



Analysis of 'Open Data' survey commissioned to support the Shakespeare Review into Public Sector Information

Field dates:

22 February – 15 March 2013

2 April 2013

David Dobrin
YouGov plc

Executive Summary

The Shakespeare Review into Public Sector Information (PSI) is looking into ways in which the public sector can improve access to data in order to promote efficiencies, stimulate economic growth and to bring wider benefits to society. In other words, the review is examining open data policy. To support this independent review, YouGov conducted an online survey aimed at ascertaining views into open data and the use of PSI. The survey was completed by the general public, as well as those who have an established interest in, and knowledge of open data.

Methodology

Who we spoke to

The survey was sent to two distinct groups of respondents. The first group of respondents were taken from YouGov's existing UK panel. 4,000 people from a nationally representative sample were asked whether they would be interested in answering a survey on 'how sharing public data could improve public services'. The survey was then sent to those who indicated an interest and 777 respondents fully participated. This group of respondents were taken from a sample of the general population and were offered a small incentive to participate. For the purpose of this report, we will refer to this group of respondents as 'YouGov Panel'.

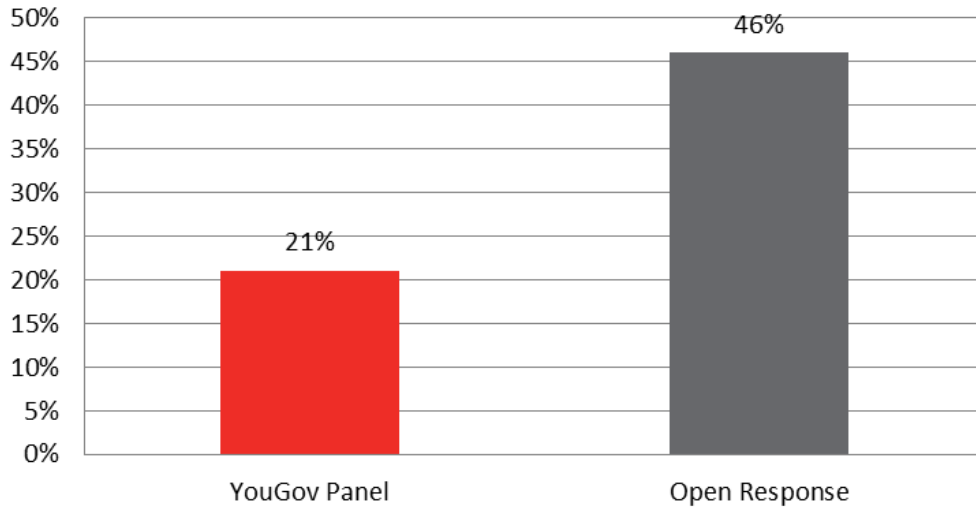
The same survey was also published publicly on the YouGov website and through the YouGov Twitter Account (c. 39,000 followers). This survey was open for anyone to complete over a 3 week period. Certain stakeholders were targeted by the Department of Business, Innovation & Skills (BIS), while others were targeted through data industry promotion. In total, 635 respondents fully participated. For the purpose of this report, we will refer to this group of respondents as 'Open Response'.

The two groups of respondents

The crucial difference between the two groups was that a far larger proportion of those who completed the open survey have a special interest in, or knowledge of, PSI and open data.

