Open Data

Six Stories About Impact in the UK

Becky Hogge

November 2015

Commissioned by
Omidyar Network

OMIDYAR NETWORK
A WORLD OF POSITIVE RETURNS

In 2010, the year the United Kingdom launched its open data portal, a Transparency & Accountability Initiative report highlighted both the promise and potential of open data to improve services and create economic growth.

In the five years since, the UK's progress in opening its data has been pioneering and swift, but not without challenges and questions about impact. It's this qualified success that prompted us to commission this report in an effort to understand if the promise and potential of open data are being realized, and, specifically, to…

… explore and document open data's social, cultural, political, and economic impact;

… shine a light on the range of sectors and ways in which open data can make a difference; and

... profile the open data value chain, including its supply, demand, use, and re-use.

The report's author, Becky Hogge, finds that open data has had catalytic and significant impact and that, in time, will likely reveal even further value. She also flags critical challenges and obstacles, including closed datasets, valuable data not currently being collected, and important privacy considerations.

Omidyar Network has long championed open data as one key ingredient to more effective, efficient, fair governance and a more empowered and engaged citizenry. This report reinvigorates our commitment to this space and our admiration for the organisations and champions working daily to maximize the value of open data.

We are deeply grateful to Becky for the impressive rigour and depth with which she approached this research. We also thank all of the interviewees for sharing their valuable time and insights for this effort. We trust you will find it illuminating and helpful as we work together to realise open data's promise and potential.

Laura Bacon
Principal, Policy
Omidyar Network

Open data was sold to governments on a prospectus of potential, not proof. The policy of releasing government data in bulk, machine-readable formats and permitting its use and re-use without restriction was incubated in the UK and US[1] in the first decade of the 21st century. The policy has since spread rapidly across the world (through informal networks and also formal platforms such as the **Open Government Partnership**), driven by a number of convictions: that open data will seed economic growth; that it will deliver public sector efficiencies; and finally that it will enhance transparency and accountability both within government and beyond it. Now, those at the top of the **open data value chain**—government data publishers—as well as the major grant-giving institutions who have funded civil society groups to advocate and engage with the open data agenda, are asking: Has it worked?

The author recognises that the UK-focus and often technical nature of this report may require orientation for non-specialist or non-UK readers, and a glossary of terms (indicated in the body text **like this**) is supplied at the end of this report.

Helping to answer that question is the task of this study. It is not an easy one. For a variety of reasons, open data impact is hard to measure. Impact comes in many forms, and can be felt at many points along the **open data value chain**. Barriers to impact may be multiple and entrenched. The theory of change that initially drove government open data policies is resistant to traditional impact measurement. And it may yet be too soon to see some of the impacts we might expect.

Much of this report is, however, good news. In almost all of the six cases it details—all drawn from the UK, and including one "counterfactual" case of impact from re-use of data that was not open—the theory of change behind government open data policy is upheld. Private actors have taken government data, and they have transformed it in ways that are useful and valuable to citizens and consumers. Far more time and money has been invested in government data than it is possible to imagine the government ever having done by itself. The impact of this investment, though not always quantifiable, is in most cases tangible and scalable, if not already "at scale".

The cases

The six case studies outlined in this report each tell us something different about open data. Their selection is intentional: Each one not only has something to say about open data's success or otherwise as a policy, it also tells us something about the study of open data impact itself.

Transport for London

Transport for London (TfL) is the local government body responsible for implementing transport strategy and managing transport services across the UK capital. **TfL** was a cautious entrant into the open data fold. But now, five years after beginning to release its key datasets, many in real time, the apps built on the back of its data number in the hundreds, reach millions of London transport users, and deliver tens of millions in monetised time savings to its core customer base, all for relatively low investment. Open data thinking is now "embedded" in the organisation, and **TfL**'s experience

with open data has led to other national transport authorities following in its footsteps. This study highlights both how the market for developing apps based on transport data is highly responsive, and how organisations (and governments) considering open data reforms must take an initial leap of faith. It examines the role of **smartphones** in **TfL**'s decision to go open, and speculates about how the market for re-using transport information might change over time.

HM Land Registry Price Paid Data

Her Majesty's Land Registry (**HM Land Registry**) is responsible for maintaining the **Land Register**, where more than 24 million "**title**s" (evidence of land and property ownership) are documented. **HM Land Registry** began a phased release of its data on property transactions—the **Price Paid Dataset**—in March 2012, and by November 2013 the entire historic record dating back to 1995 was released. The data provides much-needed transparency in a historically "murky" business, and is already being used extensively by some traditional players in the property market. Additionally, new players are consolidating around the field of **proptech**, developing digital tools that use **HM Land Registry data**, often in conjunction with other, proprietary datasets, to bring buying and selling property "out of the Stone Age". The release has also improved data quality. But the Price Paid data is only a subset of what **HM Land Registry** could release. This study examines the impacts that flow from releasing **core reference data**, provides a window on the way **trading funds** like **HM Land Registry** have set back open data reforms in the UK, and sounds a warning note against future data privatisation.

Open Corporates and Beneficial Ownership Transparency

2013 saw the UK Government concurrently host the **Open Government Partnership (OGP)** summit and hold the presidency of the G8. Prime Minister David Cameron used this joint platform to launch a positive commitment to **beneficial ownership transparency** (identifying the real owners of corporate entities—a key tool in eliminating the ability of anonymous shell companies to mask financial corruption) that has the potential to drive global norms on the issue. Behind the scenes, a coalition of civil society groups focussed on corruption, fraud, overseas development, and tax justice were driving the agenda, aided by open corporate data pioneers OpenCorporates. OpenCorporates was able to leverage its expertise in order to play a critical, if minor, role in the success of the campaign for **beneficial ownership transparency**. This is a case about the contribution open data has to make to advocacy efforts on complex issues, and illustrates how moving the needle on complex issues like corruption and governance reform requires much more than opening government data. It further argues that the role of the open data community should not just be to use the government data it is given, but to demand the government data it needs.

The Open Public Services Network and school performance in England

The **Open Public Services Network (OPSN)** is a programme based at the **Royal Society for the Arts** that uses data to encourage debate about the quality of public services in ways that engage and empower service users. **OPSN** undertook two major pieces of work exploring how government data, beyond traditional accountability mechanisms like **league tables**, can be used to evaluate schools. The resulting outputs—*A* **GCSE** *Schools Guide*[A] produced in association with the *Guardian*, and a report into lack of access to "hard" subjects like **triple science** at **GCSE** in deprived areas—succeeded in reaching both parents and policymakers, and drew significant press attention. This study focuses on the contribution open data can make to improving public services, and the crucial role non-open **personal data** has to play. It highlights the lack of sustainable business models for organisations that can interpret government data on behalf of the public.

TheyWorkForYou.com

TheyWorkForYou.com is a parliamentary monitoring website that has been running for more than 10 years in the UK. It provides an accessible, searchable version of the official record of proceedings of the UK **Parliament**, as well as the devolved **Northern Ireland Assembly**. It also provides analysis of the voting records of elected and **non-elected representatives**. One of **TheyWorkForYou.com**'s long-term impacts appears to be encouraging parliamentarians to vote less with their party and more in the interests of the people they represent. It is also important to consider the potential monetised time savings delivered to the site's users—many of whom are civil society groups and journalists. These might well be in the millions of pounds a year, although methodological constraints mean it will probably always be impossible to put an exact figure on them. **TheyWorkForYou.com** pre-dates open data reforms in the UK by a number of years, and is included in this report as a window onto the long-term impacts of outputs based on public data.

Coeliac UK and Brandbank

Gluten Free on the Move is an award-winning app designed to help coeliac sufferers adhere to the strict **gluten**-free diet that is the only treatment for their condition. It is the result of a data-sharing agreement between **Coeliac UK**—a patient self-help society—and **Brandbank**—Europe's largest provider of digital content for online retail portals. The **Brandbank**/**Coeliac UK** partnership shares several features of other cases in this report, but exhibits one important difference: The data is not open data, **Brandbank** licenses it to **Coeliac UK**. This case highlights the importance of considering the counterfactual when evaluating open data impact. Those examining open data impacts should be ready to ask the questions: Are these

A GCSE (short for General Certificate of Secondary Education) is the most common type of exam taken by school children in England at age 16. A glossary of technical and UK-specific terms is provided at the end of this report.

impacts only happening because data is open? What else about this dataset is contributing to these impacts? The case further examines whether moving to an open data model is a feasible option for **Brandbank**.

Conclusions

Taken together, the six cases demonstrate that the theory of change that drives government open data policy is working, and make a strong case for continuing to pursue the policy. But they also surface a number of issues. The first is that there are significant data gaps in the UK that may be acting as barriers to impact. These gaps divide roughly into three categories.

There are data gaps where data is collected by government agencies but is not yet released openly. Policymakers need to make good on their commitment to open data, and release the **core reference data** needed to guarantee that the UK's **national information infrastructure** exists as open data.

There is data that is not yet collected by any government agency but could be. Where data gaps like this exist, policymakers and open data advocates need to be ready to collaborate with other domain-specific stakeholders to create the "**participatory data infrastructures**"[2] that will close them.

And there are data gaps where data is collected by government agencies but is not suitable for release as open data because it contains (often sensitive) personal information. While it remains true that most **personal data** should never be released as open data[B], policymakers, open data advocates and open data developers cannot avoid engaging with the **personal data** agenda anymore. They should be ready to work with experts in the fields of data anonymization and re-identification, consent, and data-handling, in order to chart a way through this space.

Beyond data gaps, the second issue surfaced by the cases is that organisations transforming government data into social and political impacts need some help to achieve long-term sustainability. Although it's true that the rise of the internet and computer processing power is driving down the cost of running organisations that scrutinise government activity, those organisations also need to invest in people with a unique mix of data skills and governance knowledge. Prime Minister David Cameron predicted open data policy would trigger a volunteer "army of effective armchair auditors[3]" who would interrogate government data. In fact, that army has been slow to advance.

The third important issue surfaced by the cases is that even where both the open data and the expertise needed to interpret it exist, in some cases substantial advocacy work beyond the activities of infomediaries is needed to achieve impact. This last point is relevant to the entire open data community. In some domains, simply releasing data openly can have a huge impact. But in very many others—tackling corruption, improving mental health outcomes, arresting climate change, achieving gender equality—the problems we might aspire to solve are deeply complex. Expecting open data to fix them all by itself is only inviting disappointment.

The conclusion to this report sets out these findings and appropriate responses to them in greater detail. It further calls on policymakers and funders to invest in research to understand the nascent information markets open data is helping to create, how these markets might consolidate, and how the public good can be safeguarded if and when they do. Finally, it predicts a new phase of work for open data advocates, one that to some extent sets aside the collaborative approach that has characterised open data reforms in the UK to date, and instead makes calls on government for further open data reforms in the language of rights, and not economic policy.

B The author sides with those in the open data community who define open data in opposition to personal data. Nonetheless, the author recognises that in some circumstances, for example public interest registries, parliamentary records, and company ownership, personal data can and should be published as open data. For an in-depth discussion of open data and privacy, see (Zuiderveen Borgesius, Gray, & van Eechoud, Forthcoming).

Table of Contents

Introduction

Open data was sold to governments on a prospectus of potential, not proof. The policy of releasing government data in bulk, machine-readable formats and permitting its use and re-use without restriction was incubated in the UK and US[4] in the first decade of the 21st century. The policy has since spread rapidly across the world (through informal networks and also formal platforms such as the **Open Government Partnership**), driven by a number of convictions: that open data will seed economic growth; that it will deliver public sector efficiencies; and finally that it will enhance transparency and accountability both within government and beyond it. Now, those at the top of the **open data value chain**—government data publishers—as well as the major grant-giving institutions who have funded civil society groups to advocate and engage with the open data agenda, are asking: Has it worked?

Helping to answer that question is the task of this study. It is not an easy one. For a variety of reasons, open data impact is hard to measure. This report joins a growing body of research into open data impact[C], every piece of which comes prefaced with some variety of the observation that "measuring the impact of open data is perhaps the most important and most difficult task[5]". Impact comes in many forms, and can be felt at many points along the **open data value chain**. Barriers to impact may be multiple and entrenched. The theory of change that initially drove government open data policies is resistant to traditional impact measurement. And it may yet be too soon to see some of the impacts we might expect.

But the six case studies outlined in this report—all of which centre in the UK—do tell us something about open data. Their selection is intentional: Each one not only has something to say about open data's success or otherwise as a policy, it also tells us something about the study of open data impact itself. In choosing to focus on the stories contained in this report, I have tried to spread the types of impact achieved across three broad categories: 1) social; 2) political/governance; and 3) economic/commercial.

This report takes the UK as its focus[D]. The UK is recognised as a world leader in open data, and its progress against the open data agenda—and the results of that progress on the ground—are therefore valid as a subject of study for an audience beyond its borders. At the same time, situating six studies in the same geographic, economic, and political context subtly underlines the importance of context[E] when evaluating open data's impact, even while the lessons learned from each study may translate to different contexts in varying degrees.

What follows in this introduction is a brief overview of the UK's open data policy, the arguments that drove it, and the UK's performance as an open government (in data terms), as viewed by the rest of the world. The introduction then turns to a discussion of issues in measuring open data impact, drawing on examples from each case study. The conclusion of this report looks towards the future, and discusses the challenges the open data movement can expect to face over the coming years.

The majority of this report is dedicated to the cases themselves. The first two focus on government data publishers (**TfL, HM Land Registry**). The next three take as their starting point the private actors—so-called **infomediaries**—who have taken data published by government and transformed it (OpenCorporates, the **Open Public Services Network**, **TheyWorkForYou.com**). The final case, which details a data partnership with significant health impacts that occurred outside of the open data space (**Brandbank** and **Coeliac UK**), is included as a nod to the counterfactual, a reminder that a good study of any event's impact should always ask: Are these impacts happening only because data is open? Or is there something else about this dataset (its relevance, its richness) that is contributing to these impacts?

C In particular, I am indebted to work done at the Sunlight Foundation by Júlia Keserű on social impacts (Keserű, 2015), to the work of Tim Davies and others at the Worldwide Web Foundation on impacts in developing countries (Davies, 2014), to the work that went towards formulating the Common Assessment Framework for Open Data (Caplan, Davies, Wadud, Verhulst, Alonso, & Farhan, 2014), and to work at the Open Data Institute on benchmarking open data (Atz, Heath, & Fawcett, 2015).

D In fact, the focus of this report might be even more narrowly defined as "England and Wales", given most of the data profiled in the cases has very little to say about the UK's devolved nations of Scotland and Northern Ireland.

E Evgeny Morozov's critique of the open data agenda—that local politics matters more than whether a particular government "lived up to global and supposedly neutral standards of 'open government'" is relevant here (Morozov, 2013 p71)

The author recognises that the UK-focus and often technical nature of this report may require orientation for non-specialist or non-UK readers, and a glossary of terms (with terms indicated in the body text **like this**) is supplied at the end, along with references, a list of interviewees, and an appendix detailing the methodology behind this research.

Open data and the UK: Pioneer at risk?

In the UK, champions of open data can be found at the highest levels of government and in the dustiest corners of **Whitehall**. Established business leaders, entrepreneurs who run their startups from laptops in **Shoreditch** coffee shops, and earnest members of civil society groups support it in equal measure. This is not to say that there are no detractors from the policy at the grassroots, or that nobody at the top has used their power to block it. Rather, it is simply to state that open data policy in the UK has a uniquely broad appeal. That appeal can be put down in part to the dual nature of the promise behind open data[F].

Firstly, open data promised to deliver economic growth and public sector savings. Policy advocates argued[6] that, in a digital world, governments should release information in the form of data published in bulk, re-usable formats, rather than concentrating government resources on producing discrete information products tailored to the specific needs of groups of citizens and consumers. By doing this, governments would stimulate a market for data re-users (**infomediaries**) that would end up fulfilling those specific needs much better than governments could, and for much less government expenditure.

Secondly, open data promised to enhance transparency, accountability, and civic participation. The rise of the internet and computer processing power was putting powerful new ways to interrogate and interpret government data in the hands of individuals. A new breed of civil society actor was being born, the **civic hacker**, who could take government data and transform it in ways that enriched the public sphere.

In January 2010, the then **Labour** government led by Gordon Brown launched **data.gov.uk**, a portal of government data available for re-use[G]. Work on government open data policy continued after the election of a new **Conservative/Liberal Democrat** coalition in May 2010, and in September 2010 the first **Open Government Licence** was launched[H]. The UK is one of eight founding members of the **Open Government Partnership**, and published its first National Action Plan under this rubric in September 2011. Later that year, the **Chancellor** announced funding for an **Open Data Institute (ODI)**, tasked with helping deliver on the economic and social potential of open government data, and that institution officially opened in 2012. In 2013, as part of its presidency of the G8, the UK promoted a transparency agenda that saw G8 leaders sign up to an open data charter. Individual departmental commitments to open data and new government initiatives to promote its adoption and stimulate its use multiplied in subsequent years[I].

The UK tops two international open data **league tables**. The Worldwide Web Foundation's Open Data Barometer put the UK at the top of its two most recent rankings, in 2013 and 2015, although its 2015 report implied that the sale of the **Postcode Address File** during the privatisation of **Royal Mail** (see box) compared poorly to France's "willingness to focus on the availability of high value datasets[7]". The Open Knowledge Foundation's Open Data Index, which tracks the availability of government datasets across 10 different spheres including transport, budget, and mapping data also ranked the UK number one overall in 2013 and 2014, and number one in seven of the 10 individual data spheres[8].

But some open data advocates believe these rankings belie waning enthusiasm for the policy at the top of government. In

The Privatisation of Royal Mail and sale of the Postcode Address File

In June 2013, the UK's then Business Secretary Vince Cable announced that **Royal Mail**, the UK's publicly owned postal service, was to be privatised. A majority of shares were floated on the London Stock Exchange in October 2013, with 10% of shares given to **Royal Mail** employees and the government maintaining a 30% stake. In June 2015, the **Chancellor** George Osborne announced the government would sell the remainder of its stake in the company.

The **Postcode Address File (PAF)** is a database maintained by **Royal Mail** that holds a record of all known delivery points in the UK, including domestic addresses but also commercial premises. It totals 29 million addresses across 1.8 million UK postcodes, "constantly updated and verified," according to **Royal Mail**[9], by 90,000 postmen and women. It is used by "everyone from marketing firms to emergency services", who each pay a fee to **Royal Mail** to access the data.

Ahead of the privatisation of **Royal Mail**, open data advocates urged the government not to bundle the **PAF** with the sale[10]. They argued that the data would contribute more to the economy if it was made available as open data, and that it was part of the UK's **national information infrastructure** and should therefore be publicly owned. These arguments were also being made by open data champions at the **Cabinet Office**. But they did not win out: The **PAF** was eventually included in the sale.

