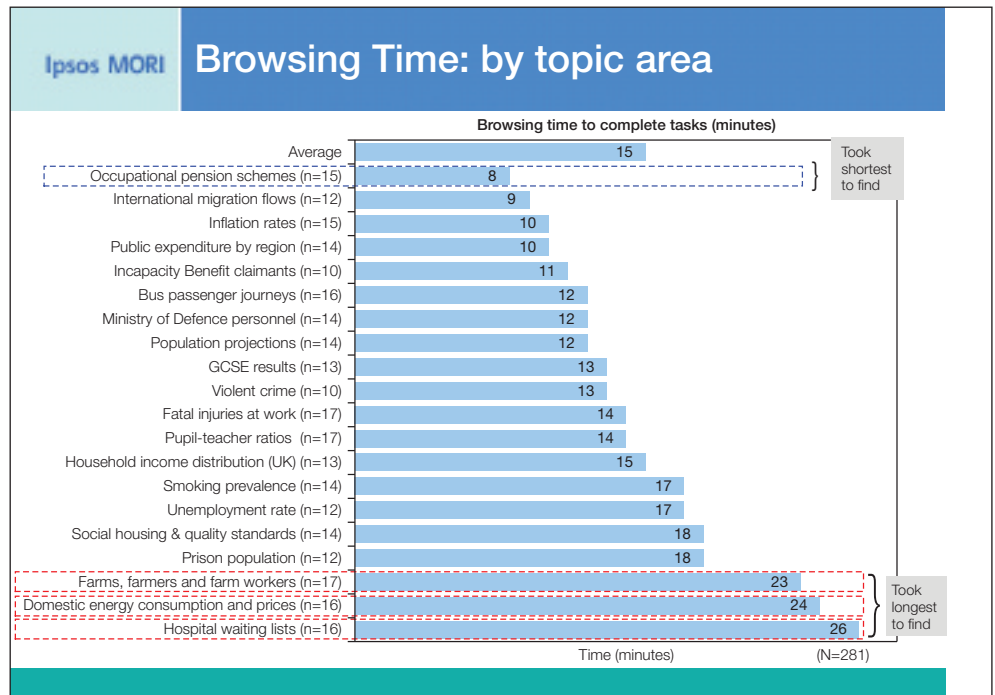
Female Novice, London

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

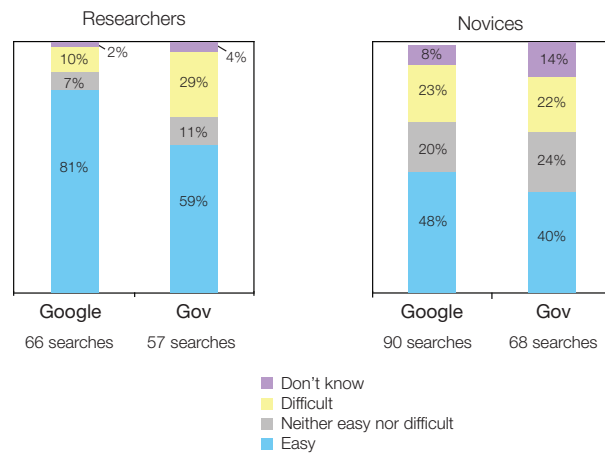
Female Novice, Edinburgh

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

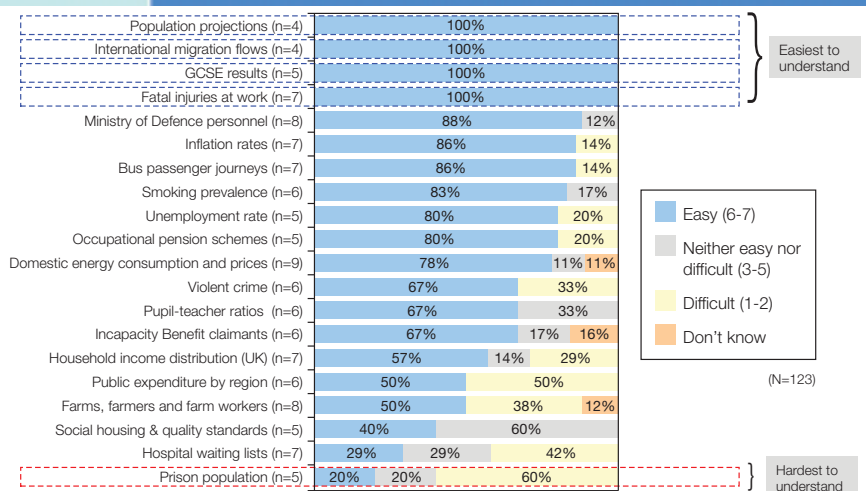
Understanding the information

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



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.