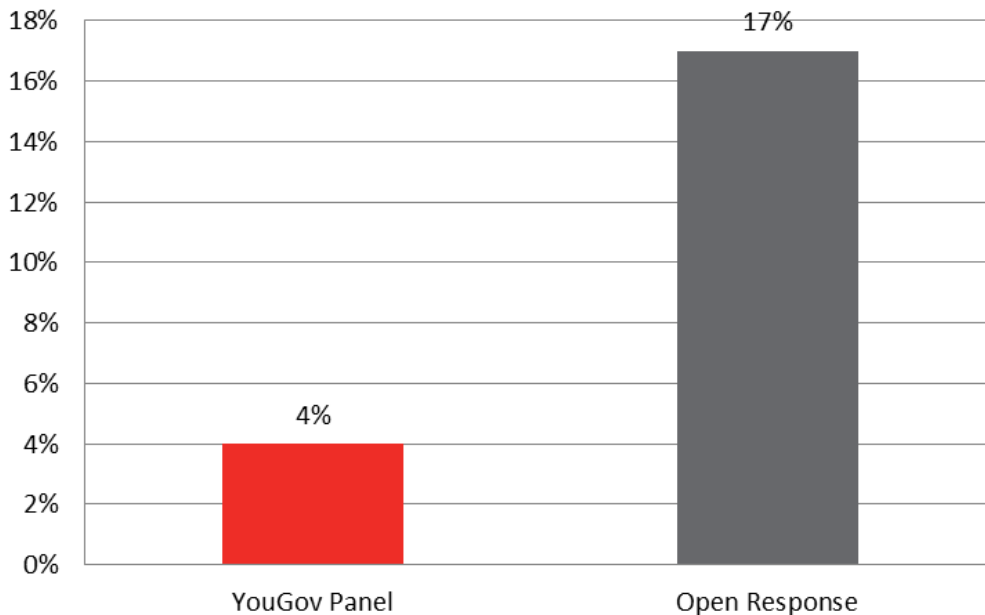
As the graph below indicates, 46% of those who completed the open survey said they have a professional, academic or special interest in open data, compared to just 21% of those from the YouGov panel.

% of respondents who said they had a professional, academic or special interest in open data



Furthermore 17% of open respondents state that they are 'highly informed' on data issues, compared to just 4% from the general YouGov panel.

% of respondents who said they were 'highly informed' on data issues



As we will see, the findings of this survey are significant particularly because of the *shared* views held by the two groups of respondents, irrespective of their existing use of, or interest in, PSI and open data.

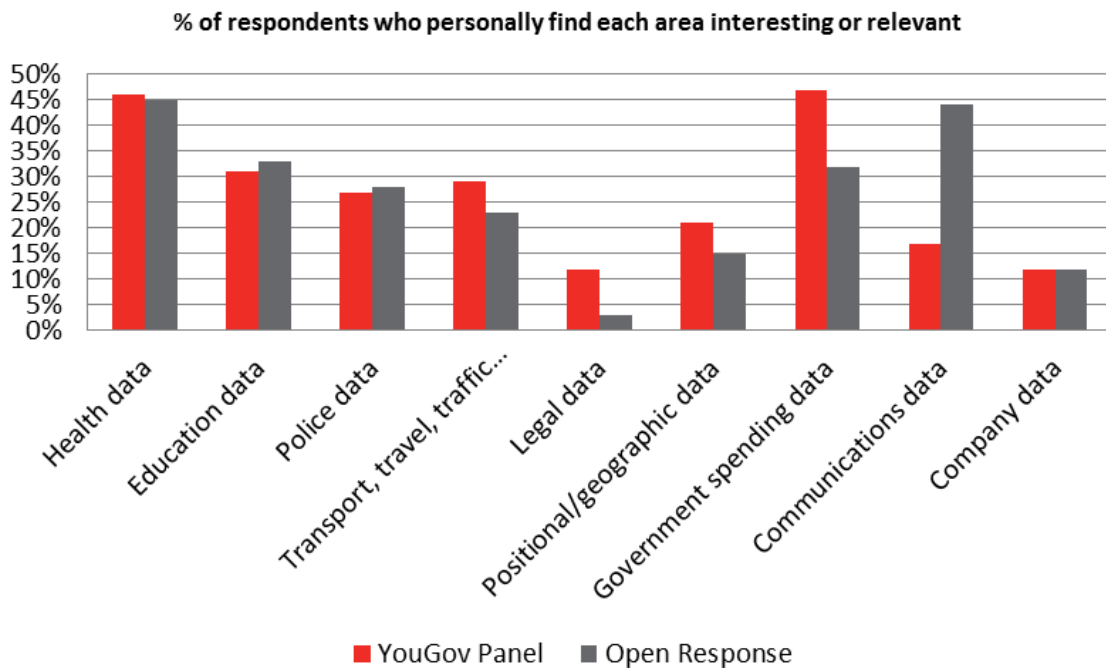
Methodological approach

Rather than conducting a traditional quantitative survey that gives respondents a limited choice of balanced answers, we wanted to provide the opportunity for respondents to elaborate on their own views and ideas to inform the Shakespeare Review. As such, this survey involved a far greater number of ‘open-box’ questions (in which respondents type their responses), than closed questions. In this respect, the survey is consultative.