In March 2014 **Parliament**'s **Public Administration Committee** strongly criticised the government for selling off the **PAF**, arguing that it had done so to boost the company's valuation, ignoring the expense that had gone into creating it and the direct value it offered the economy. The chairman of the committee said:

The sale of the PAF with the Royal Mail was a mistake. Public access to public sector data must never be sold or given away again.[11]

F My thanks to Jonathan Gray for pointing out that "where open data comes from and the objectives that are associated with it … is both highly contingent and (at least a bit) contested" (personal communication). For a more thorough treatment of the origins of open data, see (Gray, Towards a Genealogy of Open Data, 2014)

G For a history of the policy development process that led up to this move, see (Hogge, 2011)

H Unlike in the US, there was no historic assumption in the UK that government information should be free of intellectual property rights. Prior to the launch of the OGL, much government information had been published under Crown copyright.

I A useful, if incomplete, collaboratively edited timeline of open data in the UK is curated by Tim Davies (Davies, UK Open Data Timeline [to Feb 2014] n.d.)

an August 2015 interview, Sir Nigel Shadbolt, one of the policy's key independent instigators and a co-founder of the Open Data Institute, warned that some policymakers may be taking their "eye off the ball", saying the UK was still "a long way away from … the routine availability of data that provides a resource to make stuff happen[12]".

Many of the **core reference datasets** that make up the data infrastructure in the UK are administered by **trading fund**s, government agencies that meet their outgoing costs from the money they charge for services. Some of these agencies, notably the mapping agency **Ordnance Survey**, gain the majority of their receipts charging for data. The **trading fund** model has been a sticking point in the formation of government policy around open data from the beginning. Now, as the **Conservative** government elected in May 2015 considers selling off these agencies to boost the public purse, the future of open data is in the balance. The sale of the **Postcode Address File** along with **Royal Mail** in 2013 (see page 7) set a poor precedent in this area.

What is impact and how can we measure it?

The draft Common Assessment Framework for Open Data[13] (CAF) defines impact as:

The benefits to be gained from using the open dataset. Potential benefits can be studied according to social, environmental, political/governance, and economic/ commercial dimensions.

Note that benefits are not gained but "to be gained" and potential benefits still represent, according to the CAF, a valid subject of study. By contrast, the terms of reference of this report (see Appendix 1: Methodology and interview frameworks) demands that it identify "tangible impact at scale," a considerable departure from the approach of the CAF.

A slightly different approach is put forward by researchers in Brazil[14]. They identify a process that leads to open data impact as follows: Resources > Activities > Output > Outcomes > Impact. This process might be narrated thus:

1. Open data is published (resources)
2. People access the data and transform it (activities)
3. This results in new data-driven products (output)
4. Other people use those products (outcomes)
5. This has some effect in the world (impact)

The theory of change that drove the adoption of open data policies by the UK and US governments predicts this five-stage process. Rather than trying to determine how citizens and consumers want to interact with government data, government publishes the data in raw format and the market (including NGOs and nonprofits with non-economic goals) transforms the data to serve a variety of audiences. The CAF recognises the challenges this theory of change presents to those researching open data impact:

Research designs to evidence impact generally need to be specific to targets and goals, making them tricky to implement for open data, given one of the potential advantages of open data is that it allows many possible benefits, including unanticipated benefits.[15]

What all this means is that where other fields might start at stage 3 of the impact journey (output), open data interventions have two preliminary stages (resources, activity) to be considered. Asking about open data impact therefore means asking two separate questions. Firstly, was the theory of change correct, that is, did private actors take the resources offered and transform them into citizen- and consumer-facing products? And secondly, what impact did these products have?[J]

The first question is easy to answer, the second less so. Once we know what products have been developed (the output) we can guess who might use them (outcome) and what they might achieve by doing so (impact). Many of the UK's most quoted open data impact stories are happy to stop there. For example, in 2012 upwards of £250 million in annual savings on statin prescriptions were identified by a group of **civic hackers** working on newly released NHS prescriptions data, including big data startup Mastodon C[16]. No follow-up work has been done to establish whether any of these savings have actually been achieved, savings that would rely on changing the behaviour of individual GPs. Despite this, three years on, open data proponents still quote the story as an example of open data's impact[17].[K]

How can we get from identifying outputs to measuring tangible impact? Usage statistics for particular products may be available (although often they are not) in which case researchers can get at more quantitative outcomes. But that final leap, from outcomes to impact, remains for the most part in the realm

J A potential third set of questions might coalesce around the impact of the data release on the data publisher itself: Did it improve data quality? Did it transform internal practice? Did it inspire sister organisations to publish their data openly and with what subsequent effects? Impact here may be a result of any of stages 1-4 (Resources/Activity/Outcome/Use) of the 5-stage process described.

K (Deloitte, 2013) examines the NHS prescribing case in detail, including that Mastodon C were already aware of prescription savings on statins in the hundreds of millions of pounds thanks to a paper published in the British Medical Journal in 2010 (Moon, Flett, Godman, Grosso, & Wierzbicki, 2010)

of best guess. Sometimes those guesses come with numbers attached, but it is important to recognise them as guesses nonetheless.

There are several reasons for this. One is that causality is hard to prove. Drawing on work by the Institute of Development Studies[18], Sunlight Foundation's Júlia Keserű[19] writes:

Much of the existing literature seeking to measure the impact and effectiveness of transparency and open data accountability initiatives seem to face a common challenge: it is incredibly difficult to come up with definitive, evidence-based generalizations about how "x" type of initiatives produce "y" kinds of effects.

This difficulty is evidenced in this report in the **TheyWorkForYou.com** study. It suggests a link between the existence of **TheyWorkForYou.com**, its widespread use by members of the UK electorate, and an increase in vote attendance and **rebellion** by **Members of Parliament**. But as **mySociety**'s Director of Research Rebecca Rumbul points out, such a link "[could] only be established and validated by conducting a fairly large qualitative study"[20].

Secondly, as researchers at the **Open Data Institute** observe[21], "in some cases, the **open data value chain** can be extensive". Impact in this sense may be diffuse, and there may be methodological constraints in accessing data about those parts of the value chain benefitting the most from a particular product predicated on open data. So, for example, Deloitte extrapolate a value for the amount of time saved by London's commuters thanks to **TfL**'s data from three sources: an estimate of the number of times apps based on the data have been downloaded; a benchmark published regularly by **TfL** on Lost Passenger Hours due to delays; and calculations of the value of transport users' time devised by the Department of Transport[22]. We could imagine that a similar approach might be used to calculate the value of time saved by the users of **TheyWorkForYou.com**, yet the benchmarks we would need to calculate such a value (how much is a campaigner/journalist/researcher's time worth? How much time was lost to trawling through unwieldy government information resources before **TheyWorkForYou.com** came along?) may never be available.

It's also important to consider the role of other factors on the path to impact[L]. Although it may be an appropriate moment in the policy cycle to evaluate open data impact, many of its benefits are yet to bear fruit. Whereas some fields (and the **TfL** study shows us transport is one) might see immediate impact from making data available openly, others will take longer to show changes. This could be down to gaps in the data (mentioned in the case study on **HM Land Registry)**, gaps in the market for data re-users (as outlined in the Open Public Service Network case study), or the existence of other factors acting as blocks to impact (explored in the study on OpenCorporates).

Finally, it could well be the case that open data has not yet had the time or political investment it needs to achieve everything it can even by itself. A report commissioned by Omidyar Network on the potential of open data to boost economies notes that open data may "exhibit self-organized criticality":

The important mathematical property of such a system is that any addition always has the potential to cause the entire network to jump to a new state in which the connections and the payoffs change dramatically, perhaps by several orders of magnitude.[23]

The authors go on to warn against underestimating the power of open data in the long term:

Underestimations of this type seem to have been relatively common in relation to information and communication associated activities. Among the examples here are the demand for computers, the effect of mobile phone technology, the time required to sequence DNA and so on.[24]

All of these issues merit significant further study in themselves. In the end, although we may be able to speak of the potential of open data in sweeping terms, the proof of open data's contribution to economies and societies may only be visible in millions of tiny fragments. It is to a handful of those fragments that this report now turns.

L For a deeper discussion of this issue, see (Pollock, 2012)

Transport for London: Get set, go!

Summary

Transport for London (TfL) was a cautious entrant into the open data fold. But now, five years after beginning to release its key datasets, many in real time, the apps built on the back of its data number in the hundreds, reach millions of London transport users and deliver tens of millions in monetised time savings to its core customer base, all for relatively low investment. Open data thinking is now "embedded" in the organisation, and **TfL**'s experience with open data has led to other national transport authorities following in their footsteps.

Key takeaways

- **Transport data is very amenable to the open data approach. The market for developing apps based on transport data is highly responsive.**
- **TfL was a cautious entrant to the open data fold. The decision to open TfL's data was in large part an experimental one, since the business case for open data was hard to model. The risk paid off: TfL is now converted to open data, and has been able to demonstrate the benefits of an open data policy to other stakeholders in the field who may never have been in a position to take the initial leap of faith themselves.**
- **TfL knew its customers increasingly wanted to access information about transport services across a wide variety of smartphone platforms. This was a key factor in the decision to move to open data, since the alternative—developing apps in-house that served every smartphone platform—would have been an expensive undertaking.**
- **Innovation in the transport apps market may slow once personal data becomes a more important part of the mix. Apps like Google Now have the potential to lock users in to transport data-based services thanks to location data Google collects that will be unavailable to competing providers. At the same time, TfL is developing smartphone services that include a payment element and therefore must be developed in-house to keep users' payment details secure.**

Background

Transport for London (TfL) is the local government body responsible for implementing transport strategy and managing transport services across the UK capital. It oversees almost all aspects of transport in Europe's largest city[25], with 24 million journeys made across London's transport network every day[26].

As well as managing London's buses, the **Tube** network, **Docklands Light Railway**, **Overground**, and **Tramlink**, it runs the city's cycle hire scheme, its River Services, coach station, and the Emirates Air Line cable cars that cross the River Thames to the east of the city at Greenwich. It controls the city's 6,000 traffic lights and a 580km network of main roads. It regulates London's taxis and private hire vehicles, and runs the city's **Congestion Charge** scheme[27].

TfL is part of the **Greater London Authority (GLA)**. It is publicly owned, and is governed by a Board of Directors chaired by London's mayor. It is funded by "farepayers and taxpayers"[28]. In 2014/15, nearly half (47%) of its £10.9bn funding was derived from fares and other income (e.g., the **Congestion Charge**). A quarter (25%) was from grant funding via the UK's Department for Transport and the **GLA**, and the remainder was made up of borrowing, cash movements, and **Crossrail** funding[29].

TfL enjoys a reputation as an innovator in the field of transport services, and the scale of its operations means early investment in new technology often makes good business sense[30].

The data

TfL makes available 62 separate datasets[31]. These are a mix of real-time feeds (such as **Tube** departure boards, live traffic disruption, live bus arrivals, and **TfL**'s Journey Planner **API**), fixed datasets (such as timetables, station locations, and station facilities) and transparency-oriented datasets (detailing operational performance, directors' remuneration, etc.).

TfL requires data re-users to register with them in order to access any of their data. During the registration process, users agree to a set of licensing terms that, while based on version 2.0 of the **Open Government Licence**, contain some important additional conditions[32]. As well as setting fairly reasonable limits on the demand (number of **calls**) any one user can make on the data APIs, these conditions coalesce around protecting **TfL**'s branding and not passing off any products created as official **TfL** products. Users must also give **TfL** accurate information about their intended use of the data before being granted access to it. In this respect the data released by **TfL** does not conform to the **open definition**. Nonetheless, internally and externally, **TfL** refer to it as "open data", as do observers in general. When questioned on this point, Phil Young, head of **TfL** Online, responded:

As far as developers are concerned, I think they would consider it to be open data unless they're particularly narrow in their view of what they think that open data is. Really [what we stipulate is] incredibly light, and the fact that we probably get more developers working on our stuff, and more apps created on our stuff, than anywhere else possibly in the world, would suggest that it's … fairly pen.[33]

TfL's website specifies how regularly each data feed is updated, ranging from every 30 seconds (**Tube** departure boards) to annually (**London Underground** Passenger Counts data). The **TfL** data offer attempts to exclude any form of personal information. Nonetheless, one dataset did appear to pose a privacy risk: In April 2014 software engineer James Siddle demonstrated how cycle hire use statistics connected to Customer IDs could theoretically be de-anonymised in the presence of "any seemingly innocuous personal signal" (such as a Foursquare check in, Facebook post, picture, or tweet linking an individual to a cycle hire location), leading to the exposure of "a detailed record [of] someone's life in London"[34]. **TfL** said that including Customer IDs in the data had been an administrative error[35]. They have since been removed.

The path to open

TfL's journey to open began in 2007, when the development team, led by Phil Young, released a set of embeddable **widget**s. These "code snippets" allowed users to integrate **TfL** online products like live travel updates into popular web content aggregator services like **NetVibes** and **iGoogle**, as well as custom-designed websites. The releases were part of a strategy to encourage customers to check the status of **London Underground** lines at weekends, as the network was undergoing an intense program of improvement works.

Discussing his team's motivation to develop embeddable **widget**s in 2007, Phil Young points to trends among other data teams working in the public service space, and specifically BBC Backstage, a now defunct developer network co-founded by Tom Loosemore (who, as a former trustee of UK Citizens Online Democracy, the charity that runs **mySociety**, is a key player in another open data project profiled in this report: TheyWorkForYou), that had begun to experiment with releasing the BBC's scheduling data:

We never really engaged with the BBC on it, but we were observing what was going on. We were a small digital team of keen developers, so we were just as interested in this world as everyone else, and the things that could be done. And we quickly saw that our data was probably more interesting than [the BBC's] data.[36]

In 2009, recognising that web developers wanted **TfL** to go further, Phil Young and his team launched a dedicated area on the **TfL** website for web developers[M]. A timeline of **TfL**'s data releases is given in Figure 1. Vernon Everitt, Managing Director in charge of Customer Experience, Marketing and Communications and **TfL**'s open data champion, describes the journey so far thus:

2007 Launch of embeddable 'widgets' for live travel news, map and Journey Planner.

2009 Special area for developers launched on **TfL** website.

2010 **London Datastore** launched. Additional real-time feeds launched with hundreds of developers registered.

2011 **London Underground** train location and Journey Planner **API**s launched. Registered developers rise to over 1,000.

2012 Live bus arrivals **API** launched, full London 2012 Olympic and Paralympic Games transport data portal. Over 4,000 developers registered.

2013 Over 5,000 developers, 30 data feeds, hundreds of apps on the market serving millions of customers. New accessibility and roads feeds added.

Figure 1: Timeline of **TfL** *data releases.*

Between 2007 and 2010 we were feeling our way a bit. And then by 2011 we'd got the hang of it and seen that not only do you have to make the data freely and openly available, you had to do it in a form that people could consume straight-forwardly. Hence the development of more sophisticated **API***s so people could plug in and play. And then in 2012 our bus departure* **API** *was launched, and we did a whole bunch of stuff for the Olympics which gave it added impetus*[37].

TfL now view accurate, real-time travel data as a complement to transport infrastructure in their overriding goal of serving London's transport users.

M The modern version of this area is available here: https://TfL.gov.uk/info-for/open-data-users/

Transport for London: Get set, go!

Vernon Everitt and Phil Young agree that "the clear policy of [the **GLA**] helped **TfL** to prioritise the release of data and achieve it faster than would otherwise have been the case[38]". Vernon Everitt observes that "no-one needed to persuade our political masters at the **GLA** that this was a good idea because their default setting was already openness[39]".

Emer Coleman had arrived at the **GLA** in 2009 on secondment for a year from Barnett council (one of London's 32 boroughs). Tasked with responding to imminent cuts to public sector funding and a new focus on government transparency, she was seeking a way for London boroughs to collaborate. One of the policies she devised was a proposition around open data that would eventually become the **London Datastore**. This open data portal would tease out collaboration across London's boroughs, and respond both to internal pressures to save money and stimulate economic growth in the city, and to external demands coming from open data enthusiasts, and particularly the *Guardian* newspaper's Free Our Data campaign (which had been running since 2006), to put public data in public hands.

> "**TfL** have come on a tremendous journey"
>
> Emer Coleman, **Greater London Authority**/TransportAPI.

Ahead of the **London Datastore**'s release, Coleman issued an open invitation to potential users of the portal in the developer community, with the help of Paul Clarke, a well-known figure in the government open data community who was then working as a contractor at the precursor to **gov.uk**, **DirectGov**. The two events that followed attracted between 60 and 100 participants, and calls for data were overwhelmingly focussed on crime and transport. Clarke remembers the collection of people who attended:

It wasn't just the armchair enthusiast or the casual hacker, or the train enthusiast. There were people there who were very serious about trying to build businesses out of reusing and adding value to public datasets.[40]

It was becoming clear that any launch of the **London Datastore** had to have **TfL** data in there. But Coleman describes a reticence at **TfL** to release their data openly that persisted "right up to the wire", as she planned the launch of the portal:

They didn't want to. […] **TfL** *have come on a tremendous journey. It's in a very positive place now, so I don't want to keep harping back to how difficult it was then. But it's fair to say that there were a number of concerns. They wanted to monetise the data. [And] they had a concern around accountability …. So we did a lot of work explaining to them, well, that's not really how the economics of this work.*[41]

What appeared as reticence from outside the organisation, was experienced inside **TfL** as reasonable caution. "You have to remember," observes Vernon Everitt, "that transport authorities like to control things". Phil Young tells the story from his perspective:

There was a range of views inside the organisation … and people with those views held them for a reason, there was

some logic behind them …. It probably took about a year of discussions, debate, working it through with the **GLA**, *with Emer …. We did get to the place where the argument was over, and the course was set.*[42]

Today, Everitt, Young, and Coleman agree that the economics of developing information services for customers increasingly wanting access via **smartphone**s was a crucial factor in **TfL**'s decision to release their data. Channels for consuming **TfL** data were set to diversify quickly, with travellers wanting to access information services on the move:

It was likely the authority was going to spend a lot of public money trying to design apps that wouldn't meet the consumer demand and the money, the revenue that was generated, would be small. Whereas the knock-on benefit to the travelling commuter [of releasing the data openly] was going to be huge, which would reflect back on **TfL**. *And that's actually what happened.*[43]

But at the time, **TfL** was conscious that the decision to go open was in a large part an experimental one. If Everitt had tried to sit down and write a conventional transport business case, he says he would still be writing it:

You couldn't prove in advance what making the data open was going to lead to. So eventually what we decided was we'd just go for it and see what would happen.

Outcome

TfL's main aim in releasing their data freely was to spur the development of apps in the market. For the policy to be a success, they needed existing or new businesses to develop new products and services based on **TfL** data, apps that served **TfL**'s customer base, and responded to transport users' growing demand to access data about **TfL** transport services via **smartphone**.

In 2010, the year after the launch of the special developers' area of the **TfL** website, the number of users registered to consume **TfL** data was in the hundreds. The following year, 2011, it numbered more than a thousand. In 2012, the number had risen to over 4,000 and by 2013 over 5,000 users were registered to consume and transform **TfL** data[44].

TfL are able to put exact figures to how many developers are accessing their data because developers need to register with them in order to access that data. However, beyond this, **TfL** cannot ascertain directly how many apps are using their data, or how many users these apps are reaching. Examining data download and access statistics could be misleading, because "many developers feed their app estate from their own server architecture and have a single connection to **TfL** data feeds"[45].

