



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

Field dates:

4 – 10 April 2013

---

13 May 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 has conducted two online surveys aimed at ascertaining views into open data and the use of PSI.

This second survey was designed to ascertain respondent's views on the draft recommendations of the Shakespeare Review.

## **Who we spoke to**

The survey was sent to the same groups of respondents who had participated in the first survey.

The first group of respondents were taken from YouGov's existing UK panel and for the purpose of this report we will refer to this group of respondents as 'YouGov Panel'. This group of respondents were offered a small incentive to participate.

The second group of respondents was made up of those who had participated in the first survey: following its publication on the YouGov website and Twitter Account (c. 40,000 followers); because they are stakeholders and were targeted by the Department of Business, Innovation & Skills (BIS); or through targeted data industry promotion. For the purpose of this report, we will refer to this group of respondents as 'Open Respondents'.

This second survey was also open for anyone to participate, regardless of their prior contribution.

For this second survey we received submissions from 846 respondents in total – 604 from the YouGov Panel and 242 Open Respondents.

In total, 69% of those who completed this survey indicated that they had completed the first survey. If we break this down we know that 64% of the YouGov panellists indicated they had completed the previous survey, while 83% of Open Respondents said that they had completed the previous survey.

## **The Recommendations**

The survey tabled ten broad recommendations. Respondents were asked to either support or oppose the recommendations and were given the opportunity make a comment against each.

## **Recommendation 1**

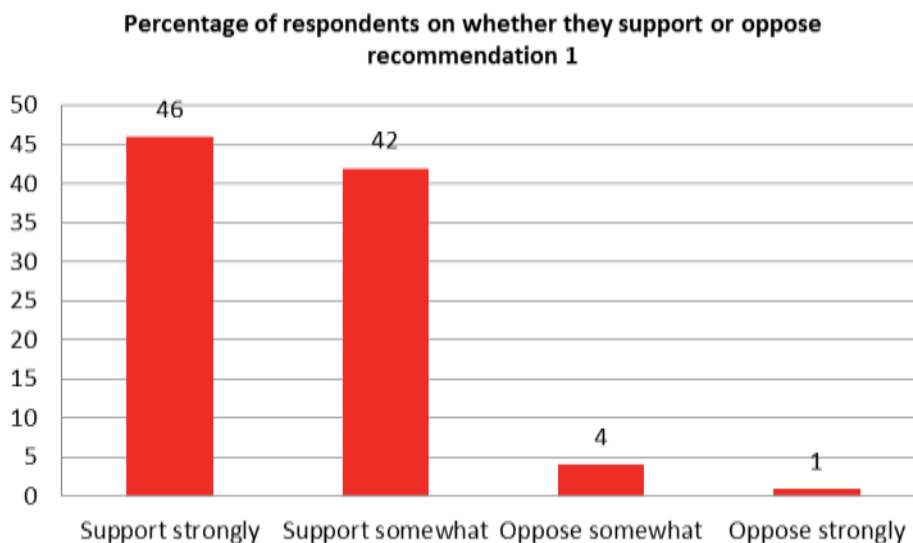
*Every department of government should make an immediate commitment to publish their single most important dataset quickly, to a high standard agreed to maximise linkability, ease of use and free access. They should also commit to maintaining that dataset and keeping it regularly updated.*

*Most important dataset to be defined as the one that is used most often by the department itself to carry out its work (or the one most requested by outside users, if that demand can be demonstrated and is significantly different); these datasets taken together will be the 'National Core Reference Data'.*

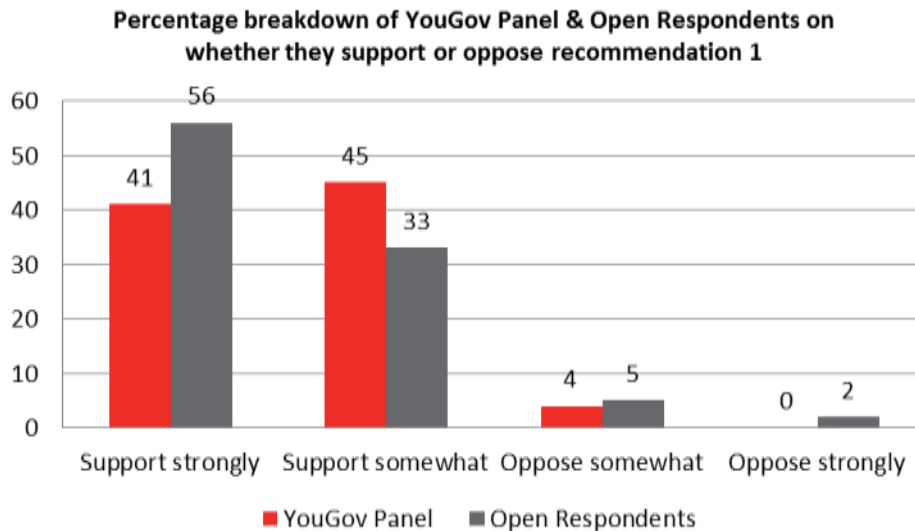
*The situation currently is that most departments have done a reasonable job at making some datasets available; my proposal says it must be their most important datasets, the ones that define their core work; and currently they are not necessarily published and maintained to a consistent format and standard, and that should now happen.*

## **Results**

As we can see from the table below, 88% of total respondents said that they 'strongly' (46%) or 'somewhat' (42%) support Recommendation 1.