Findings

What is particularly striking about the findings of this consultative survey is the degree to which the two sample groups broadly shared the same views.

Towards the beginning of the survey, respondents were asked which areas they personally find most interesting and relevant. As we can see from the graph below the degree to which each of the groups are interested in particular areas is broadly the same (with the exception of ‘legal data’, ‘government spending data’ and ‘communications data’).



As well as sharing similar interests in data topics, it is clear that the two respondent groups also have shared views on the main priorities that should be considered as part of a policy for open data. This includes: the extent to which PSI should be published; the safeguards that must be guaranteed; and ensuring that the most is made of the opportunity presented by open data.

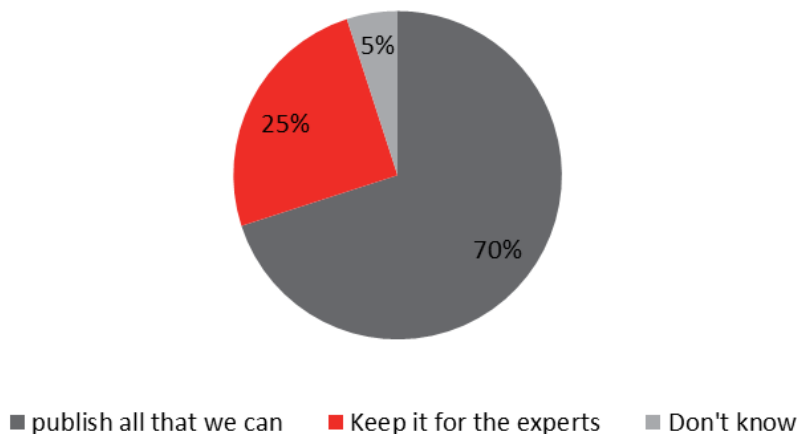
How much data should be published?

A key element of the open data debate is the *extent* to which we open up government data. How far do we go and how much data should we release? Using the example of the healthcare system, we asked respondents which of the following statements came closest to their views;

1. We should collect and publish all that we can about what happens in our health system, including the outcomes of all procedures, because that way we can make more intelligent decisions for ourselves and for policy.
2. We should be careful to keep information only to the experts because they are best placed to make best use of it, and too much transparency will create confusion and end up doing more harm than good.
3. Don't know

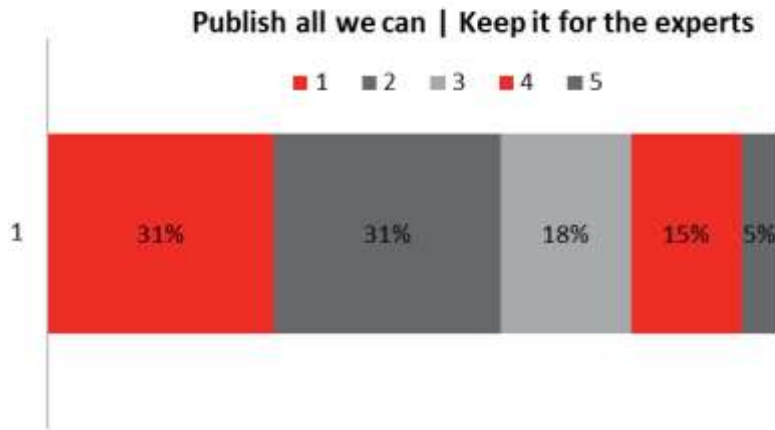
The results indicate that 70% of total respondents think that we should collect and make public all that we can about our health care system, whilst 25% believe we should just keep it to the experts so as to prevent confusion.