Each year, **TfL** attempts to ascertain take-up of its data indirectly, by counting the apps using the data across the major **smartphone** platforms. The latest count, done in November 2014, showed 362 **smartphone** apps using **TfL** data[46].

In a report published in May 2013, Deloitte provided analysis of how many people have downloaded apps powered by **TfL** data, based on a proprietary research tool provided by xyo.net. They estimated that such apps had been downloaded nearly 4 million times in 2012 (3,979,300)[47].

> "You couldn't prove in advance what making the data open was going to lead to. So eventually what we decivded was we'd just go for it and see what would happen."
>
> Vernon Everitt, **TfL**

Coleman is keen to convey the speed with which developers take new **TfL** data releases and transform them into apps, a turnaround which also impresses Vernon Everitt:

When we put the cycle hire docking station data live there were literally two products live in the Apple store 48 hours later.[48]

Although this thriving app garden, grown from the seeds of **TfL**'s data, is the main focus of most research and monitoring in this area, it's important to note other users of **TfL**'s data offer, including businesses planning locations of new stores and offices, as well as academics looking into issues such as road safety. Coleman highlights the work of the Centre for Advanced Spatial Analysis at University College London in creating and curating visualisations of **TfL** data, (see Figure 2 for one example[N]).

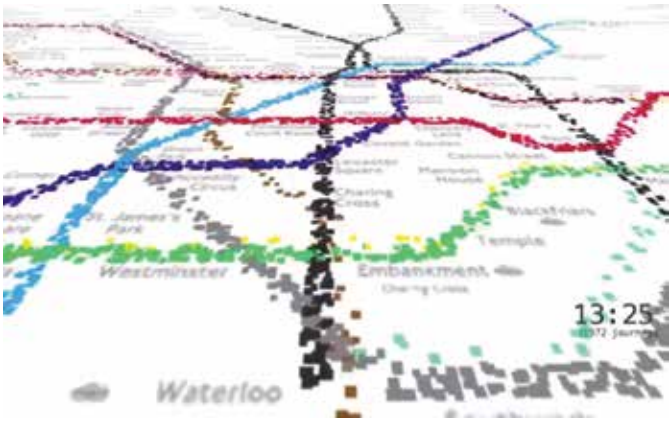


Figure 2: Visualisation of **TfL Tube** *data. Copyright Will Gallia. Reproduced with permission.*

Impact

What is the impact of **TfL**'s open data policy to date? There are multiple lenses through which to view this question. Did **TfL** save money by adopting a policy that effectively outsourced the majority of its app development? If time is money, how much "money" did **TfL** save its customers by better informing them of delays and disruptions to transport services? Has general customer satisfaction improved following the policy and what is the value of this to **TfL**, and to London? Has **TfL**'s data offer stimulated an app economy that is making a real contribution

N For more examples, see (CASA blog network n.d.)

to London and the nation's GDP? And has **TfL's** leadership in this area influenced other transport players, nationally and internationally, and with what consequence?

It is possible to conclude that **TfL** have so far saved between £15m-£42m through opening raw data to the app market, rather than developing all its apps in-house. In May 2015, **TfL** released its own in-house app for users of London's **Santander Cycle Hire Scheme**, enabling them to receive a bike release code straight to their phone, without having to use the docking station terminal[49] (for more on why **TfL** developed this app in house, see Discussion, below). A Freedom of Information request submitted to **TfL**[50] reveals development costs of £118,898.06 associated with the new app.

Number of developers registered on **TfL** website:**>5,000**

Number of apps available powered by **TfL** data: **362**

Reach of apps powered by **TfL** data: **~4 million people**

Value of time saved by users of these apps in 2012: **£15m-£58m**

Crudely, we might therefore suggest that had **TfL** made the decision to hold on to its data and develop all of its apps in-house, it would need to have outlaid development costs of over £43m to deliver all 362 apps currently powered by **TfL** open data. Alternatively, looking at reach (i.e., app downloads, of which **TfL** report there had been 29,139 for the cycle app at the time of the FOI response), **TfL** would need to have spent £16,321,501.77 to achieve the ~4m figure reported by Deloitte. This latter estimate is in a way even more crude given it takes no account of reach over time and in any case bike hire scheme users must make up only a tiny subset of London transport users (and therefore downloaders of London transport apps) overall.

Preparing **TfL's** data offer, launching the developer area of the website and generally putting **TfL**'s data "in a reliable shape[51]" so that, for example, people could query it on a regular basis, is generally understood internally at **TfL** to have incurred a one-time cost of about £1m[O]. The ongoing costs of supplying open data are "almost too hard to disaggregate[52]" from **TfL's** requirements for accurate real-time data both to manage the transport network and to power its own website.

Did the fact that **TfL**'s open data policy effectively outsourced its app development deliver a cost benefit ratio of 1:43? Or 1:16? "It's conjecture, really, isn't it?" says Phil Young:

How much would I spend on building native apps for all of **TfL***'s transport services? I don't know how much I would have spent on that. I haven't costed it, because I didn't have to do it. But let's imagine I would have spent a number of millions since 2010. It would be in that order, anyway.*[53]

How has **TfL**'s open data policy impacted the public it serves? **TfL** is a public body, so it makes sense that the impact of its open data policy on the public it serves must also be taken into account. The figure most quoted when discussing this aspect of the impact of **TfL**'s open data policy, is the one derived by Deloitte as part of the **Shakespeare Review of Public Sector Information** in May 2013. The foundation of this analysis is the idea that time is money:

By making some assumptions about the number of passenger hours saved through better access to information, and the value of an hour, it is possible to estimate the time potentially saved, and the value of that time, owing to the information released by **TfL**.[54]

Deloitte used official annual figures on Lost Customer Hours due to transport disruptions, and hypothesised how many users of apps based on **TfL** data would have avoided the delays by being better informed[P]. From this analysis, and using official Department for Transport estimates of the value transport users place on their time[Q], they calculated that overall, apps based on **TfL** data saved transport users £15m (conservative estimate) or £58m (optimistic estimate) in 2012.

Deloitte compare these annual savings to those projected for users of the first phase of the **HS2 rail project** linking London and Birmingham, which, if calculated using the same time values as the Deloitte study[R], come in at £105 million. This allows Deloitte to imply that by simply making its data open, **TfL** has delivered monetised time savings that are comparable to those of a major and politically contentious infrastructure investment project.

Using a different approach and set of figures, **TfL**'s Head of Bus Systems & Technology Simon Reed has shown that apps powered by **TfL**'s bus data will deliver £83m of customer benefit over 10 years, at a cost to **TfL** of £820,000[S 55].

Has open data improved **TfL**'s relationship with its customers? According to Phil Young, **TfL** measures its relationship with transport users in terms of trust, using metrics including customer satisfaction, user experience, progress and innovation, value for money, and perceptions of how much **TfL** cares about its transport users:

O Both Phil Young and Vernon Everitt used this figure in interview, although neither were able to point to a source confirming it.

P Starting from the ~4million "reach" figure quoted further up this report, Deloitte estimated:
20/40% (conservative/optimistic) of people who downloaded an app became regular users
Of these, 10/25% (conservative/optimistic) were able to use the app to find an alternative route in the event of a delay.

Q The value of the time saved was calculated according to DfT values of leisure/commuting time, not business time—were it calculated according to the latter figure, the overall value of time saved would be much higher.

R See note q—the original HS2 time savings value (£440m/year in 2012 prices) uses a working time value, not leisure time/commuting time as used in the Deloitte study, so Deloitte have switched the values here. Note that Deloitte are careful to make clear that they have not critiqued the HS2 study as part of their analysis. In this, I have followed their lead.

S Using TfL's own survey data, Reed calculates that 24% of daily passenger journeys by bus are supported by apps. Based on passengers' historic willingness to pay 1.44 pence per journey for SMS services where they can receive information about the time of the next bus and any service delays by typing a code shown on the bus stop into their phone, he calculates a value delivered to bus passengers of £8.3m per year (being the amount of money saved by bus users now they can get the same information for free). This analysis is interesting not least because it puts a figure on the annual running (opex) costs of the Live Bus Arrivals Data feed of £47,000. In addition to the initial development of the data feed (£350,000), this brings the total cost of providing the data over 10 years to £820,000 (against £83m of customer benefit).

These metrics add up to a trust metric, and [they're] all on the way up …. [Open data is] part of the reason or a contributory factor—it's hard to get the exact delta that you're getting out of it.[56]

Although he appreciates it is a crude measure, Vernon Everitt believes it is significant that he no longer receives complaints about **TfL**'s information provision. Referring to **Tube** strikes that took place in Spring 2015, he believes that delivering real-time information about disruptions through data feeds "alleviated at least some of the aggravation" caused to London's commuters :

It's quite hard get your arms round that and put a number on it. I think if we tried hard enough we probably could. But we just know it's working.[57]

He also credits to open data some of **TfL**'s success in managing London's transport network when the capital hosted the 2012 Olympic Games. During this time, **TfL** shared all of its transport planning documents as open data, and although they did not see a significant spike in the creation of new apps, the move "gave employers and government and organisers a sense of confidence that everything we knew, they knew,"[58] and helped make the case for everybody involved to work together to deliver a 20% reduction in regular transport demand during peak Games usage.

TfL are in the process of conducting research into what value its data offer has delivered to the London economy through stimulating app development, and expect to have initial findings to report at the end of 2015. Phil Young notes that a lot of small app development companies who started out using **TfL**'s data have gradually grown into larger tech companies, citing CityMapper and MXData as two examples.

Finally, Vernon Everitt credits recent decisions to embrace open data by other transport organisations, notably National Rail Enquiries, the service run by the Association of Train Operating Companies (ATOC) to provide transport information relating to the UK's privatised rail network, as following **TfL**'s lead. It's important to remember that **TfL** were only able to demonstrate the benefits of a more open data strategy by taking an initial leap of faith—one the private companies that make up ATOC might have felt less inclined to take by themselves. Now that more open transport data like this is coming online, Everitt anticipates an acceleration in the development of integrated transport apps like CityMapper.

Discussion

All interviewees saw no end to **TfL**'s commitment to open data. Vernon Everitt wishes to expand the program, seeing it as having a key role to play in meeting the challenges of London's growing population. Emer Coleman described the policy as "embedded" within **TfL** as an organisation, and said that TransportAPI—the open transport **data wholesaler** of which she is now a co-director—did not plan for **TfL** changing their open data policy at a business-risk level.

Paul Clark counsels against assuming that just because one set of data—transport—has immediate and obvious utility, then so will others. The "fascination" with transport data he says he witnessed while helping to organise the **GLA** developer events ahead of the launch of the **London Datastore** was "an order or two of magnitude beyond anything else …. Clearly, if you're going to try and make money out of selling apps then it was transport or nothing"[59].

Interestingly, Phil Young points to customer focus group research that indicates that a majority of transport users still want an official **TfL** travel app, although "whether we're ready to give it to them, I don't know"[60].

Transport for London: Get set, go!

Vernon Everitt was conscious that the release in May 2015 of **TfL**'s first app developed in-house for some time—the Santander cycle hire app that includes a payment component, mentioned above—had signalled to some that **TfL** was moving away from open data as a policy:

There is absolutely no question of our commitment to open data in all its forms. What I can't do is hand over [payment details] to the apps market …. I just think you have to be really careful. You're talking about people's bank details here. It's not for me or for **TfL** *to give away either the individual journey histories or the payment details of our customers. What sane organisation would do that?*[61]

This awareness of the difference between data that is appropriate to release openly and data that contains personal details is undoubtedly a good thing, and apart from the administrative error that saw Customer ID numbers briefly released with cycle hire statistics, **TfL** appears to get this right. But the interplay of **personal data** and transport data in the new Santander app does point to an issue also highlighted by Paul Clarke, about the future market for transforming open data, and how it might consolidate as apps exploiting the **personal data** of their users develop and grow. Although what Clarke terms the "data suck" of apps using open transport data (that is, the amount of data apps extract from their users) is currently quite low, this could change. Just as the rise of Facebook has consolidated audiences and affected the way news publishers function, so once services like **Google Now** become more adept at anticipating their users' needs, users may find themselves locked into such services just as many are locked into Facebook today, and data publishers may find the market for their data starkly reduced.

Len Green / Shutterstock.com

Vernon Everitt also perceives this risk:

I think it's important that we preserve the ability of apps developers to take this stuff and to make products quickly. If this became just a big corporate enterprise, I think that would work against the principles of openness.[62]

Opening public data should not equal privatising data, and so far, it has not. Yet we should be conscious of how little we know about the markets created by opening public data. If they turn out to function in any way like the markets created by the rise of the world wide web as a global communications platform, we should be wary of rapid market consolidation.

Calls to action

FOR POLICYMAKERS

- **Transport data is highly amenable to the open data approach, particularly where smartphone adoption among transport users is high. TfL achieved significant cost savings through using an open data approach to outsource app development, and its open data approach has also improved trust.**
- **A traditional business case was hard for TfL to model at the outset. This study should encourage policymakers to support transport authorities in making the leap of faith necessary to move to an open data approach.**

FOR THE OPEN DATA COMMUNITY

- **This case should form part of the open data community's advocacy toolkit. TfL's move to open its data has been shown to deliver £15m-£58m in annual monetised time savings to London's transport users, all for relatively low investment. This is comparable to savings used to justify building the first phase of the HS2 rail project linking London and Birmingham—a major transport infrastructure project.**

FOR FUNDERS

- **More research is needed into how the market for transport data re-use might consolidate as smartphone users become locked in to personalised services like Google Now. More understanding is needed of the markets created by opening public data. If they turn out to function in any way like the markets created by the rise of the world wide web as a global communications platform, we should be wary of rapid market consolidation.**

Summary

HM Land Registry began a phased release of its data on property transactions—the **Price Paid Dataset**—in March 2012, and by November 2013 the entire historic record dating back to 1995 was released. The data provides much-needed transparency in a historically "murky" business, and is already being used extensively by some traditional players in the property market. Additionally, new players are consolidating around the field of **proptech**, developing digital tools to bring buying and selling property "out of the Stone Age". **Proptech** startups attracted an estimated $1.4 billion in investment globally in 2014. PI Labs, an incubator for **proptech** startups, opened in London in late 2014.

Key takeaways

- **HM Land Registry is a trading fund, a type of government agency that meets its outgoing costs from the money it charges for services. Along with the UK's other trading funds, its future has been a sticking point in the UK's journey towards open data.**
- **In contrast to some of the other datasets studied in this report (e.g., TfL), HM Land Registry's data is delivering value most where it is cross-referenced with other, often proprietary, datasets.**
- **Releasing Price Paid data as open data has had a positive effect on data quality. HM Land Registry report an increase in the number of notifications about inaccurate data by members of the public since the Price Paid data was released.**
- **HM Land Registry is a potential candidate for privatisation. The government set a poor precedent when it privatised the Postcode Address File alongside Royal Mail (see page 7). Selling off HM Land Registry data would likely lead to a slowdown in the positive impacts described in this case: market transparency and proptech innovation.**

Background

HM Land Registry (HMLR) was created in 1862 to register the ownership of land and property in England and Wales. It is responsible for maintaining the **Land Register**, where more than 24 million **title**s (evidence of land and property ownership) are documented.

In 1993 it was established as a **trading fund**—a type of government agency that has the authority (under the Government Trading funds Act 1973) to meet its outgoing costs from the money it charges for services. Other **trading fund**s include **Companies House**, **The Met Office** and **Ordnance Survey (OS)**. **HMLR**'s receipts come mainly from fees it charges individuals and organisations to register land, and to register changes in **title** to land and properties.

The **Land Register** is the world's largest property database[63]. At the height of the UK property boom in 2007, it processed around £1 million worth of property every minute in England and Wales[64]. However, the credit crunch saw a slowdown in the UK market. **HMLR** lost £220m over three years to January 2011 and was forced to increase fees to users by 30%[65]. Those fees have since decreased, as part of **HM Land Registry**'s strategy to encourage more people to register transactions online.

HM Land Registry: The UK's trading funds, and two futures for open data

The **trading fund** model has played a controversial role in the UK's journey to open data, since at least in the cases where agencies are funded by receipts from reselling data (for example **OS**) the two approaches are in conflict. Indeed, a previous study by the author of the UK's journey to open data, records "strong resistance" from **trading fund**s and in particular **OS**, to a future of open government data:

The Power of Information Review had recommended a detailed cost-benefit analysis of the **Trading fund** *model, which was commissioned jointly by* **HM Treasury** *and the then Department for Business, Enterprise and Regulatory Reform, and published in February 2008. Despite that study finding strongly in favour of abandoning the* **Trading fund** *model, and despite the Power of Information Taskforce recommending reform on the back of its findings, positions against re-examining the* **Trading fund** *model within government remained entrenched.*[66]

In November 2010, the government announced plans to create a **Public Data Corporation (PDC)** to coordinate the release of data from the **trading fund**s that deal in data and information (**OS**, **The Met Office**, **Companies House**, and **HMLR**). The plans represented "a wholesale reorganisation of the largest **trading fund**s", but did not go into detail on how much of the data released by the **PDC** would be open, and also included the possibility of the **PDC** being part-privatised[67].

In August 2011, the Department for Business, Innovation and Skills (BIS) consulted with the public on the **PDC**'s data policy[68]. In November 2011, **Chancellor** George Osborne made commitments on open data in his **Autumn Statement** that included data releases from **Companies House**, **HMLR**, and **The Met Office**, a limited data release from **OS**, and the establishment of a Public Data Group to share best practice that would supersede the **PDC**[69]. In March 2012 BIS published a response that included terms of reference for a Data Strategy Board that would "seek to maximise the value of data from the **Public Data Group** … for long-term economic and social benefit, including through the release of data free of charge"[70].

In July 2014 the UK's **coalition government** abandoned plans to privatise or part-privatise **HMLR**, after **Liberal Democrat** business secretary Vince Cable blocked the move[71]. However, after the May 2015 elections returned a majority **Conservative** government privatisation may be back on the table: **Chancellor** George Osborne has outlined plans to privatise £23bn worth of government assets as an attempt to bolster public finances[72], and campaigners against privatisation identify **HMLR** as one of the targets[73].

In April 2015, the functions of the **Public Data Group** were integrated into a new BIS-run board, the **Digital Culture, Services Platforms, and Data Board**, and the **PDG** met for the last time[74].

The data

HMLR's **Price Paid Dataset** contains over 20 million transactions going back to 1995 for residential properties sold for full market value. It was this dataset that **Chancellor** George Osborne committed the Land Registry to releasing as open data in his 2011 **Autumn Statement**. In March 2012, **HMLR** began releasing monthly files of transactions. 2013 saw the release of historic data in two phases: Records of transactions between 2009 and 2012 were released in June 2013; and in November 2013 the full historical record from January 1995 to the current month was released. That year also saw the release of the price paid database as **linked data**. In 2014, **HMLR** worked with a third-party supplier to create an "easy to use" front end, the Price Paid Data Report Builder tool[75].

The data is not a full picture of land ownership in England and Wales—it only includes details of properties bought and sold between January 1995 and the present (see box on page 19). It is released under an **Open Government Licence** and as such it conforms with the **open definition**. It is updated every month with the new transactions that have been recorded, and corrections to older entries.