When we compare the two groups of respondents we can see that both are broadly supportive of Recommendation 1. While more Open Respondents tended to 'strongly' support (56% compared to 41%), more YouGov Panellists tended to 'somewhat' support (45% compared to 33%).



Comments

By looking at the open comments, most objections centred on the ambiguity of the definition ‘most important dataset’. One respondent stated that “just because the data is used most often, or requested most often does not necessarily mean it is the most important”.

Other comments questioned the process by which datasets are released. For example some respondents suggested asking the people what datasets they wanted released, otherwise departments may release data that they consider important but which is “completely useless to anyone else”.

**Recommendation 2**

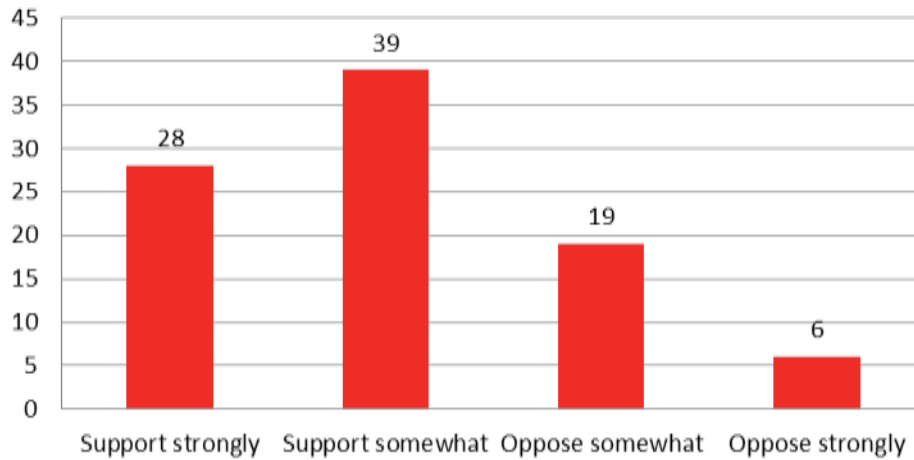
*Alongside this high-quality core data, departments should commit to publishing all their datasets (in anonymised form) as quickly as possible without concerns about quality - that is, if there is a clash between data quality and speed to publication, they should follow the ‘publish early and ugly’ principle because data scientists are well accustomed to getting value out of imperfect data.*

*Currently many datasets are held back because it is felt they are not ready because they are not of sufficiently high quality, and that resources prevent their speedy improvement. But data users say that lower quality is not as much of a problem as is non-publishing.*

Results

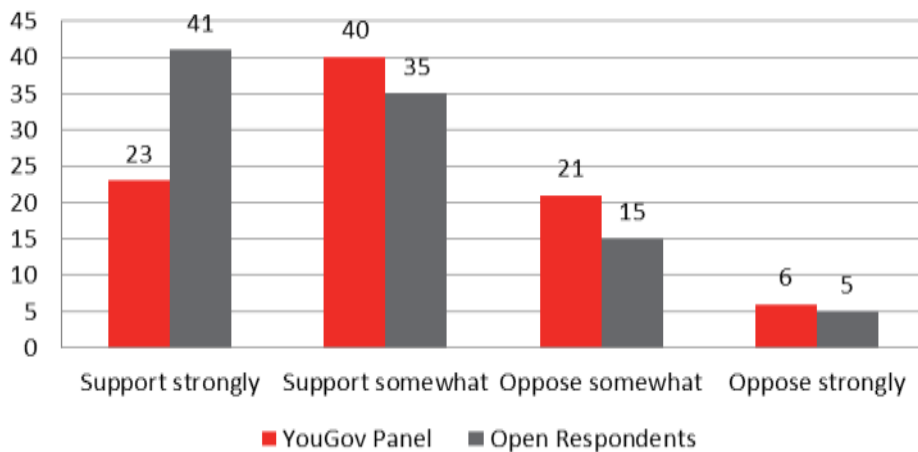
As we can see from the table below, 67% of total respondents said that they ‘strongly’ (28%) or ‘somewhat’ (39%) support Recommendation 2.

**Percentage of respondents on whether they support or oppose recommendation 2**



When we compare the two groups of respondents we can see that both are broadly supportive of recommendation 2. While more Open Respondents tended to ‘strongly’ support (41% compared to 23%), more YouGov Panellists tended to ‘somewhat’ support (40% compared to 35%).

**Percentage breakdown of YouGov Panel & Open Respondents on whether they support or oppose recommendation 2**



By looking at the open comments, most were centred on the argument that poor quality data isn’t useful and is open to misinterpretation. Comments included:

- “Poor quality data is useless and can be misinterpreted”
- “We have bitter past experience of “ugly” data and we wasted huge amounts of manpower and resources on trying (unsuccessfully) to clean it up.”
- “Get it right or don’t do it. Confusion results from wrong info.”

Others suggested that in order to overcome this potential issue, a “minimum standard” should be achieved “before rushing to publication”.