Researchers: by topic area



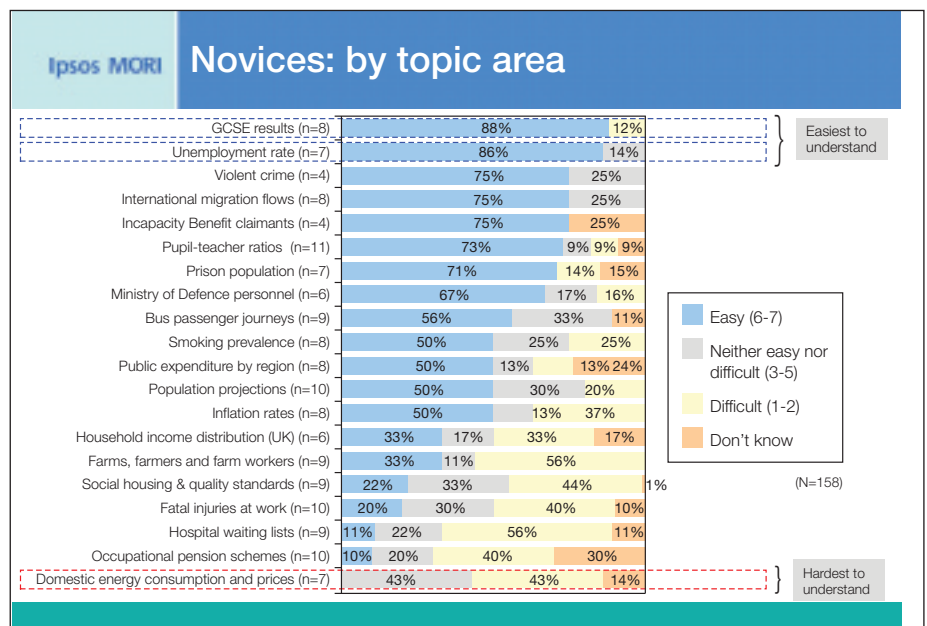
Conversely, novices were less likely to find the information they came across easy to understand. While nearly nine in ten found GCSE results (88%) and unemployment data (86%) easy to understand, no novices understood domestic energy consumption and prices data.