The quality—and timeliness—of the data relies in part on those people (homeowners and their solicitors) who register it with **HMLR**. A sample monthly update file showed approximately 2,500 entries against transactions older than a year, which might be assumed to be corrections to existing transactions. Of the (roughly) remaining 79,000 transactions assumed to be new transactions, 3,345 (4.2%) were more than six months old.

Before its release as open data, Price Paid data had been made available in bulk to paying customers as part of **HMLR**'s commercial activities. As such, **HMLR** had sought the view of the **Information Commissioner's Office (ICO)** in 2004 as to whether Price Paid data (which includes addresses) constituted biographical information, and the **ICO** had indicated that it

did not. Before its release in 2012, **HMLR** carried out a Privacy Impact Assessment[76], and reviewed the assessment again in 2013[77]. Both reviews concluded that Price Paid data was deemed not to be biographical or personal in nature. The pre-release review flagged an increase in companies using the data for direct marketing purposes as a potential issue; the 2013 review concluded that there had been no increase in direct marketing as a result of the release.

The path to open

Who was calling for **HMLR** Price Paid data to be released as open data? The fact that **HMLR** is a **trading fund** means it was a general target for open data enthusiasts. In responses to the government's **PDC** consultation, **HMLR** data was rated as the third most requested dataset to be "free for use and reuse" after **Ordnance Survey** data and Address/Postcode information[82V]. Lynne Nicholson, Head of Data Products and Services at **HMLR**, says the data release responded to "a general call for government to release more data[83]", and did not mention any particular organisations who were calling for the data to be released, beyond the Open Data Institute, which in any case was only created at the same time as the announcement that the **Price Paid Dataset** was to be made available as open data.

Data based on **HM Land Registry** data, including Property Sales[84] and Median House Prices[85] had been published intermittently on **data.gov.uk** by the Department for Communities and Local Government since 2010, and had fed into a small number of apps, notably the parking app JustPark. Before the open data release, **HMLR** Price Paid data had been available as downloadable .txt files (and also as bespoke reports) to a small group of organisations on a commercial basis. A business case for releasing the data, drafted before the **Chancellor**'s **autumn Statement** and released under FOI in February 2012, reveals that **HMLR** were attracting a modest £600,000 of income from this activity per year. It notes that:

Existing customers for this data are, in the main, companies that publish the data on their own websites. Land Registry has 30 existing subscribers who are website based businesses and 7 others which they class as consultant/research users. In a typical month they also receive ad hoc requests for specific areas and time periods of data from a small group of customers, most of which are Local Authorities—on average 5 requests a month. They also receive ad hoc requests from **Estate agent***s and solicitors, usually 1-2 a month from each sector*[86].

During interview, **HMLR** did not wish to reveal the identities of its paying customers prior to the data release[87]. A source in the industry speculated such customers might include property portals (such as Rightmove, Prime Location, and MousePrice), business-to-business publications (such as Hometrack) and

What is in HM Land Registry's Price Paid dataset, and what isn't

HM Land Registry's **Price Paid Dataset**, the subject of this study, is a subset of the information **HM Land Registry** holds on land and property ownership in England & Wales. The over 20 million transactions in the current version of the dataset appear to relate to 11,738,465 unique addresses[T]. In total, the **Land Register** consists of 24 million **title**s.

Even the **Land Register** itself is not a complete picture of land and property ownership in England & Wales, since compulsory registration of land and property is only triggered by a property transaction of some kind. And although the Land Registry was created in 1862, compulsory registration across England & Wales only arrived in the 1990s. Since then, notifying the Land Registry has been compulsory when land or property is bought or sold, and (more recently) when any mortgage is taken out on it, or when land or property is inherited. Some land and properties have remained in the same family or organisation's ownership for generations, and have therefore never been registered. In July 2012, **HM Land Registry** announced that 80% of the land mass of England & Wales was now registered[78].

The **Price Paid Dataset** is a subset of the information on the 80% of the land mass that **HM Land Registry** does hold (just under 50% of the **title**s it holds that in turn relate to 80% of the land mass of England & Wales). The **Price Paid Dataset** also does not contain information on the individuals who own the **title**, or the **title** ID. The address, postcode, and transaction value is included, as well as some basic information about the property, such as whether it is a new build. Further information may be obtained for a fee of £3 per **title**.

While it's easy to understand why the Land Registry don't include the registered owners of all the properties in the **Price Paid Dataset**, some of the dataset's other exclusions are harder to understand. Price Paid only includes property transactions between individuals where property has been sold for full market value. It does not include any information (for example) that the Land Registry holds on properties that have been inherited, or on properties owned by corporations, or on land transactions. Land Registry does make available in bulk its data on land and property owned by corporations (excluding private individuals, overseas companies, charities, and trustees), but only to paying customers. This database consists of 3.2 million **title** records.

In July 2015 David Cameron announced that **HM Land Registry** would release, as open data, data on property and land owned by corporate entities registered overseas, including the name and correspondence address of the legal owner of the company. This followed a series of reports from **Private Eye** magazine using similar data, obtained from the Land Registry under Freedom of Information laws. The data released to **Private Eye** consisted of details of 96,441 **title**s[U].

continued on page 21...

T Price Paid Dataset does not include title IDs, so unique addresses were used as a proxy. The author downloaded the dataset on 21 September 2015, and queried it for the number of entries with a unique Postcode/PAON/SAON combination.

U An FOI request for the same dataset is available at (Bowden 2015)

V As well as these categories of data, which all relate to location, meteorological data, transport data, health, education, and crime statistics were all cited by respondents as having particular societal benefits.

nationwide property consultancies (such as Savilles). Attempts by the author to interview one such potential customer (Rightmove) were unsuccessful.

The same business case document asserts that, across all the **trading funds** considered, it would be existing customers of the data that would capture a significant proportion of the initial benefit of releasing the data for free. This notwithstanding, when asked to identify stakeholders who were against the release of the **Price Paid Dataset** as open data, Lynne Nicholson, Head of Data Products and Services at **HMLR**, pointed to concerns from existing commercial customers "that opening up the data would allow more competition into the market"[88].

Indeed, the response to a consultation on government open data strategy put forward by the **Cabinet Office** in 2011, includes one from Landmark Information Group (who own Mouseprice.com). It argues strongly in favour of a presumption for open data, but notes that:

Due consideration should be made of the private sector organisations who may already offer services around any data sets which currently attract a charge from the public sector data providers which may be candidates for being re-designated as 'open'. Thus avoiding any potential negative affects to the existing markets.[89]

Users and outcomes

Lynne Nicholson is keen to underscore that, now the **Price Paid Dataset** is released openly, **HMLR** cannot track who is using it:

It's open data. We put it out there. We don't put a registration system up there. So we have absolutely no idea who's downloading it. We can put analytics around it to say how many downloads, but we simply don't know who's downloading it. That's the whole point of open data. For the Price Paid data we had roughly 30 customers [before the open data move]. Obviously they're going to be downloading it. But our downloads are, you know, thousands a month. We have no idea who the others are or what they're doing with it.[90]

According to the second privacy impact assessment, between January 2012 and March 2013, the Price Paid data was downloaded a total of nearly 78,000 times[91].

One data re-user I spoke to described how portals listing Price Paid data have proliferated on the web:

Now you can download a **.csv** *and have your own [***HM Land Registry***] property price portal up in a couple of hours …. If you Google your own house address you will see there are dozens of sites showing sold prices for your house.*

Nicholson also notes that government departments were attracted by the open data offer, then entered into confidential data-sharing agreements involving richer **HMLR** datasets as a result. Nicholson mentions "significant use by the media and researchers," making an example of the data and what could be achieved with it. The cartogram in Figure 3 (which appears

Median House Prices, 2014

Figure 3: Cartogram of median house prices in London. Released uncredited on the London Datastore under OGL.

uncredited on the **London Datastore** and was picked up by bloggers at UCL's Centre for Advanced Spatial Analysis[92]), shows the median house prices in different London boroughs.

HMLR hosted an event in July 2014 for users of its data[93]. At the event, Manuel Timita from Illustreets—a platform that takes open data from a variety of sources (including the **Office for National Statistics**, **OS**, **Police.uk**, Environmental Agency,

Figure 4: Screenshot of Illustreets' video timeline of rise in London house prices. Copyright Manuel Timita. Reproduced with permission.

Department for Transport, as well as **HMLR**) and displays it on a map to help people decide where they want to live—showed a timeline of London house prices, matched against key events surrounding the credit crunch (see Figure 4)[94].

Also at the event was Johnny Morris from **estate agents** Hamptons International, who, according to the Land Registry's report of the event, "divulged how [Hamptons International] are enriching their market view of the 'downsizer', through mixing their own proprietary data with Land Registry Price Paid Data and **Ordnance Survey** open data"[95]. However, it is not clear whether Hamptons were an original customer of the

HMLR Price Paid data, or only began using the data once it was released openly.

One **estate agent** who is using the data now it's been released by **HMLR** is Adrian Black, founder and director of YOUhome. He moved to estate agency after 10 years working in technology for Goldman Sachs, drawn to a business that looked ripe for technological change:

In every agency business you have diminishing returns and you have to become more productive and efficient and also pick up revenue from other sources. That's not happening in estate agency. Fee rates are coming down, but they're coming down slowly and the increase in property prices means agents are actually earning more. [With online portals] reaching buyers is getting easier, and yet revenues per sale are increasing.[96]

YOUhome is using Land Registry data and other proprietary sources to make their estate agency business more efficient, and more attractive to potential clients.

Sellers typically look to engage agents who offer the highest valuation and the lowest fees. But that leads to low conversion rates.[97]

Black says seeing actual data can persuade sellers to go for a more realistic asking price, can get buyers comfortable more quickly, and can achieve a quicker sale at a fair price. And it's working for their business. They report that typical central London agents charge 2-3% rates, whereas their central London rate is 0.8%. And in Bournemouth their brokers transact on average 50 properties per year per agent, against an industry average of around 25 (15 in London). "We're the Land Registry's number one fan," says Black:

This is what the government should be doing to make efficient businesses. They should open up data.[98]

Property expert Henry Pryor anticipates that more and more groups will find ways to make use of **HMLR**'s data:

Retailers are already making use of the value of property as part of the matrix that they look at when trying to segment their market. Clearly, they're trying to target different individuals with different products. To know that somebody lives in a million pound house as opposed to a hundred thousand pound house, you can make assumptions based on that information.[99]

The release of **HMLR**'s Price Paid Data has coincided with the rise of the new field of **proptech**: businesses using data and technology to innovate in the property sector. One report quotes figures from CrunchBase showing investors worldwide put a record $1.4bn into **proptech** startups in 2014, although most of the money went to US-based firms[100]. In late 2014 Pi Labs, the UK's first **proptech** incubator, was set up by **estate agent**s Cushman & Wakefield and venture capitalists Spire Ventures.

"It's kind of buzzy at the moment and so a lot of different things fall under [**proptech**]," explains one **HMLR** data re-user who does not wish to be identified. A video filmed to promote a roundtable on the future of the property market in May 2015[101] showcases companies including Splittable (which helps tenants of shared houses split their bills), Hubble (a web platform for making renting office space easier), Homeshift (a platform through which agents, tenants, and housebuyers can communicate during a house move), Fixflo (a tool for tenants and property managers), Land Technologies (a **data aggregator** that identifies land that's suitable for building on), MoveBubble (an online agent that acts on behalf of tenants), and We Are Pop Up (an agency for short-term retail lets).

Peter Thum-Bonanno sold his first **proptech** startup, the property portal Find Properly, to search engine Nestoria in November 2014. Although he is unwilling to ascribe the rise of **proptech** entirely to the release of the **HM Land Registry** dataset, he says there is a link:

I don't know if it's grown or if I've just become more aware of it, but there's definitely a rise in property technology companies who call themselves **proptech** *companies. I think a lot of it is because property seems to still be stuck in the Stone Age. The way things are*

> "We're the Land Registry's number one fan. This is what the government should be doing to make efficient businesses. They should open up data."
>
> Adrian Black, Founder, YOUhome

What is in HM Land Registry's Price Paid dataset, and what isn't, continued

Using this data, **Private Eye** were able to uncover a number of public interest stories, including the frequency with which property and investment companies use overseas corporate vehicles to own significant London landmarks, thereby achieving capital gains tax and stamp duty advantages.

An interactive map that displays the data is available on the **Private Eye** website[79]. In order to build the map, **Private Eye** used a dataset of **title plan**s released by **HM Land Registry** under the EU's INSPIRE scheme—the **INSPIRE index polygons**. Although the **INSPIRE index polygons** are released by **HMLR** under the **OGL**, the dataset contains third-party intellectual property that belongs to **Ordnance Survey (OS)**. **HMLR** states in its conditions of use[80] that this means anyone making the polygons available to third parties, or using them for anything other than "personal, non-commercial or commercial or non-commercial use within your organisation" needs to "contact **Ordnance Survey** for the relevant licence conditions". Christian Eriksson, the freelance journalist who led the project on Land Registry data, says that neither he nor **Private Eye** contacted **Ordnance Survey** before publishing the interactive map, believing his use of the **INSPIRE index polygons** constitutes fair dealing under Section 30(2) of the Copyright Designs and Patent Act 1988[81].

done still seems to be very antiquated. People think technology could help make that more efficient. And I think, compared to a lot of other industries, there's a huge amount of data out there and the Land Registry is mainly to thank for that.[102]

Thum-Bonanno's new venture is a tool to assess the efficiency of agents by using data including asking prices, sold prices, and **conveyancing** times. GetAgent uses a mix of public and private data sources, but **HMLR** data "plays a big part in that".[103] As well as using **HMLR** Price Paid data, it uses a non-open data **API** supplied by property portal Zoopla[104]. This is not uncommon, and the developers and entrepreneurs I spoke to all agreed with Vasanth Subramanian of Splittable, that **HMLR**'s **Price Paid Dataset** was just "one tool in the toolbox":

It's when it's combined with other data, expertise or products you or your organisation might possess that the real value of the dataset is realised.[105]

A developer who did not wish to be identified feared that because Price Paid data was creating value only in conjunction with other datasets, arguments for it to remain open during any privatisation process would be weakened. This would be a mistake, according to Vasanth Subramanian:

It would be a big problem if the Price Paid got sold off with **HMLR**. *Even if the dataset [is not providing amazing value in and of itself], it's a vital building block and you'd see a slow down in* **proptech** *innovation.*[106]

Impact

How can we measure the impact of **HMLR** releasing their Price Paid data? Henry Pryor believes one of the main benefits will be in bringing transparency to the housing market, and likens the change to the introduction of screen-based trading on the London Stock Exchange in the 1980s:

This data shines a light into a murky world and provides transparency which in turn provides trust and confidence in transaction values associated with the purchase and sale of residential property. We don't have it in the rental market and the market suffers as a result. But just as we saw with **Big Bang** *in the city and this idea of effectively real-time pricing, buyers and sellers can get far more confidence and as a result you get a far firmer, more mature market where confidence is high.*[107]

This echoes Adrian Black's experience on the ground: that homeowners use data to feel comfortable about the price of the property they are buying or selling, helping the market move more quickly.

But the state of the UK housing market is subject to many more forces than the state of the estate agency business, however "murky" or "Stone Age" those involved in it believe it to be. Prior to the UK's 2007 credit crunch, precipitated by the run on mortgage lender Northern Rock, housing transaction volumes had peaked at around 150,000 a year. 2015 figures (see Figure 5, next page) show they have now stabilised at around 100,000 a year[108]. Although the steady

Total UK Residential Property Transactions

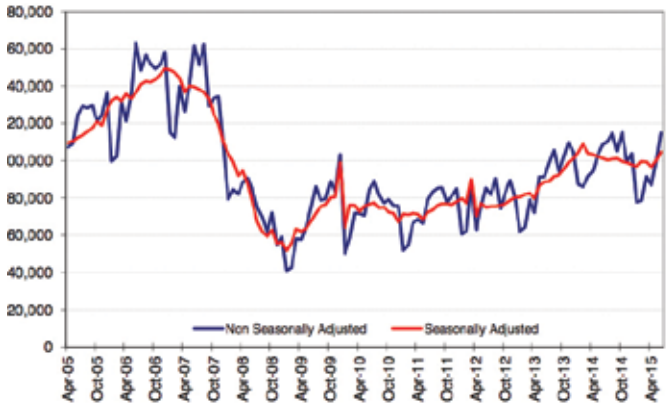


Figure 5: Total UK Residential Property Transactions. Reproduced from (HMRC 2015).

upward climb towards this figure from a low of around 50,000 a year in December 2008 appears to accelerate after March 2012, given the complexity of the market and the availability of the data to market participants through **HMLR**'s paying customers prior to this date, it seems wise to regard **HMLR**'s open data release as a correlation, and not a cause.

The OECD ranks the UK in the top 10 countries with over-valued housing markets, one of six in this group where house prices are still rising[109]. The affordability of homes in the UK is a significant political issue. Transaction volumes are also distorted by the attempts of successive governments to both maintain an economy that is back-stopped for many voters by the (inflated) value of their property, and also help first-time buyers afford a home[110].

Compared to ascertaining the release of open data's effect on the housing market, its effect on innovation in the property sector looks slightly more straightforward. Here, the rise of the **proptech** sector, and the opening of Pi Labs, look like good indicators for innovation, and many of the developers and entrepreneurs involved are quick to recognise the contribution of **HMLR**. But all of them qualify this with statements that indicate that **HMLR** data works best in conjunction with other datasets, be they other open datasets (as with Illustreets) or, more often, closed data sets (as with Peter Thum-Bonanno's GetAgent). This stands in contrast to **TfL**, where the end products created by innovators are far more likely to rely on **TfL** data alone. Any contribution that **proptech** makes or goes on to make to the UK economy will be hard to pick apart: It will be hard to say how much of this is thanks to **HMLR**'s open data policy.

One area where it is easiest to say that the open data policy has had impact is data quality. Lynne Nicholson states that increased exposure has also helped **HMLR** to maintain data quality and that they are reacting to an increased number of queries since release:

Now we're publishing information, it's more accessible, more people are viewing it, so you're bound to get more queries and questions, and requests too. I'll give you one example. Say we've captured a property as an end terrace. The owner might ring up and say 'I regard my property as semi-detached, not an end terrace'. So those are some of the types of questions we get.[111]

HMLR report an increase in the number of notifications about inaccurate data by members of the public since the Price Paid data was released, saying they received 2,256 issues related to incorrect or missing Price Paid information from the release of the data up to the publication of their second privacy impact assessment in May 2013. However, no figure for complaints about inaccurate data prior to the Price Paid release is given for comparison.[112]

Discussion

Lynne Nicholson explains how open data has shifted the focus of **HMLR**:

Our focus is changing from commercial activity to the release of data so that organisations can develop products and services from our data. Our remit is not to compete with the private sector and not to compete with organisations who are downloading our data.[113]

HMLR is committed to making the data it releases easier to digest, and has invested in a reporter tool for Price Paid data that it believes is driving more non-expert users to its data.