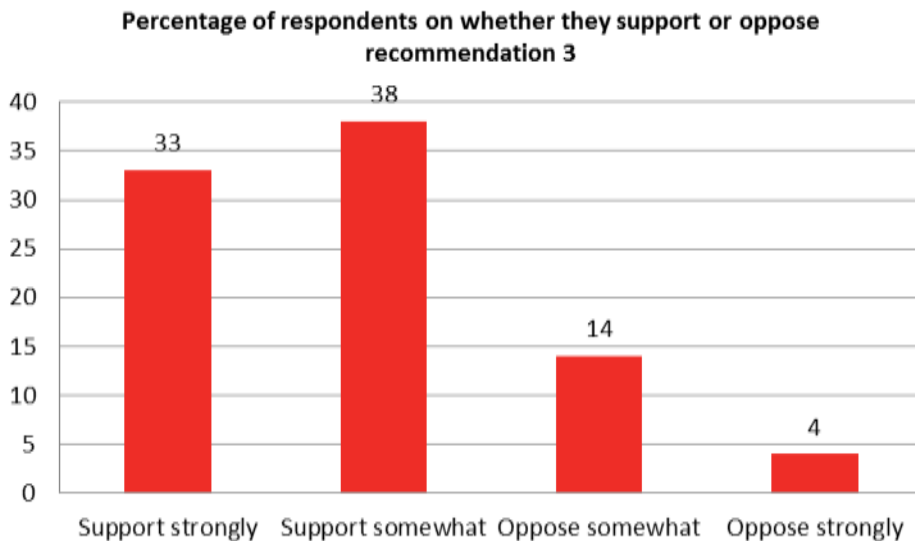
### **Recommendation 3**

*Recommendations 1 and 2 taken together define a twin-track policy for a simultaneous 'high quality core' AND a 'publish early and ugly' policy. This twin-track policy will maximise the benefit within practical constraints (with the further recommendation that departments take pride in adding as many datasets as possible and as quickly as possible from track 1 and track 2).*

*This approach reduces the excuses for poor or slow delivery; it says 'get it all out and then improve'.*

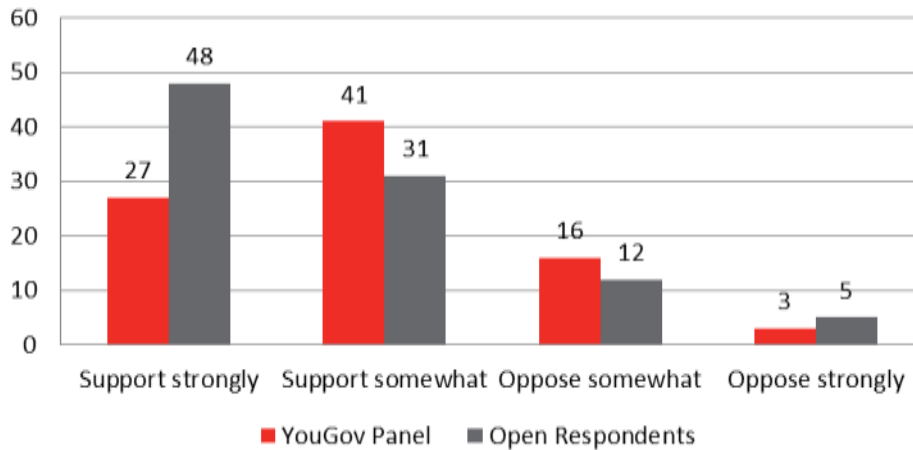
### **Results**

As we can see from the table below, 71% of total respondents said that they 'strongly' (33%) or 'somewhat' support (38%) Recommendation 3.



When we compare the two groups of respondents we can see that both are broadly supportive of recommendation 3. While more Open Respondents tended to 'strongly' support (48% compared to 27%), more YouGov Panellists tended to 'somewhat' support (41% compared to 31%).

**Percentage breakdown of YouGov Panel & Open Respondents on whether they support or oppose recommendation 3**



By looking at the open comments, some respondents suggested that this approach could give certain departments a “ready-made excuse for poor quality work”. Others went further by suggesting that releasing poor quality data “causes more problems than it solves”.

**Recommendation 4**

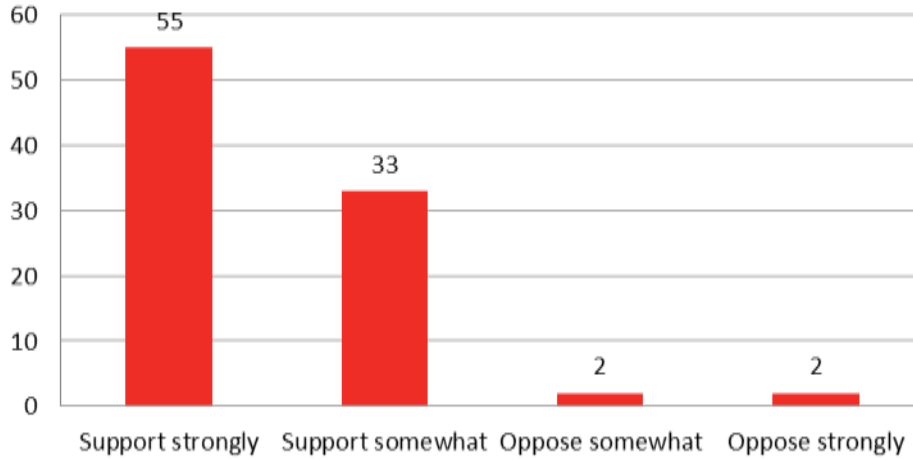
*Building on existing activities, there should be an immediate programme of investment in basic data science through our academic institutions, covering both genuinely unfettered 'basic research' and research of 'practical immediate value' to the national data strategy. We cannot rely only on markets and government departments to maximise the potential of this relatively new and fast-developing field in which we are positioned to be a world leader.*

*At the moment, America invests massively more than us and continuously reaps the benefits in world-leading business applications of science and technology; yet Britain is capable of being first in this field, given our strength in data science and the fact we have large, coherent datasets. For example, nowhere in the world has such good health data, due to the scale of the NHS as a single provider. There is huge potential here for building social and economic value if we are willing to invest smartly.*