% of respondents on the amount of health system data that should be published



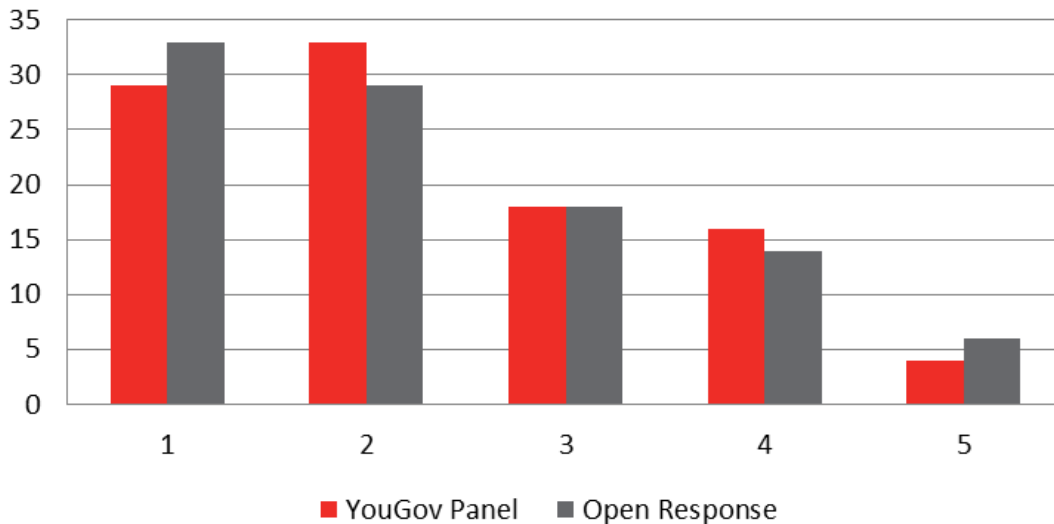
In order to assess the views of our respondents in further detail, we offered them the opportunity to place themselves on a continuum of opinion, where 1 is 'publish all we can' and 5 is 'keep it for the experts'. While the results indicate a fairly balanced distribution of opinions, a majority of respondents are in favour of publishing 'all we can'.

% of respondents who placed themselves on a continuum where 1 is 'publish all we can' and 5 is 'keep it for the experts'



Furthermore, when we compare the two groups of respondents, it is clear that they also share similar views across this continuum and are almost equally in favour of publishing 'all we can' as opposed to keeping the data 'for the experts'. This is demonstrated in the following graph -

% breakdown of YouGov Panel & Open Respondents across the continuum, where 1 is 'publish all we can' and 5 is 'keep it for the experts'

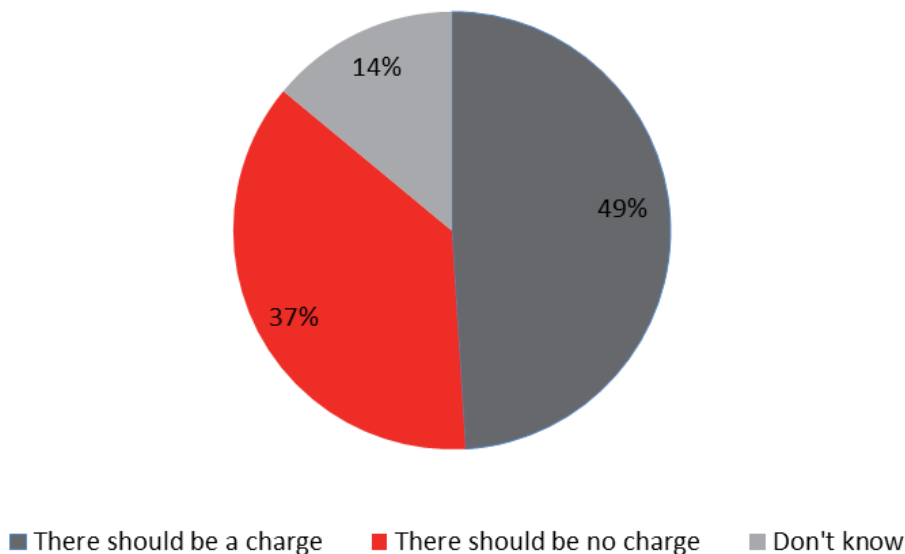


It is significant to note that the vast majority of respondents are in favour of releasing and publishing as much information as possible. Although this question relates specifically to healthcare, it may be indicative of a trend towards welcoming open data across other sectors.

Charging for the Data

We also asked respondents about whether they think there should be a charge for the use of PSI data. Using the Postal Address File as an example, we asked respondents whether or not there should be a charge for this data. As the following graph indicates, views are fairly balanced with a slight majority in favour of a charge.

% of total respondents on whether there should be a charge for the Postal Address File



There are two principle concerns amongst those who believe there should be a charge for public data, such as the Postal Address File. The first is centred on maintaining the quality and availability of core datasets. The second is more complex and is centred on whether a charge should be dependent on who is using the data and for what purpose; for example, it should be free for individuals, but not for businesses that will use that same data to make profit.