As for the **proptech** sector, for the moment innovation rather than consolidation looks set to rule. "You've got lots of startups like ours that are nibbling away at small bits of the value chain," says Peter Thum-Bonanno:

So you have companies that are trying to make getting a mortgage easier and more straightforward. We're trying to help you find an **estate agent** *.... There's no one yet who is doing the whole process for you from finding an* **estate agent**, *finding you a buyer, helping you move, helping you manage your home. It seems like that's the aspiration of all these small* **proptech** *startups and it will be interesting to see if there will be a consolidation in the industry, allowing you to manage your entire home online, or if it's going to continue to be piecemeal in the future. I'm not sure.*[114]

HMLR's data is delivering value to existing stakeholders in property price information (**estate agent**s and homeowners) and stimulating innovation that looks set to contribute to UK economic growth. Putting an exact figure on the impact of the data, however, may never be possible.

In the absence of quantitative methods, it is likely that future data release policies will be governed by other factors. Attitudes to **HMLR**'s current data release policy, and those of other **trading fund**s such as **Ordnance Survey**, varied wildly between the data users interviewed. Innovators ranged from being "very grateful" to **HMLR** for the data release, to feeling

frustrated that more data was not being either collected (e.g., number of bedrooms, square footage) or released (e.g., full land ownership records). Meanwhile, at the more traditional end of the spectrum, Henry Pryor did not see why businesses who were able to monetise **HMLR** data were not required to pay for it.

This contrast in attitudes persists to the highest levels of government. Claudia Arney, chair of the **Public Data Group** between June 2012 and its final meeting in April 2015, describes **trading funds** navigating their open data strategies in an environment of competing pressures: strong pressure to open up data; and equal and opposite pressure to provide evidence that any change to the status quo represented true value for money and maintained the integrity of the data in question[115]. She believes the integration of the **Public Data Group** into the **Digital Culture, Services Platforms, and Data Board**, which will include representatives from the **Cabinet Office** and **HM Treasury**, will allow **trading funds** to plot their future path with greater certainty. But the election of a **Conservative** government and the broader focus on the health of public finances in UK politics may yet shift the balance.

Number of transactions captured by **HMLR**'s Price Paid Dataset: **20 million**

Number of downloads of **HMLR Price Paid Dataset** between January 2012 and March 2013: **78,000**

Global investment in **proptech** startups in 2014: **$1.4 billion**

Any sell-off of **HMLR** may not be as swift as anti-privatisation campaigners fear. The replacement as Chief Land Registrar of former **Student Loans Company** Chief Executive Ed Lester with the former Chief Executive of Thurrock and Brentwood councils, Graham Farrant, could be read as a subtle shift away from immediate privatisation towards a strategy of delivering efficiency savings before any sell-off. Confirming Farrant's appointment in a written statement to the **House of Commons**, Matt Hancock (then a minister at BIS and now Minister for the **Cabinet Office**) wrote "Graham has the skills and knowledge to manage the organisation through its transformation into a modern, digital organisation.[116]"

What will be of consequence is less if or when the sell-off comes, than whether **HMLR** data is included in the package. Regrettably, the last government set a precedent here when it included the **Postcode Address File** in the sell-off of **Royal Mail**, a move that attracted strong criticism from the open data community[117] and from **Parliament**'s Public Administration Select Committee[118] (see page 7).

Calls to action

FOR POLICYMAKERS

- **Policymakers need to recognise the contribution HM Land Registry's Price Paid data is making both to the traditional property market and to innovation in the sector. Regardless of the future of HM Land Registry as a public body, this data should remain in public hands.**
- **The impact of HM Land Registry's Price Paid Dataset suggests that more of HM Land Registry's data, such as records of land and properties owned by corporations registered in the UK, should be considered for release as open data.**

FOR OPEN DATA ADVOCATES

- **HM Land Registry may be a target for privatisation. The open data community needs to pressure the UK government to prevent the bundling of HM Land Registry data with any sale. In particular, the open data community should ensure that government is aware of all the stakeholders in HM Land Registry data as open data, including estate agents and proptech innovators such as those interviewed for this study.**

FOR FUNDERS

- **Funders could encourage open data advocates to network more with innovators in the proptech sector, in order to cement alliances against data privatisation at HM Land Registry.**
- **Any campaign against data privatisation will be greatly supported by economic analysis that speaks the language of those making the decisions at HM Treasury. However, it is also important to recognise that open data reforms are in some sense a leap of faith (see TfL study).**

Summary

2013 saw the UK Government concurrently host the **Open Government Partnership** (**OGP**) summit and hold the presidency of the G8. Prime Minister David Cameron used this joint platform to launch a positive commitment to **beneficial ownership transparency** (identifying the real owners of corporate entities—a key tool in eliminating the ability of anonymous shell companies to mask financial corruption) that has the potential to drive global norms on the issue. Behind the scenes, a coalition of civil society groups focussed on corruption, fraud, overseas development, and tax justice were driving the agenda, aided by open corporate data pioneer OpenCorporates.

This case study shows how OpenCorporates was able to leverage its expertise in order to play a critical, if minor, role in the success of the campaign for **beneficial ownership transparency**. It also discusses two issues in open data impact as a driver for social change: the continuing role of traditional advocacy and the limits of so-called **exhaust PSI**.

Key takeaways

- **Moving the needle on complex issues like corruption and governance reform requires much more than opening government data: The success of just this one chapter in the campaign for beneficial ownership transparency relied upon a concurrence of factors including political mood, external events, and focussed, concerted efforts from traditional advocacy organisations.**
- **OpenCorporates' expertise, gained from years of handling government data (open and not open) about companies, had a critical, if minor, role to play in this success. OpenCorporates could not have gained this success on its own.**
- **The role of the open data community should not just be to use the government data it is given, but to demand the government data it needs.**

Background

In 2010 the World Bank published a report showing that of 213 grand corruption investigations across 80 countries, 150 involved corporate vehicles that shielded the true beneficiaries of financial transactions. In these 150 cases, the total proceeds of corruption amounted to approximately $56.4 billion. The report, T*he Puppet Masters: How the corrupt use legal structures to hide stolen assets and what to do about it*, opens with the example of a tender put out by the Kenyan government to replace its passport system:

Despite receiving a bid for €6 million from a French firm, the Kenyan government signed a contract for five times that amount (€31.89 million) with Anglo-Leasing and Finance Ltd, an unknown UK shell company, whose registered address was a post office box in Liverpool. The Kenyan government's decision was taken despite the fact that Anglo-Leasing proposed to subcontract the actual work to the French company. Material leaked to the press by whistle-blowers suggested that corrupt

OpenCorporates: Open data as one small part of the picture

Total proceeds of corruption across 150 grand investigations involving corporate vehicles that shielded the true beneficiaries of financial transaction: **$56.4 billion**

Jurisdictions covered by OpenCorporates' database of company information: **105**

Companies listed in the OpenCorporates database: **85 million**

Number of Goldman Sachs subsidiaries registered in the Cayman Islands: **739**

senior politicians planned to pocket the excess funds from the deal. Attempts to investigate these allegations were frustrated, however, when it proved impossible to find out who really controlled Anglo-Leasing.[119]

One of the report's key recommendations was that information available at company registries should be improved and made more easily available. The recommendation echoed that made by a similar report published by the OECD in 2001, *Behind the Corporate Veil: Using Corporate Entities for Illicit Purposes*, which called on governments "to ensure they are able to obtain information on the beneficial ownership and control of corporate entities"[120].

OpenCorporates is the largest open database of companies in the world. Founded by "open data veterans"[121] Chris Taggart and Rob McKinnon, it launched at the end of 2010 covering 3.8 million UK past and present companies. As Chris Taggart told the **Open Data Institute** in 2012:

We take messy data from government websites, company registers, official filings and data released under the Freedom of Information Act, clean it up and using clever code make it available to people.[122]

The launch of OpenCorporates predates the decision by **Companies House** to release all the data it holds as open data. But **Companies House** has made more basic datasets available for several years, and it was this data, combined with other government data sources (for example government spending data and **Health and Safety notice**s) that fuelled OpenCorporates in the beginning. Taking the same mixed input approach, OpenCorporates has now expanded its coverage to over 105 jurisdictions and 85 million companies today.

The path to impact

In 2012, Global Witness, a civil society group that investigates and exposes the links between natural resources, corruption, and conflict, began dedicating more resources to moving the needle on the beneficial ownership issue. Though their original advocacy target was EU Money-Laundering Regulations, which were about to be updated, Global Witness soon became aware that political forces in the UK were aligning to make the UK's hosting of the G8 summit at **Lough Erne** a favourable venue for pushing reform. If they could get the G8 leaders to commit to **beneficial ownership transparency**, they could begin establishing a global norm.

The UK government was under pressure to respond to several financial scandals, including angry press coverage of the tax avoidance in the UK of major US firms Amazon, Starbucks, and Google, and the 2012 US Senate investigations into British bank HSBC's involvement in money-laundering for Mexican drug cartels. Speaking at the World Economic Forum in Davos in January 2013, David Cameron called on the G8 to get its own house in order on corporate transparency, citing the "long and tragic history of some African countries being stripped of their minerals behind a veil of secrecy", and outlining a three-pronged agenda for his presidency of the G8: Trade, Tax, and Transparency.[123]

Robert Palmer, head of Global Witness' banks and corruption campaign, calls the speech "a real moment in the campaign, where … the direction of travel of the UK agenda [became clear]"[124]. In it, Cameron cites the work of the economist Paul Collier, Professor of Economics and Public Policy at the University of Oxford, and sources at the **Cabinet Office** confirm

that Collier's work had a big influence on the Prime Minister. Also influencing the Prime Minister according to these sources was the fact that G8 leaders "were unlikely to turn up to the meeting with their cheque books open"[125]. A coup like the one accomplished by Tony Blair at the Gleneagles summit in 2005, where G8 leaders committed significant funds in aid to Africa, was unlikely. A "new way of thinking"[126] was therefore needed.

Also looking to repeat the successes of the UK's last G8 presidency were a group of civil society organisations involved in the original Make Poverty History campaign that targeted Gleneagles. The Enough Food for Everyone IF campaign was a coalition of major development NGOs, including Save the Children, Oxfam, Christian Aid, and CAFOD. A subset of this group, chaired by David McNair (then at Save the Children and now at the ONE Campaign), and including Robert Palmer (though Global Witness were never formally part of the IF campaign), began holding weekly calls on the beneficial ownership issue in the lead-up to the G8. Robert Palmer remembers:

It was one of the most impressive group of campaigners I have ever been around …. The G8 campaign was a real collaborative effort between groups like Global Witness who could bring our policy knowledge … and groups like Christian Aid, and the ONE Campaign, who had political access and campaigners and supporters they could mobilise.[127]

At **Lough Erne** in June 2013, the G8 leaders committed to a core set of principles "to prevent the misuse of companies and legal arrangements" that included measures on beneficial ownership[128], and David Cameron committed the UK to a central registry of company beneficial ownership information[W]. He said he would consult on whether to make the registry public. Reporting on the results of the G8 summit to **Parliament**, Cameron stated that there were "strong arguments for public registers of beneficial ownership all over the world"[129]. Ensuring that the registry was open to all became the new focus of the civil society campaign.

The data

Chris Taggart, the CEO of OpenCorporates, was in touch with the NGOs working on **beneficial ownership transparency**, but he wasn't on the weekly calls that led up to the G8. Thanks to a small grant from the Alfred P. Sloan Foundation, OpenCorporates had been delving into the corporate structure issue separately to Global Witness and the NGOs involved in the IF campaign.

OpenCorporates was investigating a generalised approach to how complex corporate structures could be modelled and constructed using diverse official and regulatory data, and represented within the OpenCorporates dataset. The project kicked off with three separate datasets: shareholder data from New Zealand's company registry, to which OpenCorporates already had access through an **API key**;

and two datasets from US regulators[X]. At the heart of the project was the idea that in the age of the internet, "you can have corporate structure set ups appear that can be as complex as the computers that can cope with them"[130], and that society could only combat the proliferation of complex corporate structures through better data.

To make this case, OpenCorporates worked with data visualisation studio Kiln on a series of visualisations to underscore the complexity of the structures behind a handful of the world's more recognisable global companies, mainly in the financial sector. One of the visualisations—representing Goldman Sachs—is shown in Figure 6.

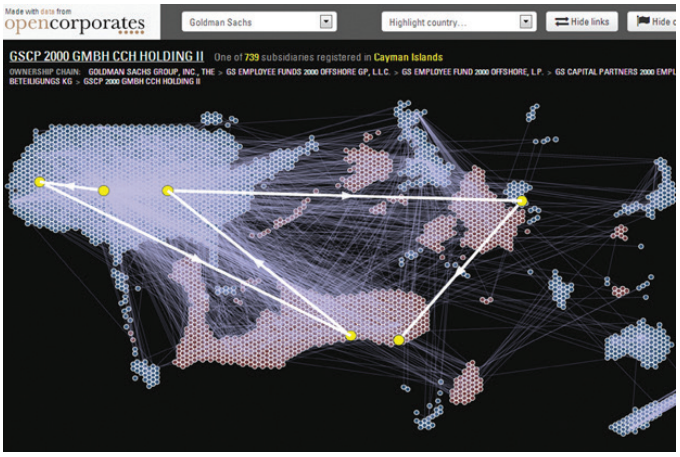


Figure 6: Detail from OpenCorporates and Kiln's visualisation of complex corporate structures. Copyright OpenCoporates. Reproduced with permission.

Displaying the country of incorporation in a map-based way required "some significantly difficult maths," says Taggart. But it lends the message a visual power: The landmass shown below the United States in Figure 6 is the Cayman Islands, an offshore tax haven in which Goldman Sachs has 739 registered subsidiaries[Y].

Outcome

The OpenCorporates briefing was published after the G8 announcement, in July 2013. Robert Palmer remembers it well:

I think it was one of the things that was really powerful about what Chris could do. He had company data, he had the raw data that he could use to demonstrate the ridiculousness of some of these sorts of situations.[131]

The visualisation was reported extensively but almost exclusively by the technology press[Z]. Yet although it may not have enjoyed mainstream impact in and of itself, the work OpenCorporates had put in behind the scenes had a key role to

W The author notes that the G8 Open Data Charter, which commits G8 governments to an "open by default" policy, was also launched at Lough Erne. The campaign for beneficial ownership transparency is treated as separate from the campaign for the open data charter for the purposes of this report.

X For details of the datasets used, see (Taggart 2013). Chris Taggart gives more detail of these datasets and the work that went into obtaining them: "[They were] the [US Securities and Exchange Commission]'s EDGAR website (specifically the Exhibit 21 subsidiary filings) and the Federal Reserve's Bank database of Bank Holding Company corporate structures. Both were problematic: in the case of the Exhibit 21 filings, these are more-or-less unstructured filings, of a wide variety of forms and formats; in the Fed data, it meant reengineering data from complex PDFs of hundreds of pages long." (personal communication)

Y For a discussion of the design element of the OpenCorporates visualization, see (Kiln n.d.)

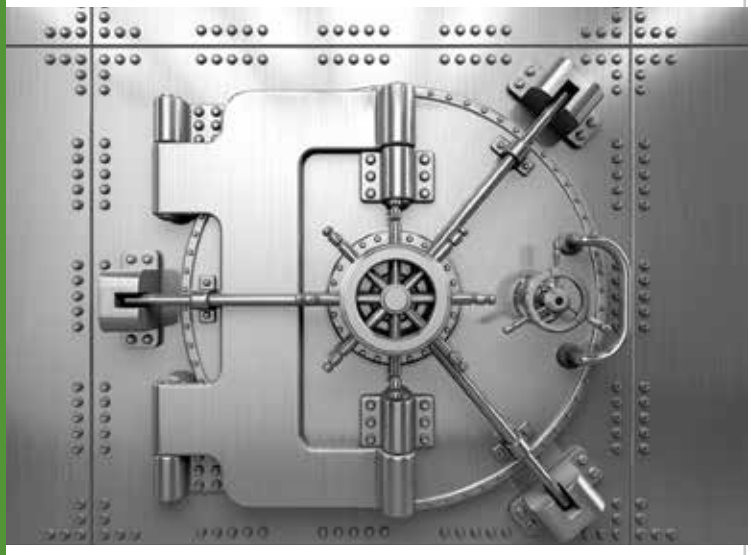
Z Articles were published in GigaOm (Meyer 2013), Wired (Solon 2013), Tech President (McKenzie 2013), The Verge (Kastrenakes 2013), Computer Weekly (Computer Weekly 2013), and Gizmodo (Campbell-Dollaghan 2013).

play in the period between the G8 announcement and the UK's hosting of the **OGP** Summit in October 2013.

> "If it was just OpenCorporates, of course it wouldn't have happened. But the example they were showing and the technical understanding that they had as a result of that, was really critical."
>
> David McNair, Save The Children/ONE

Devising the generalised data structure for corporate relationships as part of the Sloan work was "one of the hardest things I've ever done in my life," says Taggart. But a by-product of this sort of handling of open government data is the ability to speak the language of internal government bureaucracies. This turned out to be a key advocacy tool. The labour put into establishing, for example, how to represent the multiplicity of different structures, how to define what a subsidiary was in the face of different national interpretations, how to represent changes in a structure over time, paid off once OpenCorporates became more deeply involved in the campaign for a public beneficial ownership registry in the UK post-G8[AA]:

Going down to that sort of level is really important …. It's not just the fact that there should be a public beneficial ownership register, nor even that there should be a public beneficial ownership register that is open data. How is the data stored? What's being done on it? How is that being recorded? What level of granularity? What should we do here? What should we do there? Understanding the problem is a nontrivial part of the puzzle here.[132]



David McNair agrees:

The added value that I saw OpenCorporates bring was that very, very detailed knowledge of how this database would work …. If it was just OpenCorporates of course it wouldn't have happened. But the example they were showing in terms of demonstrating that it's possible to build a database, and the technical understanding that they had as a result of that, was really critical for making the standard robust.[133]

Robert Palmer echoes this sentiment, calling what OpenCorporates did "data-based advocacy". He recalls a moment when the Department for Business were consulting on whether directors' and shareholders' full dates of birth should be published on the register: OpenCorporates were able to demonstrate using real data that were dates of birth to be partially redacted, investigators would be unable to identify individual directors and shareholders robustly in cases numbering in the tens of thousands.

Palmer also argues that OpenCorporates was instrumental in pushing NGOs to demand the registry be made publicly available:

One of the biggest impacts that OpenCorporates had on the campaign was to insist that the new beneficial ownership information be provided as open data and this was a key part of the eventual NGO position and the final government announcement.[134]

Impact

In October 2013, David Cameron announced at the **OGP** summit that the UK's registry of beneficial owners would indeed be public. It was a major success for the campaign.

Commitments to **beneficial ownership transparency** are now spreading as a global norm, just as Robert Palmer and his fellow campaigners had hoped. But the campaign is by no means won. Indeed, when the G20 announced its commitment to **beneficial ownership transparency** in November 2014, language around making national registries public was very much missing[135], to the disappointment of transparency NGOs.

In the UK, Chris Taggart will be watching for the impact of the new rules:

What's going to be interesting is the number of companies that dissolve and remove themselves from the register before the beneficial ownership requirements come in.[136]

Discussion

The expertise OpenCorporates has gained across years of handling government data about companies, and especially the work it did on complex corporate structures, played a critical, if minor, role in achieving the overall impact discussed in this study: the Prime Minister's commitment to a public register of beneficial owners in the UK.