**Results**

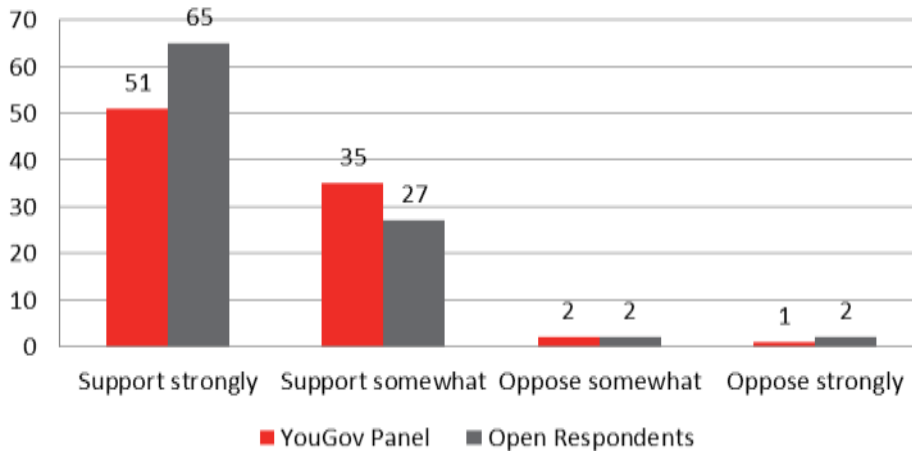
As we can see from the table below, 88% of total respondents said that they strongly (55%) or somewhat (33%) support recommendation 4.

**Percentage of respondents on whether they support or oppose recommendation 4**



When we compare the two groups of respondents we can see that both are broadly supportive of recommendation 4. While more Open Respondents ‘strongly’ support (65% compared to 51%), more YouGov Panellists support somewhat (35% compared to 27%).

**Percentage breakdown of YouGov Panel & Open Respondents on whether they support or oppose recommendation 4**



Most of the comments around this recommendation indicated that respondents were supportive of this recommendation on the basis that “the social and economic value can be guaranteed to benefit the British people”. Furthermore many respondents who supported this recommendation also argued that the UK should focus heavily on research and investment and should stop being short-sighted in its approach.



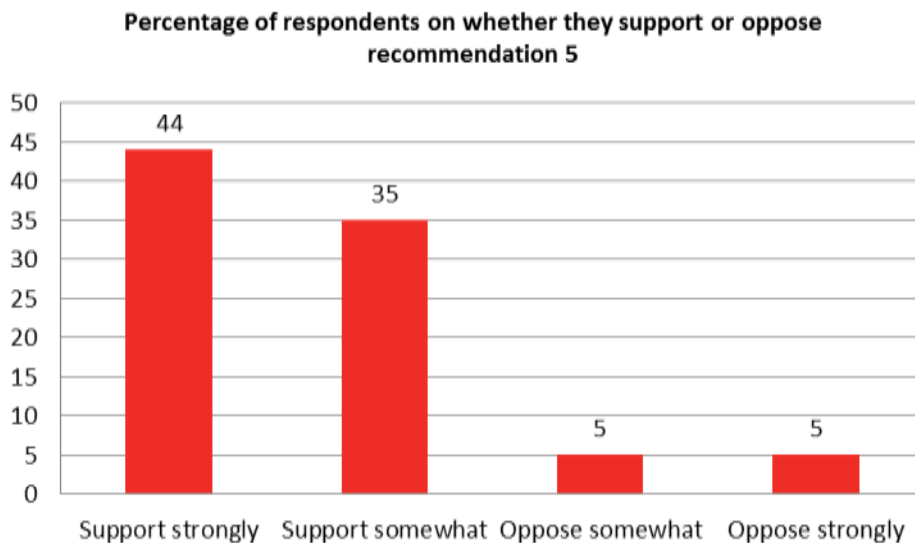
## **Recommendation 5**

*We should have a clear pragmatic policy on privacy and confidentiality that increases protections for citizens while also increasing the availability of data; we can do this by putting in place guidelines for publication that, if correctly followed, pushes responsibility for (mis)use on the end (mis)user, strengthening application of punitive consequences. Especially sensitive datasets should be accessible only to those who can demonstrate sufficient expertise in the area and whose activity with the data is traceable.*

*We currently have an unrealistic degree of expectation of any data holder to perfectly protect all our data, which has led to a situation where data scientists are presumed 'guilty unless proven innocent' - an attitude that inhibits innovation. Following 'Best practice' guidelines should be enough, so long as we are willing to prosecute those who misuse personal data. Otherwise we will miss out on the enormous benefits of Public Sector Information, including open government data.*

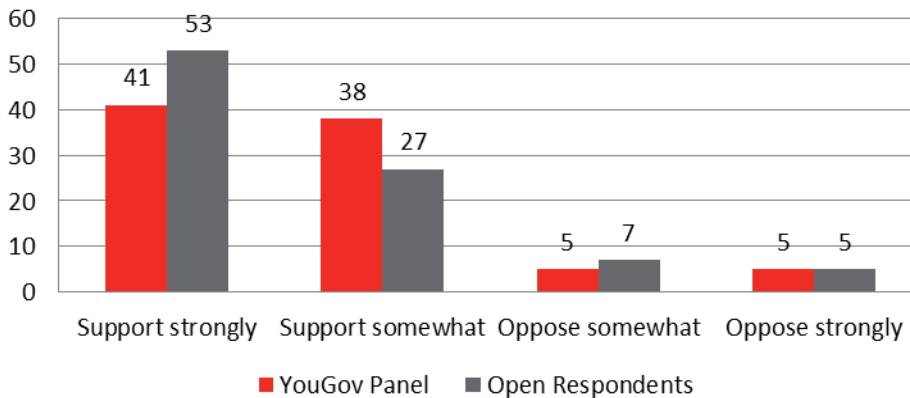
## **Results**

As we can see from the table below, 79% of total respondents said that they strongly (44%) or somewhat (35%) support recommendation 5.