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

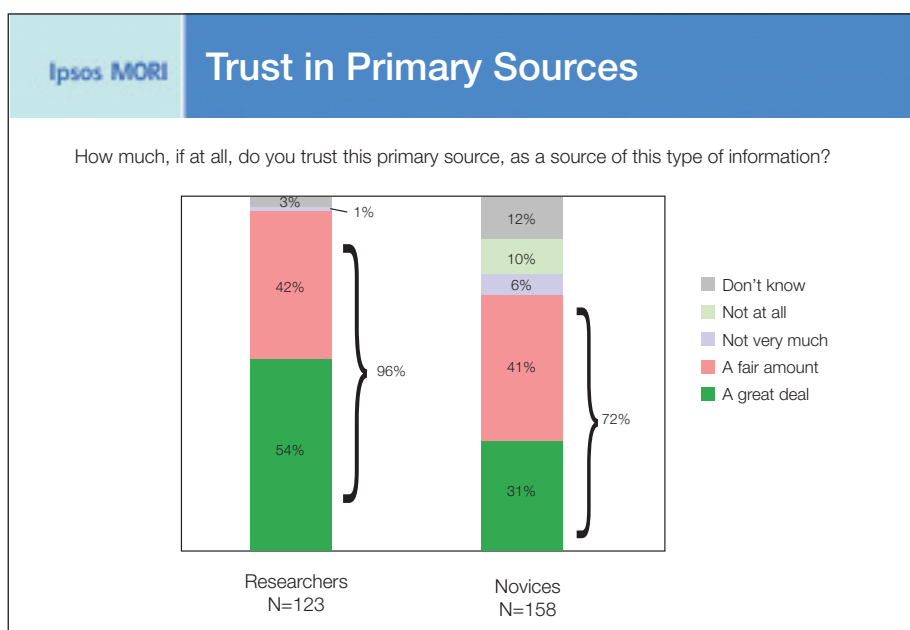
Male Novice, Edinburgh

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

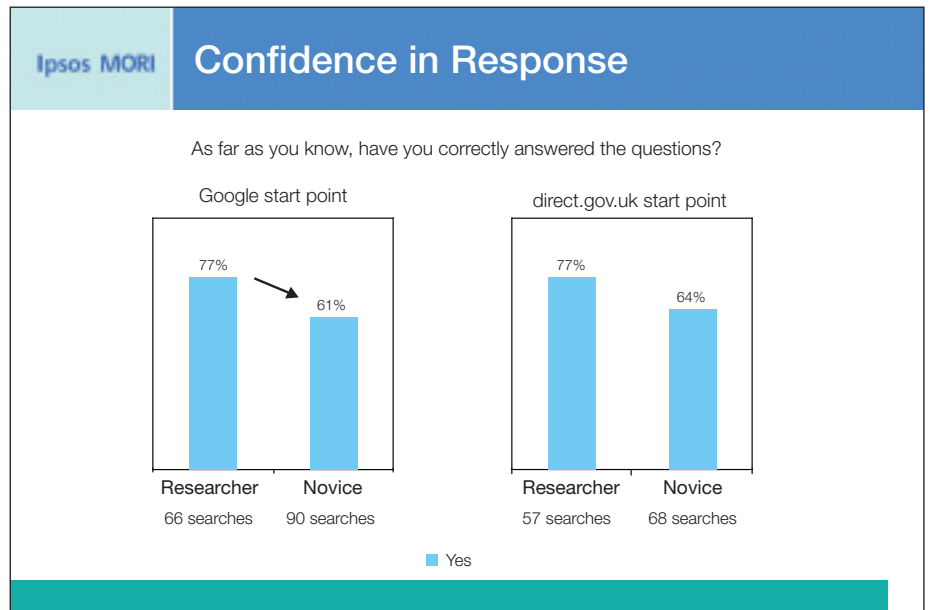
In fact, when asked to what extent they trusted their primary sources, researchers are most likely to trust their sources for inflation rates (86% trust a great deal) and public expenditures (83% trust a great deal). In both these cases, researchers listed government sites as their primary source (statistics.gov.uk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

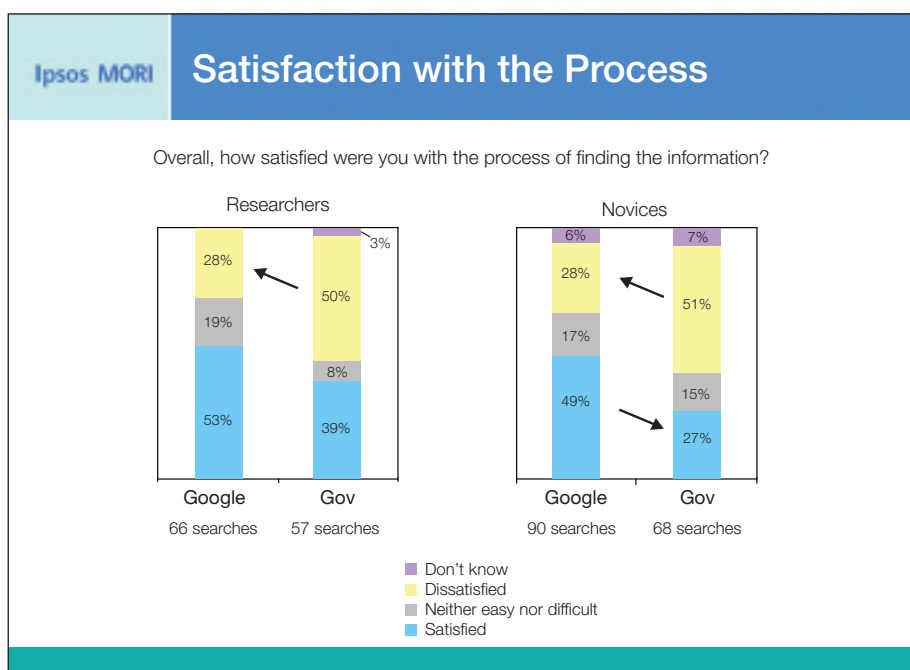
Overall, 70% of participants felt they had correctly answered the question. Not surprisingly, researchers are more confident in their response, with more than three-quarters (77%) reporting that they had found the right answer. Novices were less sure, with 62% reporting the same. Starting points had 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	67
Core Questionnaire	71

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