AA It's important to note that OpenCorporates was also involved – as were many others in the open data community—in more general campaigning around the UK's OGP Action Plan during this period. (Martin Tisné, Director of Policy at Omidyar Network, private communication)

The success of this chapter in the campaign for **beneficial ownership transparency** relied upon a rare concurrence of factors including political mood, external events, and focussed, concerted efforts from traditional advocacy organisations. The combination of the UK hosting the G8 as well as chairing the **Open Government Partnership** was also a lucky break. Open data expertise had some role to play in this success, but it could not have gained it in isolation. This last point is perhaps the point of this study: Moving the needle on complex issues like corruption and governance reform requires much more than opening government data. Chris Taggart says he has "so much respect for the other actors involved", and the lobbying work they did:

That stuff's really hard. I mean, the stuff I do, the data, the coding stuff, that's way easier than that sort of stuff.

Could the G8/**OGP** summit campaign have succeeded without OpenCorporates? The G20 commitment contained no language on beneficial ownership registers being public, and it's tempting to speculate that this could be because no equivalent organisation to OpenCorporates was involved in the campaign running up to the G20. However, it's likely that individual and entrenched political cultures in the G20 group played a larger role: Insiders say China was the major block in making beneficial ownership registries public.

The story of OpenCorporates' role in the campaign for **beneficial ownership transparency** has been picked up by Jonathan Gray and Tim Davies in their arguments for **participatory data infrastructures**[137]:

We think the case of campaigning for a public registry of beneficial ownership in the UK is an interesting example of where calling for the disclosure of existing datasets was not enough—as the information needed was not routinely collected by the public sector. Civil society organisations had to undertake a more creative, sustained and holistic engagement with shaping and influencing the development of data infrastructures as socio-technical systems.[138]

Their argument, that most government open data is an administrative by-product and may not therefore capture aspects of its subject needed by data re-users such as advocates, is radical and relevant.

We contend that a politics of public information fit for the digital age should go beyond the question of what datasets are disclosed when and to whom, and should additionally ask what information should be collected and generated in the first place.[139]

> "[Political lobbying] is really hard. I mean, the stuff I do, the data, the coding stuff, that's way easier than that sort of stuff."
>
> Chris Taggart, OpenCorporates

Calls to action

FOR OPEN DATA ADVOCATES

- **The next phase of open data advocacy involves influencing the data the government collects. The role of the open data community should not just be to use the government data it is given, but to demand the government data it needs.**
- **Traditional advocacy organisations, with their political experience, strategic thinking and networks of influence, make powerful allies in campaigns to open up government data.**

FOR OPEN DATA PRACTITIONERS

- **A by-product of daily handling of open government data is the ability to speak the language of internal government bureaucracies. That ability can be a powerful tool in advocacy.**

FOR FUNDERS

- **Traditional advocacy efforts can be complemented and enhanced by open data approaches. This might mean a new presentation of the facts based on open data sources, or simply the inclusion in civil society coalitions of open data practitioners able to speak the language of internal government bureaucracies thanks to their daily handling of government data.**

Summary

The **Open Public Services Network** undertook two major pieces of work exploring how government data, beyond traditional accountability mechanisms like **league tables**, can be used to evaluate schools. The resulting outputs—A **GCSE** Schools Guide produced in association with the *Guardian*, and a report into lack of access to "hard" subjects like triple science at **GCSE** in deprived areas—succeeded in reaching both parents and policymakers, and drew significant press attention. This study focuses on the contribution open data can make to improving public services.

Key takeaways

- **OPSN told both a local and a national story with the data it had. The GCSE Schools Guide tool let users see what the data had to say about schools near them. The *Lack of Options* report took a broad, national perspective.**
- **The long-term impact of OPSN's work may best be viewed in the same light as OpenCorporates' contribution to the campaign for beneficial ownership transparency: just one piece of a complex puzzle.**
- **Some of the data OPSN used in this case derived from the National Pupil Database, which is not open data. Much of the data needed to meaningfully evaluate public services will never be released—and indeed should never be released—as open data because it contains personal information about service users. It is in the public interest that an understanding of this data does not reside wholly within government. But vested interests and negative press attention driven by a vocal and effective privacy lobby in the UK serve to make this kind of data-sharing policy "a nasty political space".**
- **Neither official government interpretations of public services data nor interpretations provided by the press fully meet the needs of public service users. But business models for organisations that could serve the public's needs are thin.**
- **Prime Minister David Cameron predicted open data policy would trigger a volunteer "army of effective armchair auditors[140]" who would interrogate government data in the public interest. In fact, that army has been slow to advance.**

Background

The **Open Public Services Network (OPSN)** uses data to encourage debate about the quality of public services in ways that engage and empower service users. Hosted at the Royal Society of the Arts, it is run by Charlotte Alldritt, a former government policy advisor, and chaired by Roger Taylor, once a *Financial Times* journalist who in 1999 co-founded Dr Foster, a ground-breaking information service that focussed on the performance of hospitals and that later attracted controversy when it was part-sold to the government[141].

OPSN published its first major report—*Empowering Parents, Improving Accountability*, about the performance of schools in England—in September 2013. To time with its publication, the *Guardian* newspaper released a **GCSE** Schools Guide portal that allowed parents to search through **OPSN**'s data to discover which schools performed best on which subjects in England, and how accomplished schools were at improving individual pupils' outcomes in exams. In 2015, **OPSN** followed up this work with a report highlighting subject deserts, Local Education Authorities (LEAs) in England where challenging subjects such as **triple science** and **modern languages** were sparsely offered or not offered at all.

The data

For its first report, *Empowering Parents, Improving Accountability*[142], **OPSN**'s aim was to see what existing data sources published by the government could tell parents about schools. It drew on open datasets published by the Department for Education (DfE) on **data.gov.uk**, including exam performance data, and pupil and school characteristics data.

The report also used inspection reports published by **Ofsted**, the government's schools regulator. It called on **Ofsted** to release the data contained in these reports in an accessible format that enables analysis, noting that "at present it mainly exists by school in .pdf format"[143].

For its second report, *Lack of options: how a pupil's academic choices are affected by where they live*[144] **OPSN** used data from the DfE's **National Pupil Database**. This dataset is not open data, and the pupil-level data it holds is deemed "personal" under the **Data Protection Act**. The DfE is authorised to share it at various levels of detail under strict terms and conditions with named bodies (including schools, local authorities, and other government departments). Third parties can apply to have access to the data—again, at various levels of detail—to conduct research or provide information services for the purpose of promoting the education or well-being of children[145]. The DfE provides guidance on how analysis of the data may be reported in order to protect the privacy of the individual pupils on the database[146].

OPSN accompanied both reports with data releases of its own, releasing the data that drove its analyses for others to re-purpose and re-use. The data is released in Microsoft Excel's Open XML format (**.xlsx**), under a **CC-BY licence**.

The path to impact

The aim of the project was to demonstrate that public service transparency, driven by open data can "support a richer and more multi-dimensional approach to accountability than is offered by the current exam results **league tables** and **Ofsted** report-based regime"[147]. **OPSN** believes that data on schools and other public services represents "a multitude of truths[148]" that are not captured by current accountability frameworks. In its report **OPSN** quotes research that shows that such frameworks have "relatively little influence on the choices people make"[149]."What good looks like varies against what you're trying to measure,[150]" says Charlotte Alldritt. Different service users want to ask different questions to evaluate public services according to their own contexts.

For the 2013 report, **OPSN** convened a panel of experts representing qualifications authorities, school governors, pupils, teachers, and others. Together they devised a set of features they thought represented the way parents thought about the quality of education offered by schools: facilities; climate of learning; curriculum and pupil outcomes. **OPSN** then set about interpreting the data sources available to see how fit they were at addressing variations in these features. The data they came up with included new measures, such as rates of uptake of different subjects at **GCSE**. This data was republished by the *Guardian* on a postcode-driven searchable portal that allowed parents to compare local schools, *The Guardian* **GCSE** schools guide[151].

It was during the research for *Empowering Parents, Improving Accountability* that Roger Taylor says he began to notice how many schools were not showing **GCSE** results for challenging subjects like the individual sciences (Physics, Biology, and Chemistry—commonly shortened to "**triple science**") and **modern languages**. This observation led to the second of **OPSN**'s education research projects, using **National Pupil Database** data to ascertain which schools were not offering these subjects across England.

The report identified six LEAs where 30% or more of schools had no pupils enrolled in **triple science**: Medway, Slough, Newcastle upon Tyne, City of Kingston Upon Hull, Knowsley, and North East Lincolnshire. In only 41 of England's 151 LEAs did all schools have at least one pupil enrolled in **triple science** (see Figure 7).

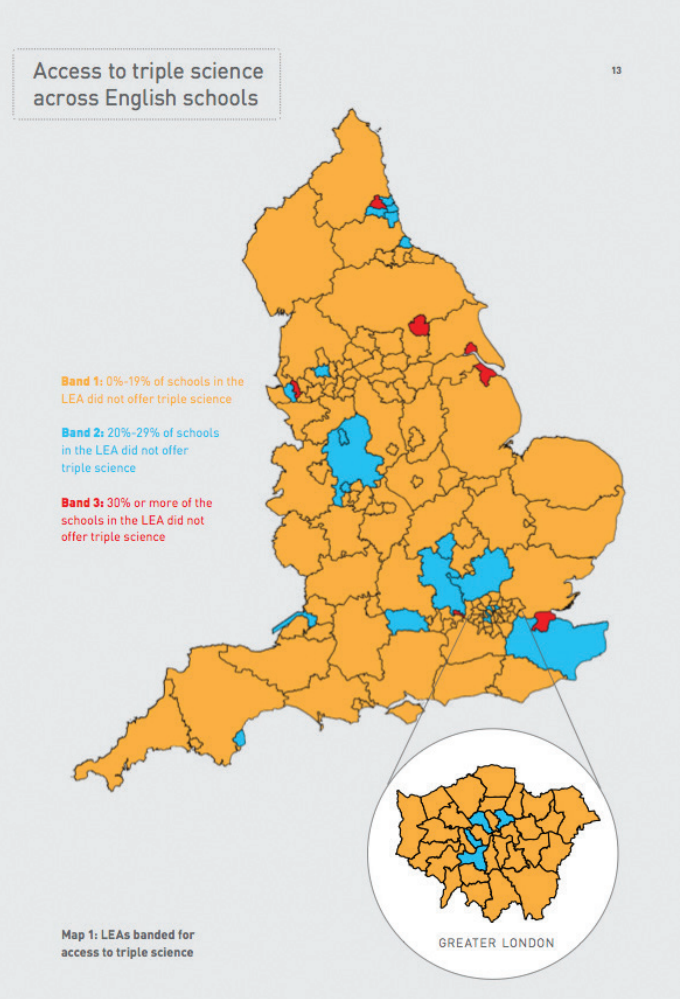


Figure 7: Access to triple sciences across English schools, taken from Lack of Options report. Copyright OPSN. Reproduced with permission.

The report then shifted its focus to all science **GCSE**s (including dual-award science, which covers subjects in biology, physics, and chemistry and is worth two **GCSE**s). This uncovered one LEA where some pupils were taking no science subjects at all (Knowsley). Mapping this data against deprivation data (available as open data on **data.gov.uk**) showed—with notable exceptions—that there was some relationship (see Figure 8). The report concluded that "fewer science **GCSE**s per pupil tend to occur in poorer areas"[152].

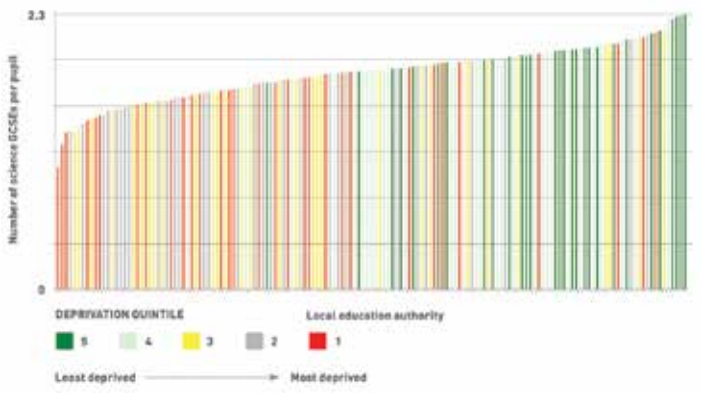


Figure 8: Number of Science **GCSE***s per pupil in LEAs coloured by deprivation quintile, taken from Lack of Options report. Copyright OPSN. Reproduced with permission.*

Impact

The **Cabinet Office**, which profiled the *Guardian* **GCSE** Schools Guide as an open data case study a month after its launch, reports that the portal attracted 20,000 users on its first day of release.[153] Given the number of pupils receiving **GCSE** results across England, Wales, Scotland, and Northern Ireland in 2015 was around 700,000, this is a significant proportion of the project's target audience. The **GCSE** Schools Guide has also been picked up by the World Bank[154] as an example of open data use cases in the education sector.

The 2015 *Lack of Options* report received substantial press coverage when it was published in 2015. Its headline, that some LEAs in England were subject deserts, was reported extensively by the BBC[155] and picked up by specialist education[156] news outlets and local papers[157,158,159] in the areas highlighted as performing poorly. The *Daily Mail*[160] also gave the story prominence, highlighting its critique of official **league tables**.

> "We cannot continue to allow generation after generation of pupils to be let down simply because of the accident of where they were born."
>
> Chris Skidmore MP

Chris Skidmore **MP**, a former member of the Education Select Committee, tabled a bill in the **House of Commons** in response to the *Lack of Options* report that would guarantee pupils the opportunity to study **triple science**, and quoted the report's findings extensively. In closing his speech, he stated:

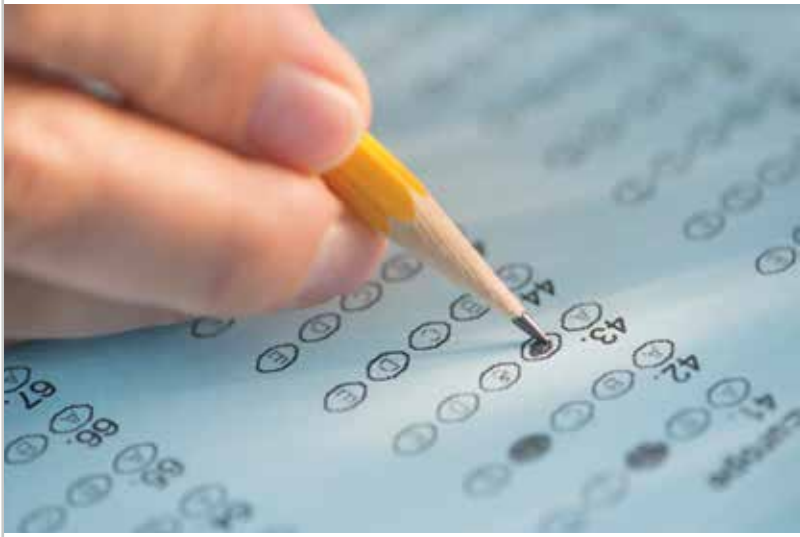
Poverty of aspiration, which lowers horizons and dims lights that should be burning brightly, still reaches into areas of our education system, and into places where education is most needed to transform young lives. We cannot continue to allow generation after generation of pupils to be let down simply because of the accident of where they were born or what school they attend.[161]

The **OPSN** succeeded in engaging its target audiences: the public, the media, and policymakers. This is impressive in itself. But beyond that, how can we understand its impact? Will **OPSN**'s intervention affect subject availability in deprived areas?

Like much proposed legislation put forward by **backbencher**s, Chris Skidmore's bill did not make it far through the UK's legislative process. In any case, an expert convening in response to *Lack of Options* attended by the author in June 2015 concluded that the problems highlighted by the report were far more likely to be addressed by broader education reforms. In the end, the *Lack of Options* report may prove to be just one piece of evidence put forward on the long journey towards policy reform[AB].

Discussion

The success of **OPSN**'s interventions in the education space mirrors that of an American project undertaken by ProPublica in 2011. *The Opportunity Gap* app used US Department of Education data to show that where some states, like Florida, offer rich and poor students roughly equal access to high-level courses, other states, like Kansas, Maryland, and Oklahoma, offer less opportunity in deprived districts. In his analysis of this project, ProPublica's Scott Klein writes:



AB For an interesting take on the role of evidence in policymaking, see (Maybin, 2013)

We … worked really hard at making sure the app told a "far" story and a "near" story. That is, the app needed to present the reader with a broad, abstract national picture—specifically, a way to compare how states did relative to each other on educational access. But given that abstraction sometimes leaves readers confused as to what the data means to them, we also wanted readers to be able to find their own local school and compare it to high and low-poverty schools in their area.[162]

OPSN's work also tells two different stories. The appeal to the public (through the **GCSE** schools guide) and the appeal to the press and policymakers (through the *Lack of Options* report) are subtly different, and point to an understanding that drives **OPSN**'s work: Service users want to ask different questions of public service data than service providers. At the heart of this approach lies a challenge to models of thinking about public service improvement through open data.

"People need simple heuristics to guide them in their lives and understand what is going on[163]," says Roger Taylor, Chair of the **OPSN**. He is deeply sceptical about the model of public service improvement through open data vaunted by David Cameron, who, when he announced new measures to publish local government spending in 2010, predicted the birth of "a whole army of effective armchair auditors looking over the books[164]".

High-profile cases of public spending waste have been exposed thanks to open data, for example, when hundreds of millions of pounds in potential annual savings on prescriptions for statins were identified in 2012[165] (for more on the impact of this case, see Introduction). But since the government's big push to publish more government data in order to improve public services in 2010, several commentators have noted that Cameron's army are, for the large part, still yet to advance.