When we compare the two groups of respondents we can see that they are broadly supportive of recommendation 5. While more Open Respondents 'strongly' support (53% compared to 41%), more YouGov Panellists support this recommendation 'somewhat' (38% compared to 27%).

**Percentage breakdown of YouGov Panel & Open Respondents on whether they support or oppose recommendation 5**



There was an array of comments around this recommendation. Some who supported the policy argued that the UK “errs on the side of caution too much”. Other argued that “the cult of perfectionism has invaded the public sector. If a genuine mistake is made then there should be no witch hunt”.

However a number of respondents were concerned about who is ultimately responsible for privacy breaches. Some respondents stated that responsibility does not only rest with the end user. One respondent said that “fully informed laws must cover the data issuers and end user”. Another respondent said that “the data holder should be held accountable as no doubt the data holder has made promises to data providers which it should meet”.

**Recommendation 6**

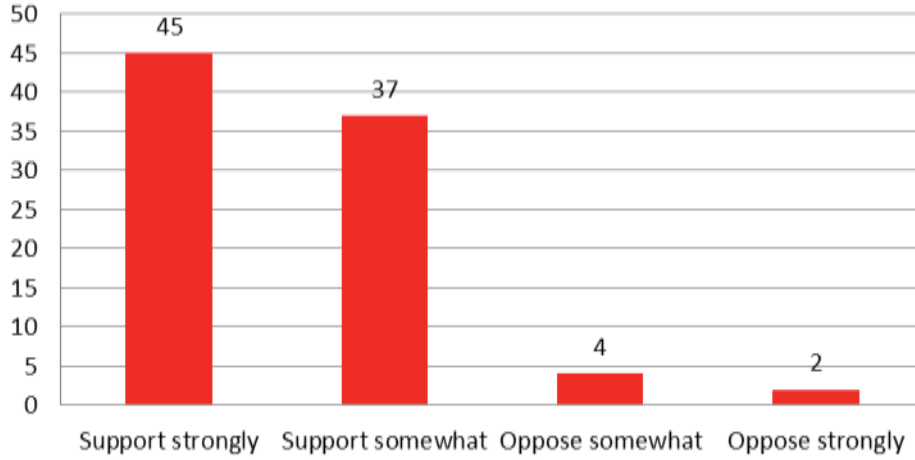
*We should have a mechanism for driving the implementation of the national data strategy throughout the public sector, and its oversight. This should include clarity about what data is/can be available, with a feedback loop for its improvement; it will be continuously accessible to citizen and business- user influence. The idea is to be an exemplar of the democratic crowd-sourcing of decisions.*

*We have several committees, boards, overseers and champions of data; but no easily understood, easily accessed, easily influenced mechanism for making things happen. This is ironic given it's all about 'information'. We should create a single channel for driving Public Sector Information, including open government data through the system.*

**Results**

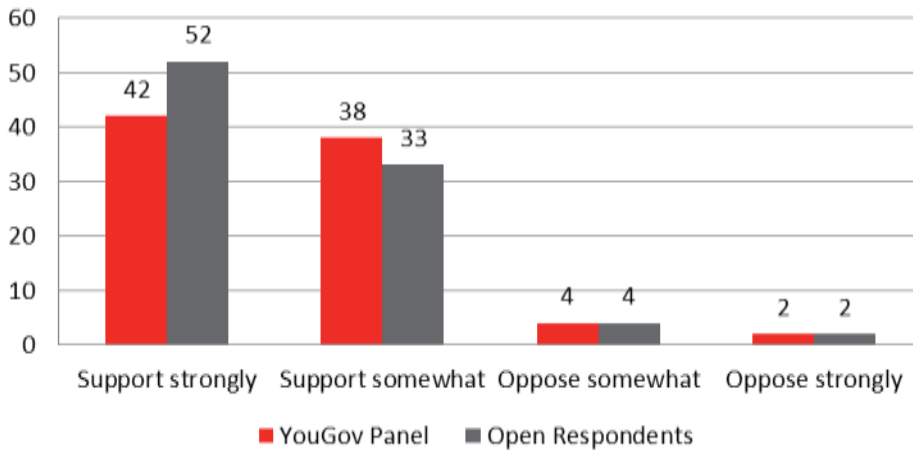
As we can see from the table below, 82% of total respondents said that they strongly (45%) or somewhat (37%) support recommendation 6.