Representative comments of those who believe there should be a charge include:

- 'We need to ensure that essential information continues to be collected and maintained to a high quality and that costs money'
- 'Charge should sensibly reflect use and likelihood of user making financial gains from the data'
- 'Some should be free, other more commercially viable data should be used to recover investment'
- 'Why should everybody pay for a service that only specific individuals or companies derive benefit from?'
- 'Maybe we could be more discerning about how it's going to be used and charge if the user wants it for financial gain. But charging system needs to be simple'

The primary concern amongst those who do not believe there should be a charge, is that charging presents a barrier to entry and thus prevents the maximum possible effective use of open data for social and economic purposes.

Representative comments amongst those who do not believe there should be a charge include:

- ‘The benefits of greater use would in time outweigh the income generated through licensing’
- ‘The evidence is fairly conclusive that charging for datasets provides a pretty significant de-motivator for reuse. Much of the innovation that might happen never gets to the brainstorming stage, because developers won't spend time exploring datasets (and imagining possible applications for them) if they can't access them easily’
- ‘Charging would be an unnecessary barrier to entry’

It is clear that while respondents are in favour of open data, a more nuanced policy needs to be developed on whether (and if so, to whom) it should be charged for.

As well as the cost of releasing this data, respondents also highlighted a number of concerns about the widespread release of data, which must be safeguarded against.

The Main Concerns

Whilst our respondents hold favourable views toward the release of open data, they also have certain concerns that could potentially present obstacles to its effective and safe release. The concerns revolve around three key themes: *privacy*, *openness* and *security*. Representative comments relating to each of these three themes include:

Privacy

- ‘Privacy is important, and there are ways to ensure it. We can't be scared of the risks that we allow the opportunity to improve to pass us by’
- ‘It is not difficult to produce huge benefits from open data without compromising privacy at all: yet most people seem to regard these two outcomes as mutually exclusive. This is a mistake’
- ‘I do feel that ensuring the data is anonymous is important to avoid the requirement to obtain consent from individuals (which may be withheld and thus skew data)’
- ‘Benefits outweigh the risks’
- ‘Frankly, I don't have any privacy issues. If the data is anonymised (and is checked properly by humans to ensure small populations can't be identified), then there is no problem’

Openness

- ‘Opening up all non-personal public data, to strengthen democracy, increase innovation and encourage competition’
- ‘Openness and availability of data is essential’
- ‘Encourage free and easy access to data’
- ‘That the owners of that data agree that it can be used by whoever wishes to’

Security

- ‘To protect seriously private and important personal and state data- but enable those who can build new services and improve old ones via data insights, to do so easily and effectively, all in cause of efficient public services’
- ‘I am concerned about information getting into the wrong hands’
- ‘A difficult situation given the level of Cyber-crime currently. However if implemented correctly there should be no greater danger to exposing private data than there is now and possibly much less’

The ‘concerns’ as priorities

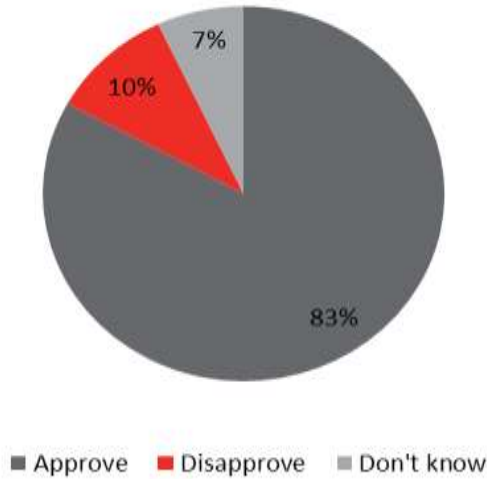
However, a majority of the total respondents feel these concerns can be overcome. Respondents would like to see these three themes (*privacy, openness* and *security*) recognised as fundamental components of any official policy on open data.

In our survey it was suggested that a proposal to make data as ‘open’ as possible would have three crucial conditions:

1. No data is ever published with names, addresses, dates of birth etc. and as far as possible we will ensure technological guards against personal identification.
2. We have new rules against mis-use of personal data as well as stricter enforcement of existing rules.
3. Every citizen has the right to opt out of having their data published even under these conditions.