In November 2012 the think tank Policy Exchange blamed the lack of armchair auditors on the fact that data like the Combined Online Information System of government spending (COINS) was "unusable".[166] And in April 2015 the Institute for Government pointed out that much of the data armchair auditors would need to hold the government to account was inaccessible, or of poor quality.[167]

But the data may only be half the problem. Roger Taylor says that while good data analysis doesn't take a large corporation to do, it does need significant resources:

Analysing complex datasets and trying to get useful signal out of all the noise is expensive and time-consuming. It isn't something you can do in your armchair …. It really does come down to small numbers of extremely talented and able people.[168]

Number of users of the *Guardian's* **GCSE** Schools Guide tool on the first day of its release: **20,000**

Drawing on his experience with Dr Foster, he worries that business models to support such operations "are very thin". Although he is clear that the Department of Health "never once attempted to use its influence to affect anything we said," after it paid £12m for a 50% stake in the company in 2006, "basic economics" meant the organisation ended up focussing on how to meet the information needs of NHS organisations and healthcare professionals, not how to meet the information needs of the public and patients:

Professionals always err towards the more cautious side, with the result that what you tend to get from more professional organisations are really complex presentations of data, with massive amounts of caveats about over-interpretation. They tend to prefer presenting the raw data rather than [stating] what it means, with the result that it means nothing to the public and they really can't act on it.[169]

This situation disempowers service users:

All this information about hospitals or doctors or whatever: if it can't be used to answer the question "should I have this treatment?", "should I let this doctor operate on me?" —if it can only be used inside the system—then all that does is simply allow the public to watch through the glass window and see how the professionals sort it out among themselves.[170]

Taylor adds that traditional media approaches to public service evaluation often mislead the public because "the way our media dialogue works is through over-simplification". But he also believes that even when the media take a more data-literate approach, the media business model doesn't support the sorts of personalised information services people need:

[To] get this information to the point where it works for an individual … and have a business model working you'll probably need a pretty intimate connection with that individual …. An old fashioned publishing model, where you try and earn revenues from publishing ratings or advice sheets or that kind of stuff—it's really hard to make that work.[171]

It's important to note that the *Lack of Options* report was produced using non-open data: a version of the **National Pupil Database** for which researchers need the approval of the Department for Education in order to access. Taylor believes that, through the open data agenda," we've set a standard for data release that is wholly inappropriate to most of the information that is of interest in terms of public services":

> "All this information about hospitals or doctors or whatever: if it can't be used to answer the question 'should I have this treatment?', 'should I let this doctor operate on me?' then all that does is simply allow the public to watch through the glass window and see how the professionals sort it out among themselves."
>
> Roger Taylor, **OPSN**

For anything to do with mental health, education, to make sense of what is going on you need to understand about the outcomes for people, what happened to them and data about their personal circumstances. There's no way [you can] put that data together in usable formats and then simply put it out under an **Open Government Licence**.[172]

"It is in the public interest that an understanding of what [the data says] does not reside wholly within government," says Taylor, but getting this data out of government and into the hands of approved third parties, for example on terms similar to those of the **National Pupil Database**, is hard. "It's a nasty political space to try and move forward on," says Taylor. He believes that vocal and effective privacy campaigners in the UK mean this type of government data sharing attracts negative press attention, while behind the scenes, vested interests such as professional bodies and government departments use their power to block progress. By contrast:

Open data is really straightforward. So you see why all the political focus has gone in that direction. But unfortunately, we aren't actually going to get many useful outputs from that process.[173]

Calls to action

FOR POLICYMAKERS

- **Policymakers should recognise the role personal data has to play in improving public services. It will never be appropriate to make such data generally available as open data, but neither is it appropriate for an understanding of what such data has to say about public services to reside wholly within government. Successive governments in the UK have damaged public trust on issues around the security and exploitation of sensitive personal data. If policymakers are serious about using data to improve public service, then strategies to regain that trust[AC], as well as tackle those vested interests that fear third-party scrutiny of public service delivery, should be top of their agenda.**

FOR OPEN DATA ADVOCATES

- **Open data advocates should recognise that releasing personally identifiable, sensitive data as open data is not appropriate.**
- **Open data advocates should partner with privacy advocates to encourage policymakers to engage in this "nasty political space" in ways that are technically literate and respect privacy.**

FOR FUNDERS

- **Funders need to be realistic about the capacity of volunteer "armchair auditors" to improve public services. Targeted support for skilled infomediaries should be part of the open data funding mix.**

AC For more on this see Conclusion. The contentious case of care.data, a scheme to centralise and share medical records previously held by individuals' GPs that collapsed under the weight of public criticism in 2014, has shown at a minimum that policymakers should prioritise communicating with the public clearly about how their data will be shared and with whom, and actively seek, rather than assume, the public's consent.

Summary

TheyWorkForYou.com is a parliamentary monitoring website that has been running for more than 10 years in the UK. It provides an accessible, searchable version of the official record of proceedings of the UK **Parliament**, as well as the devolved **Northern Ireland Assembly**. It also provides analysis of the voting records of elected and **non-elected representatives**. TheyWorkForYou predates open data reforms in the UK by a number of years, and is included in this report as a window onto the long-term impacts of outputs based on public data. One of TheyWorkForYou's long-term impacts appears to be encouraging parliamentarians to vote less with their party and more in the interests of the people they represent. It is also important to consider the potential monetised time savings delivered to the site's users—many of whom are civil society groups and journalists. These might well be in the millions of pounds a year, although methodological constraints mean it will probably always be impossible to put an exact figure on them.

Key takeaways

- **The greatest impact of TheyWorkForYou may be on Parliamentarians themselves. mySociety suspects, and some data confirms, that Parliamentarians have changed the way they go about their work in response to TheyWorkForYou's vote monitoring and analysis tools, both by turning up for more votes and rebelling against their party more often.**
- **Just as TfL's open data policy saves commuters' time, TheyWorkForYou.com saves time for its users—many of whom are civil society groups and journalists. As well as expecting websites with a social/political mission to achieve long-term positive social/political impacts by themselves, we can also expect them to save time for people trying to achieve those impacts through other means.**
- **Although it may look on the face of things like postcodes have nothing to do with Parliamentary monitoring, in fact postcode data is one of the key datasets that drive TheyWorkForYou.com. Early on in the website's history, mySociety had access to postcode data via a licence from Ordnance Survey. It now uses open data on postcodes provided by Ordnance Survey and the Office for National Statistics[AD]. But mySociety's lack of access to Royal Mail's Postcode Address File (see page 7) means it cannot deliver accurate information to all its users on who their elected representative is.**
- **mySociety views maintaining TheyWorkForYou.com as a "residual", and hence low priority, activity, and now focus most of their work on helping international NGOs. The long-term sustainability of projects like TheyWorkForYou.com is brought into question by this case.**

AD Although in the case of postcode information (for Northern Ireland) provided by the Office for National Statistics, this data does not conform to the open definition, as it is made available for non-commercial use only.

Background

TheyWorkForYou.com provides accessible, searchable data on the members and proceedings of **Parliament**, the main legislative body of the United Kingdom, as well as the **Northern Ireland Assembly**. Previous versions of the website also offered information on the members and proceedings of the Scottish **Parliament**. **TheyWorkForYou.com** provides a wide range of information, including members' voting records, speeches, and registered interests. The front page of the website invites users to answer the question "Does Your **MP** represent you?" by filling in their postcode to access analysis on the way their **constituency**'s **MP** votes and see their latest appearances in **Parliament** (See Figure 9 and Figure 10).

Figure 9: Screenshot of front page of TheyWorkForYou.com. Copyright mySociety. Reproduced with permission.

Figure 10: Screenshot from TheyWorkForYou.com. Copyright mySociety. Reproduced with permission.

TheyWorkForYou.com was launched in June 2004 by a group of volunteers "who thought it should be really easy for people to keep tabs on their elected **MP**s, and their unelected Peers, and comment on what goes on in **Parliament**"[174]. Individually, the original volunteers had already developed a number of civic-minded websites. Since 2006, **TheyWorkForYou.com** has been run by **mySociety**, a nonprofit social enterprise based in the UK that develops web platforms that "give people the power to get things changed".[175]

The data that powered **TheyWorkForYou.com** was **scrape**d from **Hansard**, the official record of parliamentary proceedings, published on **Parliament**'s own website. **TheyWorkForYou.com** launched despite the fact that this activity constituted a copyright infringement: The volunteers did not have the right to reproduce **Hansard**, which was covered by Crown Copyright. Later on, and in cooperation with some of the **TheyWorkForYou.com** volunteers, click-use licences were developed at the **Office for Public Sector Information (OPSI)** which among other things legitimised the site's activities[176AE].

The data

TheyWorkForYou.com uses multiple data sources. When asked about the most important datasets that drive the platform, **mySociety**'s Matthew Somerville, the site's lead developer, identifies the following datasets[177]:

1. **Postcode and constituency boundary data, for mapping people to their representatives**
2. **People, for each of the legislatures covered**
3. **Transcript data, for each of the legislatures covered, and multiple different types of data in that (e.g., actual debates, written answers, future business, committees, etc.)**

He continues:

The vote analyses are important, but are fundamentally just derived from the transcript data that includes the voting information. I'd possibly add our own dataset of everyone signed up for email alerts, as the daily emails we send based on information gleaned from the other datasets does presumably help power the site in a way[178].

AE Today, the official Parliament website is published under an Open Parliamentary Licence, which encourages re-use. In 2014, the development team behind the official Parliament website began releasing Parliamentary proceedings as open data.

Over 10-plus years operating **TheyWorkForYou.com**, Somerville has seen changes in how this data is provided. In terms of postcodes and **constituency** boundaries, since 2010, **TheyWorkForYou.com** has been using **Ordnance Survey**'s **CodePoint Open** open data product to match people's postcodes to their constituencies in Great Britain, and the **Office for National Statistics**' postcode product to perform the same function for users living in Northern Ireland. The latter source does not conform to the **open definition**, since it is provided for non-commercial use only. Before these products became available, **mySociety** had access to postcodes via a licence from **Ordnance Survey**.

The people data has been constructed over the years from various data sources, and **mySociety** makes it available under an open licence[179].

The transcript data for the **Northern Ireland Assembly** is accessed via the assembly's AIMS open data portal launched by the **Northern Ireland Assembly** in 2012. Since 2014[180] the development team of the official UK **Parliament** website, Parliament.uk, has also been making its data available via **API**. Despite this, **TheyWorkForYou.com** still **scrape**s transcript data from **Hansard** every morning to populate the site with the latest proceedings of the UK **Parliament**, rather than access the data via the **API**.

Tom Steinberg, outgoing CEO of **mySociety**[AF], explains that continuing to **scrape** official websites rather than taking advantage of new open data **API**s is mainly down to prioritising resources at **mySociety**:

If we've not used some of it [open data] it's quite often because the cost of us rewriting the software to use it is just not worth it when we can carry on screen **scraping**.[181]

The original **scraper** that provided **TheyWorkForYou.com** with transcripts of proceedings in the Scottish **Parliament** was developed by a volunteer, Mark Longair. But for the last couple of years **TheyWorkForYou.com**'s Scottish site has not been updated. Somerville explains:

[The] Scottish **Parliament** *changed their site a couple of years back and we haven't had time since to update the* **scraper**, *and not had anyone volunteer to help out, so [it's not been] updated since then. Don't think they have an* **API** *of any sort as yet.*[182]

The path to open

TheyWorkForYou.com played a key role in forming the UK's policy around open data. In 2007 Tom Steinberg was commissioned to co-write an influential review of the opportunities presented by opening up public sector information, a move he says traces directly back to his involvement with **TheyWorkForYou.com**:

Back in the **Labour** *era, Number Ten [i.e., the Prime Minister's office] was willing to talk to us as a group of people because TheyWorkForYou existed. Why were they willing to talk to us? Because they used it, they used it to look up their own parliamentary data and they couldn't use the* **Parliament** *website at the time because it was too [poorly designed]. And so unarguably TheyWorkForYou was a key to get in through the door of Number Ten. And it led pretty directly to the Power of Information review.*

One of the most interesting things about **TheyWorkForYou.com** is how central postcode data is to its operations. **mySociety** was part of the range of voices calling on **Ordnance Survey** to free its data before the 2010 **CodePoint Open** release, particularly boundary data, which is crucial in identifying a user's elected representative.

The way **mySociety** uses postcode data immediately personalises the experience of using TheyWorkForYou, and lowers the barrier to accessing the workings of democracy in a country where research shows only 22% of people can name their elected representative.[183] This observation underscores the utility of postcodes across a vast range of online applications.[AG]

Average monthly visits to **TheyWorkForYou.com**: **200,000-300,000**

Average monthly visits to **TheyWorkForYou.com** from people working at the Houses of **Parliament**: **4,000-6,000**

Proportion of users surveyed who say they are using it as part of their job: **30%**

Proportion of users surveyed who say they are getting information on their elected representative for the first time: **21%**

Users and outcomes

Somerville reports that **TheyWorkForYou.com** sends out around 30,000 emails a day to people who have subscribed to a range of notifications, from when their **MP** speaks in **Parliament** to when specific words are mentioned.

The most recent in-depth research into **TheyWorkForYou.com**'s user base was published by Tobias Escher of the Oxford Internet Institute in 2011[184]. It showed that the site receives between 200,000 and 300,000 visits every month

AF Tom Steinberg founded mySociety in 2003 and acted as its CEO until August 2015. He was interviewed for this report in June 2015.

AG Most postcodes in the UK match precisely to constituency boundaries, meaning the OS CodePoint Open product is generally sufficient to serve TheyWorkForYou.com's users. But there are exceptions. A 2013 Twitter conversation between Matthew Somerville and a TheyWorkForYou.com user reveals one occasion where the data did not match: Although inhabitants of the user's road all shared a postcode, those living in odd-numbered houses were represented by one MP, and those living in even-numbered houses by another (mySociety 2013) (Matthew Somerville tweets as @dracos). Access to the Royal Mail's Postcode Address File (see page 7) would allow TheyWorkForYou.com to provide all its users with accurate information about who represents them in the UK Parliament.

(see Figure 11—the peaks in May 2009 and May 2010 are respectively due to the **MP**s' expenses scandal, when many newspapers published links to **MP**s' pages on **TheyWorkForYou.com**, and the publication of an election quiz that matched voters to Parliamentary candidates based on each of their stated political beliefs, developed by **TheyWorkForYou.com** volunteers in the run-up to the 2010 elections).

TheyWorkForYou.com visitor statistics

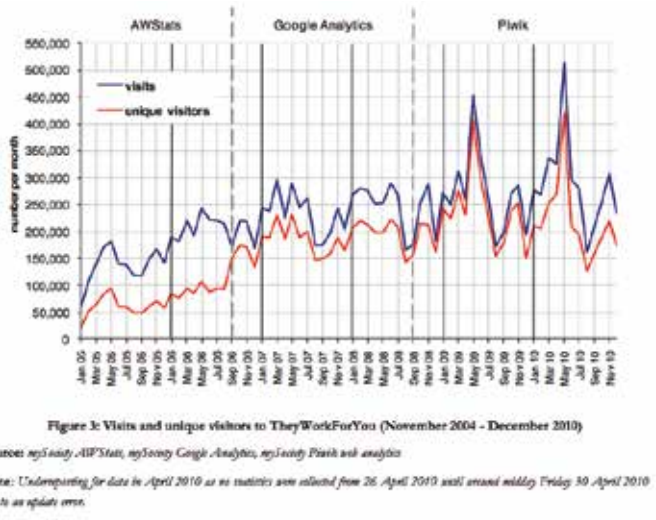


Figure 3: Visits and unique visitors to TheyWorkForYou (November 2004 - December 2010)

Source: mySociety AWStats, mySociety Google Analytics, mySociety Piwik web analytics

Note: Underreporting for data in April 2010 as no statistics were collected from 26 April 2010 until around midday Friday 30 April 2010 due to an update error.

Figure 11: TheyWorkForYou.com—visitor statistics. Taken from (Escher 2011). Copyright Tobias Escher. Reproduced under terms of CC-BY-NC 2.0 licence.

> "I don't think the political classes have got used to the idea of news or information that sticks."
>
> Tom Steinberg, **mySociety**

Escher reports that about half of **TheyWorkForYou.com**'s audience are regular users, and about one in five use it every month. 30% of respondents to a survey of 903 site users developed by Escher say they use the website for work, and usage patterns concentrated in UK business hours substantiate this. About 2% of visits come from IP addresses that indicate the user is working in the Houses of **Parliament**, and a further 2-3% indicate the user is working for the government. Workers at the BBC accounted for 0.5% of visits in 2010.

Impact

Tom Steinberg thinks that the site's impact on elected members is probably greater than on any other of its stakeholders:

I don't think the political classes have got used to the idea of news or information that sticks. They're very habituated to the idea that there's a bad news story today and it's gone tomorrow. All politicians can cope with that. However, a Wikipedia page or a TheyWorkForYou page … they're sort of permanent. They mean that if you're going to go and meet someone who doesn't know you, in the future, then you have every reason to suspect that that person will find out about you from this. So what these static pages say I believe influences what politicians do[185].

Several anecdotes support the claim that Parliamentarians are becoming increasingly aware of how their activities are reflected on **TheyWorkForYou.com**. In 2006 The Times newspaper speculated that some **MP**s were making interventions in **House of Commons** debates simply to keep their "appearances" tally on **TheyWorkForYou.com** sufficiently high, an event which prompted **mySociety** to rethink how it represented **MP**s' activity[186]. Steinberg relates a story he heard from someone who once had a meeting with an **MP** cut short so that the **MP** could go and vote in the **House of Commons**. When they asked whether the vote really mattered, the **MP** replied "They all matter since TheyWorkForYou". In 2008, Gordon Prentice **MP** complained to the leader of the **House of Commons** that **TheyWorkForYou.com** did not cover **MP**s' work on committees and therefore gave a "distorted impression" of the efforts **MP**s make at Westminster[187].

Philip Cowley, Professor of Parliamentary Government at the University of Nottingham, has published data showing that "**rebellion**" (that is, **MP**s voting against the party line, or—specific to Cowley's data—government **MP**s voting against government policy) is on the up in British politics. An Economist article from 2012 cites this data[188] and argues that vote-tracking websites are part of the reason why, as they allow voters and **constituency** parties to keep better track of their **MP**s:

One ringleader of [2011]'s huge **Conservative rebellion** *over a referendum on EU membership says that it was "incredibly easy" to persuade* **MP***s to join because they were already under intense pressure from their* **constituency** *associations. Gone are the days when an* **MP** *could vote with the government, then sign a contradictory early-day motion or two to muddy the waters, he adds. These days it is "harder to bluff".*

In recognition of the site's important accountability role, most of the new resources **mySociety** now devotes to **TheyWorkForYou.com** go into improving voting records and voting analysis. A recent grant from the Joseph Rowntree Reform Trust saw **mySociety** add about 60,000 more pages of vote analysis to the site.

The impact on the site's other users, be they civil servants, journalists, campaigners, or simply citizens interested in the workings of **Parliament**, is less visible. But Steinberg does not doubt that it is real:

I have no doubt that in some way [TheyWorkForYou] must oil the wheels of tons and tons of campaign groups and people who are kind of just generally in the political world … if it saves them five minutes here, that's five minutes they can spend on something else …. However, what I can't do is say "Here is someone who said 'I was going to lose my campaign, and then I used TheyWorkForYou and then I won it'".

For these communities, **TheyWorkForYou.com**'s impact is probably best understood through the same lens as the impact apps derived from **TfL**'s data have on London's commuters: It saves them time. One respondent to Escher's survey notes:

It is a hundred times easier to search than **Hansard** *itself. I've spent two hours on* **Hansard** *before coming here and have just found what I want in under a minute.*

Crudely, then, one might try and follow Deloitte's analysis of **TfL** and assume that 30% of the 200,000-300,000 monthly visits to **TheyWorkForYou.com** are work-related. If every one of these visits represented a 119 minute time-saving (and since time wasted sitting in traffic can be compared to time wasted looking for something and not finding it) one could use the Department for Transport's average value of working time (£34.12 per hour) to conclude that **mySociety** was saving this group of users roughly £5,075,350 worth of time every month, or just under £70 million worth of time each year. Taking a more realistic assumption, say that each visit represented Steinberg's 5 minute time saving, annual time savings delivered are just over £2.5million.