**Percentage of respondents on whether they support or oppose recommendation 6**



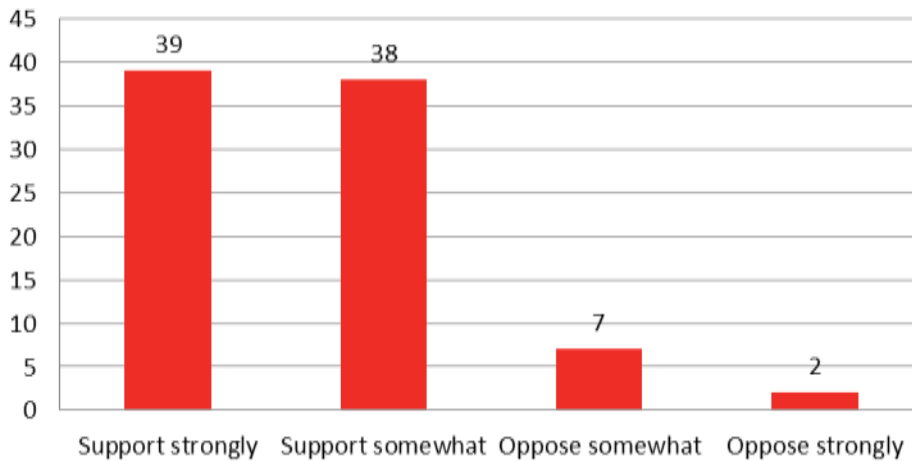
When we compare the two groups of respondents we can see that they are broadly supportive of recommendation 6. While more Open Respondents ‘strongly’ support (52% compared to 42%), more YouGov Panellists support this recommendation ‘somewhat’ (38% compared to 33%).

**Percentage breakdown of YouGov Panel & Open Respondents on whether they support or oppose recommendation 6**



Some respondents were concerned that “the views of a limited group would drive the control and access to information”. Others stated that we need to speed up the process for requesting and searching index data, rather than “messaging around with boards and governance”.

Percentage of respondents on whether they support or oppose recommendation 7



**Recommendation 7**

*We should develop a model of a 'mixed economy' of public data so that everyone can benefit from some forms of two-way sharing between the public and the commercial sectors. Data that is derived from the activity of citizens must be seen as being at least co-owned by them and returning value to them, though the investment of business in collecting and processing the data should also be respected.*

*There are government initiatives such as Midata a government led project that works with businesses to give consumers better access to the electronic personal data that companies hold about them. The project recognises that data about citizens belongs to them and that they should have a way of claiming and using their ownership.*

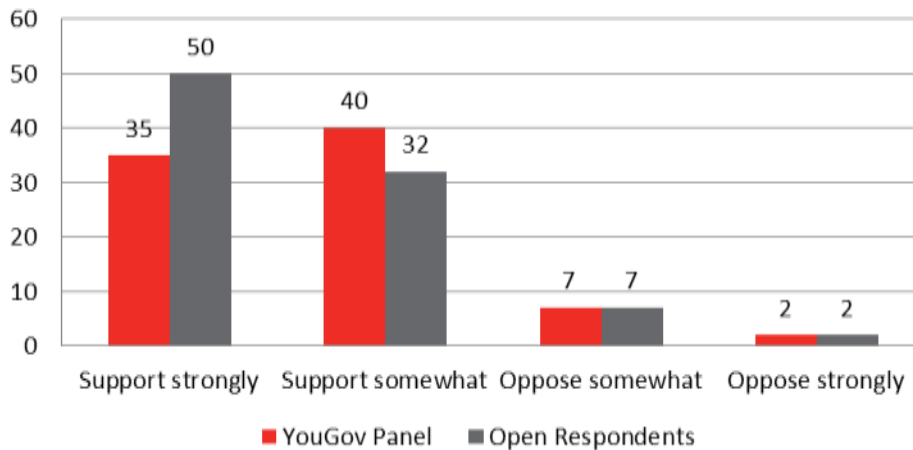
*Government should be able to make a claim on certain data, and make it open, if a strong case can be made for public value which is not by other means returning to the public.*

**Results**

As we can see from the table below, 77% of total respondents said that they strongly (39%) or somewhat (38%) support recommendation 7.

When we compare the two groups of respondents we can see that they are broadly supportive of recommendation 7. While more Open Respondents 'strongly' support (50% compared to 42%), more YouGov Panellists support this recommendation 'somewhat' (40% compared to 32%).

**Percentage breakdown of YouGov Panel & Open Respondents on whether they support or oppose recommendation 7**



Some respondents said that pushing too much for access to commercial company data could add to the costs of publishing, and impact small businesses. Other respondents questioned who would decide what commercial data should be made available, and whether firms should be compelled to release their data.

Respondents who opposed the recommendation outright stated that “the government should NOT have a claim over any privately held data, unless there is reasonable suspicion of criminal activity”. Others were not opposed, but emphasised that it was their right to know what information companies held about them.

**Recommendation 8**

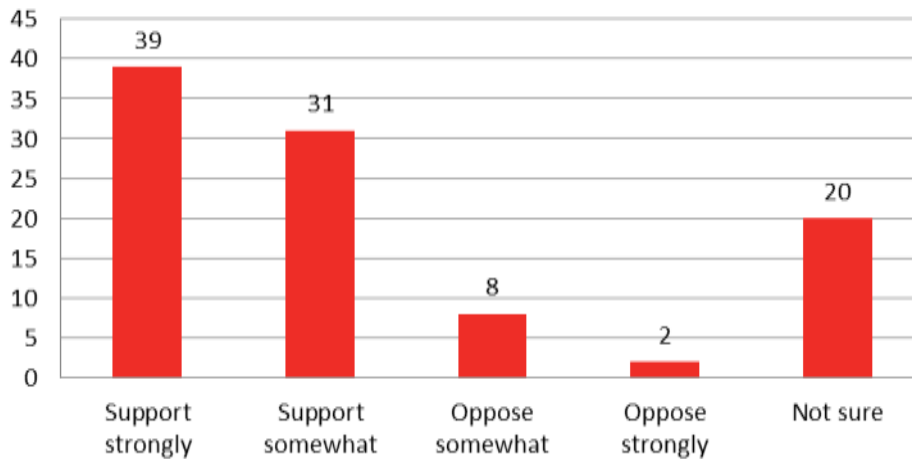
*We should challenge the current quasi-commercial Trading Fund model (for Companies House, Land Registry, the Met Office and Ordnance Survey) in favour of a basic information utility or scientific institution model, in which Trading Funds should be responsible for transparency of data production (that is, collecting and publishing data in a way that can be seen to be reliable and authoritative) and only provide 'added value' services where the market is likely to fail.*