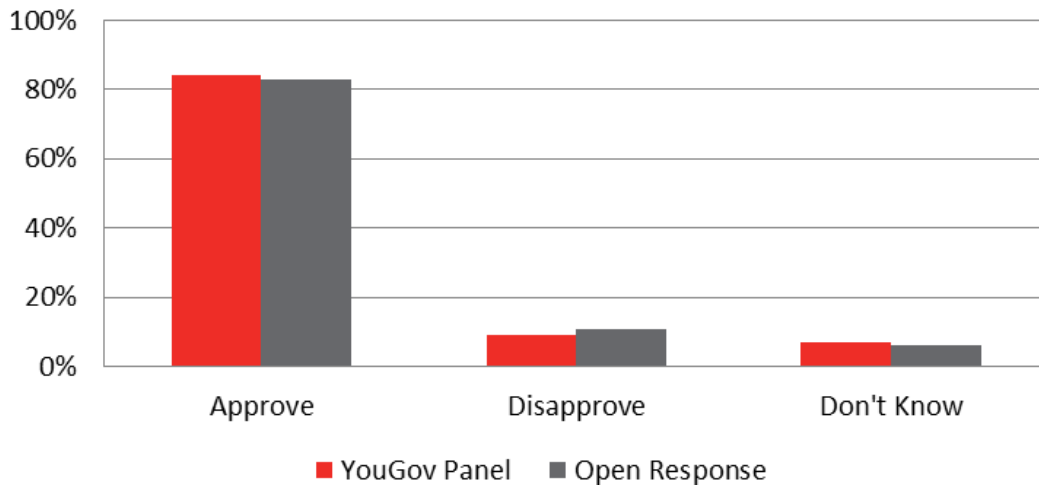
On the basis that these conditions would be guaranteed, we asked respondents if they would approve or disapprove of a general policy for open data. The graph below indicates that a vast majority of respondents would approve of a general policy if these conditions were guaranteed.

% of respondents who would approve or disapprove of a general policy for open data if the specified conditions were guaranteed



Significantly, as the graph below indicates, the two groups of respondents shared almost identical opinions relating to a general policy of open data under the specified conditions.

% of YouGov Panel & Open Respondents who would approve or disapprove of a general policy for open data if the specified conditions were guaranteed



How do we make the most of the opportunity?

As well as asking respondents about their views on releasing open data and the guarantees that must be made, we also wanted to hear their views about how to make the most of the potential opportunities. Respondents were asked to comment on how we can ensure that we have the capability to make the most of the opportunities for data sharing and enterprise. As you would expect, a key theme to emerge was 'education'. Other themes to emerge included utilising the skills and expertise of the private sector, as well as inspiring people by demonstrating the potential social and economic benefits.

Below is a representative selection of respondent's suggestions for how to ensure that we make the most of the opportunities presented by open data.

Education

- 'Promote those with information management skills - test, teach and coach in this discipline as we'd expect to with any other core competency'
- 'By influencing the Universities to run courses that focus on key skills needed - employability emphasis. By adding requirements to recognised accredited skills development'
- 'Large scale education issue that needs to start in schools - encourage government to create posts for the appropriate kind of people. Some kind of expertise centre'

Using expertise of private sector

- 'Work with private sector - particularly local people, local entrepreneurs and SMEs - to understand the information areas they are likely to find useful in the short-medium term'
- 'I think that the ODI among other initiatives are in part an answer to this problem. Making it easier to start businesses by better funding options and a less punitive tax regime would grow the sector even more and improve skills as a natural consequence'
- 'There are already people who work with data and have been trained by their own companies. You could use them to train others'

Ensuring the data is usable

- 'The benefits in an open data approach depend often on core reference data and on the opportunity to cross reference and share / integrate different data. Having this data in silos and in multiple formats with varying levels of access prevents the real value being unlocked'
- 'The word 'usable' in the previous selection is important. You can make as much open data available as you like but if it is not easily 'usable' then it is of little value'
- 'Do not fly tip data without standards'
- 'My focus is to create interoperability through the release of data standards alongside the data itself. Therefore most data and outcomes can be improved by

implementing a policy that requires standards to be released alongside the data sets’