*Currently the Trading Funds do a reasonably good job of collecting, using and sharing data. But many think it would be even better if they could focus on transparent collection and distribution, and where appropriate scientific processing, rather than holding on to it for quasi profit-making purposes.*

**Results**

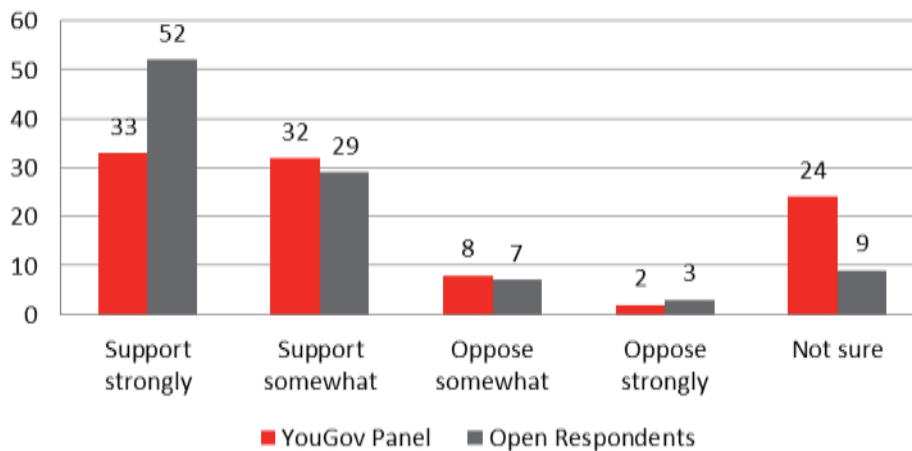
As we can see from the table below, 70% of total respondents said that they strongly (39%) or somewhat (31%) support recommendation 8. Interestingly 20% said that they were not sure.

**Percentage of respondents on whether they support or oppose recommendation 8**



When we compare the two groups of respondents we can see that they broadly share similar views. While more Open Respondents ‘strongly’ support (52% compared to 33%), more YouGov Panellists support this recommendation ‘somewhat’ (32% compared to 29%). By breaking the results down between the two groups, we can see that the significant number who indicated that they were not sure was driven by the YouGov Panellists (24% compared to 9%).

**Percentage breakdown of YouGov Panel & Open Respondents on whether they support or oppose recommendation 8**



Some respondents said that there should be a “reasonable cost” to businesses that will gain a commercial advantage from the data. Others who supported this view argued that any charge should be to cover the cost of acquiring and producing the data, rather than for active profit making.

Others argued that all the data held by trading funds “should be entirely free for any use”.

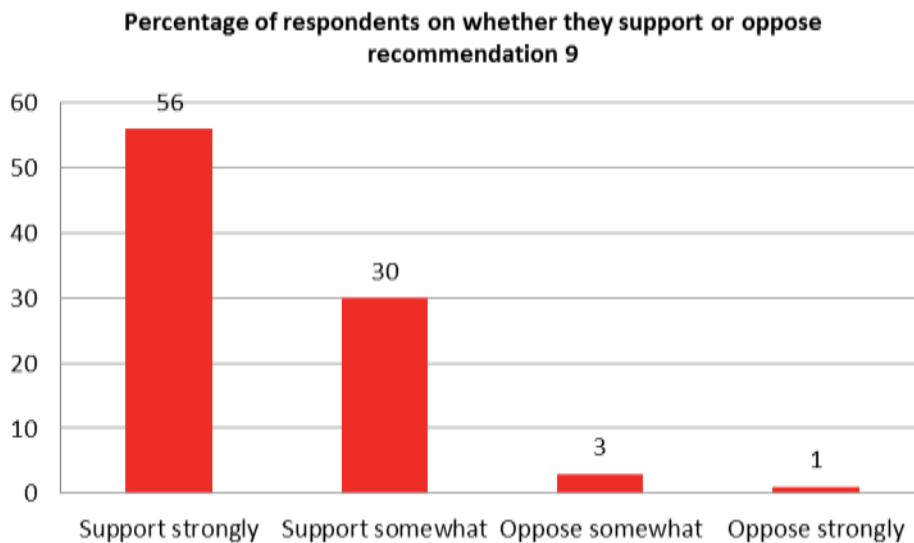
## **Recommendation 9**

*We should expect systematic and transparent use of data in the formulation, implementation and monitoring of government policy, and formally embed this in the democratic process.*

*Although government does publish some data as evidence for policy, for example in impact assessments, practice varies, and the wider consultation process is not generally considered to be effective. We should deepen and broaden the role of data in policy making.*

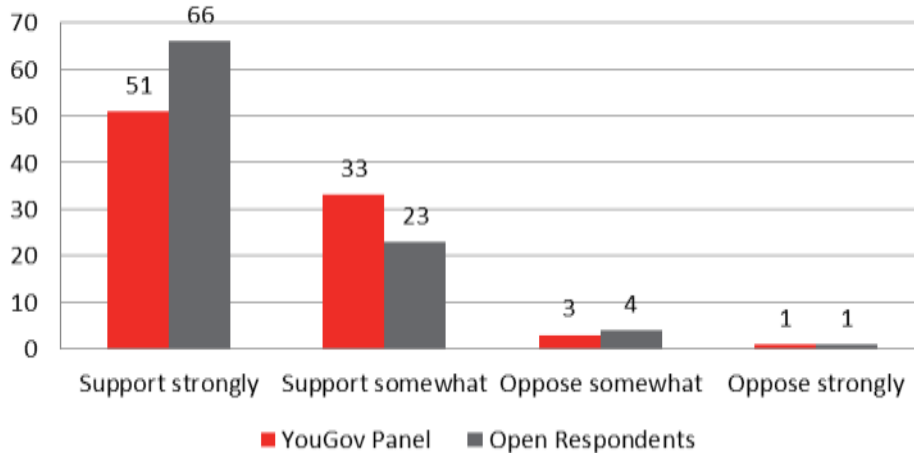
### **Results**

As we can see from the table below, 86% of total respondents said that they strongly (56%) or somewhat (30%) support recommendation 9.



When we compare the two groups of respondents we can see that they are broadly supportive of recommendation 9. While more Open Respondents ‘strongly’ support (66% compared to 51%), more YouGov Panellists support this recommendation ‘somewhat’ (33% compared to 23%).

**Percentage breakdown of YouGov Panel & Open Respondents on whether they support or oppose recommendation 9**



Most comments focussed on the fact that “data should support policy, not set it” because “data can be interpreted in different ways”.

**Recommendation 10**

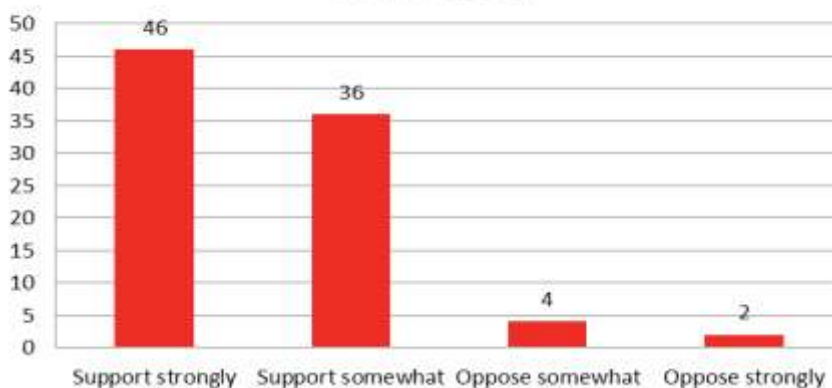
*We should look at new ways to gather evidence for the economic and social value of opening up Public Sector Information and government data. This evidence should be used to underpin a bold strategy of investment in an infrastructure of data in order to make the UK the world leader in this field, thereby gaining the greatest advantage in this new wave of the digital revolution.*

*Currently we can measure the costs of producing and publishing data, but we have no model for evaluating the economic or social benefits 'downstream', and so we may be undervaluing these activities, leading to underinvestment of resources.*

**Results**

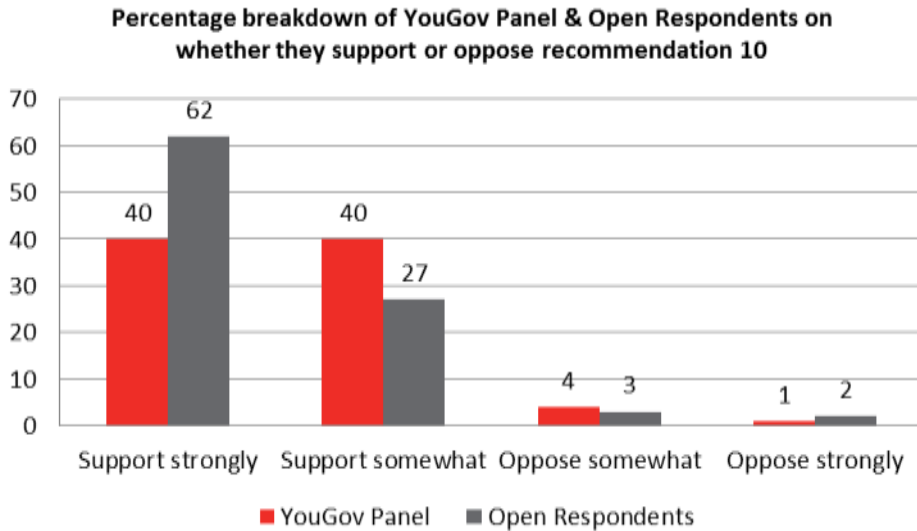
As we can see from the table below, 82% of total respondents said that they strongly (46%) or somewhat (36%) support recommendation 10.

**Percentage of respondents on whether they support or oppose recommendation 10**





When we compare the two groups of respondents we can see that they are broadly supportive of recommendation 10. While more Open Respondents 'strongly' support (62% compared to 40%), more YouGov Panellists support this recommendation 'somewhat' (40% compared to 27%).



Most comments centred on ensuring that a return could be achieved before heavy investment is made, rather than simply wasting taxpayer's money.

### **Concluding Thoughts**

The results indicate that all Recommendations received the majority support of both groups of respondents. It is interesting to note that the Open Respondents are far more supportive of the Recommendations (than the YouGov Panellists).