Demonstrating potential prospects

- ‘highlight the positives, make it something that the bright and best want to be part of, publicise’
- ‘inspire young people at school with the possibilities’
- ‘evidence-based storytelling’
- ‘educate, enable, implement and reiterate based on the evidence surrounding the use of that data’

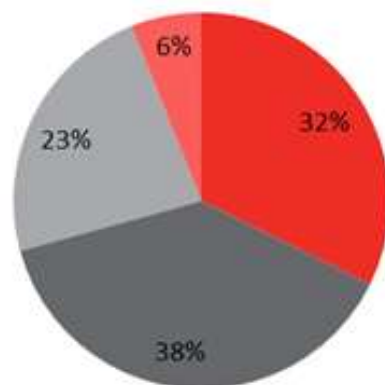
The Trading Fund Model

The results indicate that, of those respondents who know about the ‘Trading Fund model’, the majority believe that the model should be fundamentally changed or at least reformed, if we are to fully realise the benefits of increasing open data.

Following a brief introduction to each of the funds, we asked respondents how much they knew about the existing Trading Fund model.

To those who said they know ‘quite a lot’ or ‘quite a bit’ about how the trading funds operated (15% in total), we asked whether they think the Trading Fund model should be changed, reformed or not interfered with at all.

% of respondents who said they knew ‘quite a lot’ or ‘quite a bit’ about how trading funds operated, on whether they think the trading fund model should be changed, reformed or not interfered with



■ Fundamentally changed ■ Slightly reformed
■ Not be interfered with ■ Don't know

Significantly, 70% of respondents who indicated they 'knew a lot' or 'quite a bit' about the Trading Fund model said that it should be 'fundamentally changed' or 'slightly reformed'. When respondents were given the opportunity to explain why they believed the model should be 'fundamentally changed', it became clear that these people believe that the current Trading Fund model will inevitably inhibit the release, and the potential prospects of, open data. Here is selection of the comments given:

- 'There is a basic conflict between the Trading Fund model for data-holding services and the idea of making data open'
- 'The idea that a small number of actors working on a closed dataset can somehow outperform a large number of actors working on an open dataset is demonstrably incorrect. We live in an age where we have ubiquitous access to the web, commoditised access to supercomputing (the cloud), and the ability for dedicated specialists to convene around issues anywhere in the world. Trading funds have to adapt to this - it's not an option'
- 'Trading funds are forced to raise revenue by charging, and hence restricting use. In many cases the effort involved in this is out of all proportion to the returns'
- 'The trading fund model inevitably results in a focus on what will benefit the TF business rather than considering the national good'
- 'Charging marginal cost for data collected to support the public task is not possible under the current set up'

Of those respondents who believe the Trading Fund model should be 'slightly reformed', it is clear that whilst they do not believe the current model would inevitably inhibit the release and potential for open data, they do see a need to at least moderately reform the status quo.

Here is selection of the comments given:

- 'The problem is not structure - it is culture and lack of implementation of what is a perfectly good regulatory system for the TFs'
- 'Evolution not revolution'
- 'It is not realistic to fundamentally reform how the trading funds work. All those organisations (with the possible exception of OS), could do more to make their data open and available, without jeopardising their funding model, or Trading Fund status. The disruption of changing the business model for Land Registry, Companies House, Met Office, just to get open data from these organisations, is too disruptive, when those organisations could move to open data by default business models, within the current Trading Fund system'

- 'Works quite well already and reform would mean privatisation which is inappropriate'
- 'There is more economic benefit to be had from reforming some of this data'

Concluding Thoughts

The results of our survey indicate that a significant number of respondents hold favourable views regarding the release and publication of open data. However, in order to make the most of the opportunity presented by open data, a number of issues must be addressed.

This is particularly the case in relation to clearly demonstrating to the public the benefits of open data and ensuring that open data is published in a usable format. Furthermore, amongst those who understand or have knowledge of the current Trading Fund model, many believe that it should be fundamentally changed, or at least moderately reformed, if we are to fully realise the benefit of open data.

What is perhaps most striking from this survey is the degree to which the two groups of respondents share similar views. This suggests that even those who do not have knowledge of open data really understand the benefits and the crucial issues that must be addressed for any policy to be effective. The public have a strong desire to achieve more value from the existing data and information held within public sector bodies. On the condition that certain safeguards can be met, our survey suggests that the public will support a strong and resolute approach to making effective open data a reality.