The analysis is flawed, of course. Unlike the Deloitte analysis, which takes advantage of **TfL**'s data on Lost Passenger Hours, we cannot assess how many hours working users of **TheyWorkForYou.com** might have spent trawling through **Hansard** and other information sources, not finding what they were looking for. It is included here simply to remind readers that although they may not make great stories, incremental time savings are an important and entirely valid form of impact to consider.

Third sector users have gone on record in praise of the website. **TheyWorkForYou.com**'s own FAQs contain a lengthy testimonial from Jo Brodie of Diabetes UK, part of which is reproduced below:

TheyWorkForYou.com *contributes to my 'current awareness' of what is being said about diabetes and insulin (access to treatment, statistics, etc.) and other related health topics (for example organ transplantation and stem cell research as that's very relevant for diabetes and its complications too). The email alerts and RSS feeds mean the information lands rather helpfully in my intray …. It's a great site—thank you*[189].

mySociety's team also receive private messages of thanks from policy experts, elected and unelected members, third sector workers, and others working in public affairs[190].

Escher notes from his survey data that **TheyWorkForYou.com** is reaching people who don't usually engage in civic activity:

While there are clearly TheyWorkForYou users who are already politically active and who are also [organised] in groups, the online survey shows that one in five users (21%) has neither been politically active (online and/or offline) within the last year nor been a member of any group and importantly, has got information for the first time on what his or her representatives are doing through the use of TheyWorkForYou[191].

TheyWorkForYou.com has also played a key role in helping other Parliamentary Monitoring Organisations (PMOs) around the world. Three years ago it changed its mission, and it now works internationally to support partners who deploy its technology in countries around the world. The code that runs **TheyWorkForYou.com** is open source, and has been adapted to create similar websites in New Zealand, Ireland, and Australia. In 2013, working with Kenyan PMO Mzalendo, **mySociety** released Pombola, a "less-UK centric" codebase that it actively helps PMOs to install in order to deploy parliamentary monitoring websites in their own countries. Pombola is already in use in Ghana, South Africa, Nigeria, and Zimbabwe.

> "It is a hundred times easier to search than **Hansard** itself. I've spent two hours on **Hansard** before coming here and have just found what I want in under a minute."
>
> Anonymous site user

TheyWorkForYou: Taking the long view

Discussion

TheyWorkForYou.com predates open data reforms in the UK by several years and can in fact be said to have contributed to open data policy's inception. This case study is included in the report because (although not all its data inputs are strictly open data) it is an example of an "output" project that repurposes public information and that has existed for over a decade. As such, it may provide a window through which to view the more long-term impacts of such public data re-use outputs.

Indeed, the increase in political **rebellion**, attributed by experts in part to the increased visibility of **MP**s' voting records that **TheyWorkForYou.com** provides, is a positive indicator that **TheyWorkForYou.com** is having a long-term impact. If politicians are choosing to serve the people who elect them more, and the party political machinery less, then this is a good outcome for representative democracy.

Of interest too is the amount of time **TheyWorkForYou.com** may be saving civil society groups and journalists (and, of course, corporate lobbyists too). The monetised time savings used in this case study should not be relied upon and do not bear repeating, since unlike in the **TfL** case they are based on speculation about the amount of time such users may have lost in seeking out information elsewhere, and not recorded Lost Passenger Hours. The calculations have rather been included here in order to remind readers that as well as expecting websites with a social/political mission to achieve long-term positive social/political impacts by themselves (in this case better representative democracy) we can also expect them to save time for people trying to achieve those impacts through other means.

Although Matthew Somerville sees no reason why **TheyWorkForYou.com** won't survive another 10 years ("**Parliament** will certainly still be around"[192]), it should be of some concern that Tom Steinberg describes the site as a "residual" from before **mySociety** changed its mission to focus on helping international NGOs develop similar websites through projects like Pombola. Somerville sees lots more work for mySociety to do with **TheyWorkForYou.com**, and says that even if the official website becomes easier to use, many of **TheyWorkForYou.com**'s most important functions—like vote analysis and email alerts—may never be provided by **Parliament**:

There are many things the site still doesn't cover, but it's not like we have had countless volunteers [clamouring] to add select committees, or what have you.[193]

This observation further calls into question how realistic it is to anticipate sustained activity from **civic hacker**s (or indeed David Cameron's "whole army of effective armchair auditors"[194]) in response to the release of government data as open data. In fact, that pool of volunteers may be more limited than first imagined. **mySociety** has done recent, funded, work on TheyWorkForYou, the 60,000 additional pages of vote analysis supported by the Joseph Rowntree Reform Trust. But the fact that the site does not serve its Scottish users as at July 2015 because no one has volunteered to update the tool so that it can **scrape** the Scottish **Parliament**'s new website should be of major concern.

Calls to action

FOR OPEN DATA ADVOCATES

- **The Deloitte study on the impact of TfL data on transport users, and its analysis that TfL are delivering transport users annual monetised time savings of between £15m and £58m, is one of the most quoted stories of open data impact in policy circles. The Deloitte methodology can be borrowed to discuss impact in other spheres too.**

FOR OPEN DATA DEVELOPERS

- **Shiny new projects are fun and inspiring, but established sites like TheyWorkForYou.com need volunteer developers too. Open data developers should consider whether knowing you're contributing work to a project that is having long-term positive impacts for representative democracy makes up for some of the hassle of working on someone else's code.**

FOR FUNDERS

- **Projects like TheyWorkForYou.com need long-term commitment, and may not be able to rely on volunteer efforts to keep going.**

Summary

Gluten-Free on the Move is an award-winning app designed to help coeliac sufferers adhere to the strict **gluten**-free diet that is the only treatment for their condition. It is the result of a data-sharing agreement between **Coeliac UK**—a patient self-help society—and **Brandbank**—Europe's largest provider of digital content for online retail portals. The **Brandbank**/**Coeliac UK** partnership shares several features of other cases in this report, but exhibits one important difference: The data is not open data, **Brandbank** licenses it to **Coeliac UK**. The case is included in this report to underline the importance of considering the counterfactual when evaluating open data impact. Those examining open data impacts should be ready to ask the question: Are these impacts only happening because data is open? What else about this dataset is contributing to these impacts?

Key takeaways

- **Coeliac UK has transformed Brandbank's data: It now serves a market Brandbank would probably never reach on its own, with important social and health impacts.**
- **Brandbank operates in a highly standardised and regulated environment. Barcodes[AH] were the backbone to this project's success. Barcodes enabled Coeliac UK to bring its information on gluten-free foods to a whole new platform—smartphones—and transformed the way coeliacs could shop for food.**
- **Coeliac UK found Brandbank, asked for access to the data, and got it. Open data policy played no role. Those examining open data impact should be prepared to ask the question: Would impact still have occurred if the data in question was not open data?**
- **Moving to an open data approach might benefit Brandbank's business: The suppliers and brands who pay to be part of the Brandbank database do so to achieve "high visibility" for their products on online retail platform and other third parties.**

AH Specifically, the GTIN barcode data standard, maintained by GS1 (GS1 n.d.)

Background

Coeliac disease is an auto-immune disease caused by intolerance to **gluten**. Complications associated with untreated **coeliac disease** include osteoporosis, small bowel cancer, and non-Hodgkin's lymphoma. Around 1 in 100 people are thought to suffer from **coeliac disease** in the UK, although only a quarter of them are diagnosed. **Coeliac disease** is caused by a reaction of the immune system to **gluten**—a protein found in wheat, barley, and rye. When someone with **coeliac disease** eats **gluten**, their immune system reacts by damaging the lining of the gut.

Coeliac disease has no cure, and the only treatment for **coeliac disease** is a strict **gluten**-free diet for life[195]. Maintaining this regime is tough, since many staples of the Western diet include **gluten** (e.g., bread, pasta), processed foods, soups, and sauces can contain hidden **gluten**, and foods intended to be **gluten** free can easily become contaminated in kitchens where **gluten** is present (e.g., through cooking equipment, especially toasters and deep fat fryers).

Coeliac UK was founded in 1968 by Peter Benenson, the founder of Amnesty International and himself a coeliac sufferer, and Elizabeth Segall, mother to a coeliac child[196]. It is "probably one of the best national coeliac societies in the world" and

Coeliac UK and Brandbank: A view from the other side of open

was one of the first patient self-help societies to be established in the UK[197]. Today, it has over 60,000 members, who pay a membership fee of between £12 and £30 per year depending on their circumstances. The charity receives no funding from government, and supplements its membership income with fundraising activities. The charity's work includes supporting its members in maintaining their **gluten**-free diet through various information sources and local groups, campaigning on policies that affect its membership such as food labelling and prescription of **gluten**-free foods, and funding research into the condition.

Brandbank was established in 1998 and is based in Norwich. It describes itself as a "world leader in providing rich digital content" and its core business is in supplying detailed product data and images of food and pharmaceutical products to online retailers. Its retail clients include Boots, Tesco, and Waitrose and it currently holds details of 250,000 products from over 6,000 different suppliers in its database. In January 2015 **Brandbank** was acquired by global information and measurement company Nielsen Holdings in a multi-million dollar deal[198].

In June 2014, **Coeliac UK** launched its *Gluten-Free on the Move* app, allowing users to scan food product barcodes to ascertain whether they were safe for coeliacs to consume[199]. The app contains details of around 15,000 branded and supermarket own-brand products, and is powered in part by **Brandbank** data.

The data

Brandbank captures "full back-of-pack data as well as high resolution images of each product"[200]. The company's preferred method is to capture data from a physical sample of the product. Suppliers send **Brandbank** a sample of the product, and **Brandbank** then photograph it in a professional studio to provide all the images of the product that consumers will later see on e-retail platforms such as Amazon.

While these images are in post-production, data entry teams capture the detailed nutritional and **allergen** information online suppliers and retailers are bound by law to communicate to consumers, so that consumers buying the product online can make as informed a choice as if they were holding the package in front of them in a shop. The backbone of this data capture is the product barcode standard **GTIN**, maintained by **GS1**, which sets set rigorous standards around, for example, when changes to products necessitate new barcodes.

Once the supplier has approved the images and data captured by **Brandbank**, it goes live on **Brandbank**'s database of over 250,000 products. Online retailers then access the data to populate their retail platforms, using a number of approaches including **Brandbank**'s own **API**.

The **Brandbank** dataset is not open data. **Brandbank** owns the intellectual property in the images and data it captures, and it licenses this data to the retailers that use it. Agreements with the retail platforms who partner with **Brandbank** are subject to commercial confidentiality. When asked to elaborate on the **Brandbank** business model, and specifically whether **Brandbank** charges retailers for access to their database, Beau Archer, Technical Pre Sales Consultant at **Brandbank**, responded:

Giving retailers access to our product content is core to the **Brandbank** *business. Ecommerce is becoming a more important revenue engine for Brands, and [they] therefore want to ensure that the consumer has visibility of their products. Visibility of products comes through the retailer and 3rd parties.*[201]

Brandbank was unwilling to go into further detail on their business model, but this statement could be taken to read that suppliers pay to be in the database, and retailers are given access for free. Indeed, the **Brandbank** website indicates suppliers pay **Brandbank** an annual subscription.[202]

Beau Archer states that **Brandbank** generally "[does] not control how retailers use [its] content"[203]. The exact terms under which **Brandbank** licenses its data to retailers are also subject to commercial confidentiality. But **Brandbank** has developed its own browser extension for consumer use, Shopper Goggles, which may shed some light on the priorities of the company when licensing its data. The free tool allows users to layer **Brandbank** data over retail platforms not partnered with the company, and contains (among others) the following licensing conditions:

You are permitted to download Shopper Goggles solely for your own personal and private use. You must not:

- *reverse compile, disassemble, reverse engineer, decompile, copy, duplicate, modify or adapt any software or other code or scripts forming part of Shopper Goggles; or*
- *publish, distribute, sub-license or sell Shopper Goggles to anyone else; or*
- *use Shopper Goggles in contravention of any applicable law.*[204]

Even if large retailers do access **Brandbank** data for free, this may not be the case for all of its data re-users. **Brandbank** differentiates between 3rd party users of its data:

As a supplier has a focus on high visibility of its products, **Brandbank** *wants to ensure that we are appealing to all ranges and types of 3rd parties. This includes the large retailers, however there are other types of businesses that may require this data such as an eRetailer, nutritional apps etc., which we would call syndication.*[205]

Users

Coeliac UK began producing lists of foods suitable for coeliacs in 1968[206]. One of the key benefits of becoming a member of **Coeliac UK** today is receiving a copy of their annual *Food and Drink Directory*, which lists thousands of brands and supermarket own-brand products that are suitable for coeliacs to eat. As well as detailing foods available on prescription and those certified **gluten**-free, the directory focuses on the hard-to-judge products that could contain hidden **gluten**, such as ready-meals, breakfast cereals, soups, and sauces. In this section of the directory, **Coeliac UK** list foods that have no **gluten**-containing ingredients, and that are produced in environments where sufficient care is taken to avoid cross-contamination with **gluten**. Throughout the year, **Coeliac UK** posts updates to the directory, where food products have changed in some identifying detail (e.g., name/weight) but are still safe, or where **Coeliac UK** has received information that food products are no longer safe.

For many years, **Coeliac UK** liaised directly with suppliers and food manufacturers to ascertain which products could be listed in its directory. The directory is also available electronically, in the members area of the **Coeliac UK** website. But although a **smartphone** app associated with the data was an obvious next step for **Coeliac UK**, it was held back by how hard it was to match products to the most obvious input for such an app: their **GTIN** barcode identifiers. Kathryn Miller, Food Policy Lead at **Coeliac UK**, says the department at the supplier organisation who dealt with product barcodes was always different from the one they were dealing with to ascertain ingredients and manufacturing processes, making matching data hard.

Number of people in the UK diagnosed with **coeliac disease**: **>160,000**

Percentage of a sample of coeliacs who report being **glutened** in a six month period: **54%**

Members of **Coeliac UK**: **60,000**

Number of downloads of *Gluten-Free on the Move*: **70,500**

Increase in products listed thanks to data-sharing partnership: **+50% (~5,000 products)**

Coeliac UK's relationship with **Brandbank** began in 2013. One of **Brandbank**'s directors is a coeliac and Beau Archer says the initial approach from **Coeliac UK** represented "a perfect fit to work together"[207], since **Brandbank** captures all data that appears on product packaging, including **allergen**s such as **gluten**. Crucially for **Coeliac UK**, **Brandbank**'s database is built around the **GTIN** barcode standard, so an app based on their data would allow users to scan barcodes of products in-store, and find out immediately whether the product was listed in the directory as suitable for them to eat.

After gaining access to the **Brandbank** database, **Coeliac UK** hired a data company to help them work out a set of rules they could apply to the data to identify candidate products for the app. Once these rules are applied to the data, **Coeliac UK** performs extensive manual checking based on its in-depth sector knowledge. This both ensures the products identified are suitable for coeliacs, and checks that no products have been excluded by the filter that are suitable. The resulting app, *Gluten-Free on the Move*, also includes details of cafes, restaurants, and other venues accredited by **Coeliac UK** as catering well for coeliacs. It launched for **iOS** and **Android** platforms in June 2014.

Coeliac UK and Brandbank: A view from the other side of open

Outcomes

In the month of its launch, *Gluten Free on the Move* won "Best Audience Migration to Mobile Technology" at the 2014 UK Mobile & App Design Awards. In September of the same year, Kathryn Miller won TalkTalk's Healthy Living Digital Hero Award[208].

> "Love using the scanner on my new **Coeliac UK** app it's genius! Thank you for making **gluten**-free life simpler"
>
> *Gluten-Free on the Move* user

Coeliac UK reports that so far, the app has been downloaded 70,500 times[209], indicating it may be reaching considerably beyond **Coeliac UK**'s member base. The app provides free services including food label guides, a **gluten**-free checklist and eating out tips, but only **Coeliac UK** members can access the food directory and venue guide elements. **Coeliac UK** reports that the app is typically used on 5,000 devices each week, and that it is used in excess of 2,000 times a day resulting in more than 200 food products a day being matched and 100 venues being looked at.

Coeliac UK is yet to develop the app for other platforms such as Windows mobile, to the ire of some of its members[210]. In August 2015 **Coeliac UK** announced a new feature letting users download the full dataset to their phone, so they can refer to it in the absence of a network connection.

Impact

A study[211] published in 2013 involving 287 coeliac sufferers in North East England found that 54% had inadvertently consumed **gluten** in the past 6 months, even though 84.7% of respondents reported that they intended to keep to the very strict **gluten**-free diet that is the only treatment for their condition. Only 15.7% of respondents reported never having inadvertently consumed **gluten** since diagnosis.

Clearly, even with the best intentions, sticking to the strict **gluten**-free diet required to treat **coeliac disease** is hard. The same study found that perceptions of difficulty—often reinforced by the frequency with which sufferers make mistakes and get **glutened**—can feed into negative self-efficacy that in turn discourages some sufferers from maintaining their treatment.

The study reports that:

When asked what would make sticking to the diet easier, the most frequently cited responses related to better quality, choice, cost and availability of **gluten***-free food, followed by improved awareness and understanding, clearer and universal product labelling and clearer information when eating out.*[212]

The *Gluten Free on the Move* app contributes to several of these goals, not just in terms of information but in terms of non-specialist products the app helps coeliacs consume with confidence.

For the purpose of this report, **Coeliac UK** shared a number of tweets praising the app.

"Love using the scanner on my new @Coeliac_UK app it's genius! Thank you for making gluten-free life simpler"

"Impressed with @Coeliac_UK app for mobiles. Something to have a play with over next day or two. Thanks to all involved in its design."

"Whoa ho I have the long awaited @Coeliac_UK app. #makeslifeeasierforcoeliacs out and about. Yippee"

"Just downloaded the @Coeliac_UK app which is out today. Fantastic to have the gluten-free food & drink directory in electric on-the-go format"

"@Coeliac_UK the #gf (girlfriend) has already given it a go and is impressed. Will make #gf (glutenfree) life much easier! Great work, thanks"

"Great app will definitely help us getting our 5 year olds diet right @Coeliac_UK #glutenfree"

"@Coeliac_UK many thanks for info packed, easy to navigate app. Will certainly make my #glutenfree life much easier and safer. Great charity!"

*"New @Coeliac_UK app is AWESOME! Venue map and barcode scan function are superb. Been practising already! *very happy face*"*[213]

Although some app users find it frustrating that more products are not listed by the app, the partnership with **Brandbank** has meant that **Coeliac UK** has already been able to increase the number of products listed as suitable from 10,000 to more than 15,000.

Kathryn Miller says an increase of this magnitude is probably a one-time thing, rather than a rate they can expect the directory to expand at year-on-year. Nonetheless, the aim of **Coeliac UK** is to use the **Brandbank** data to gradually continue the expansion of listed products. Right now, **Coeliac UK** is "just getting to grips with"[214] the possibilities using **Brandbank** data has for their mission. For example, monthly updates, which used to number between 10 and 50, now number in the hundreds. The vast majority of these are additions (new products that are suitable for coeliacs—overall the market for **gluten**-free foods is growing rapidly[215]) and amendments to existing products (weight, brand name, etc.), rather than deletions (foods no longer suitable for coeliacs that once were).

Discussion

The case of **Brandbank** and **Coeliac UK** has been included in this report as a signpost to those examining open data impact that they should consider the counterfactual. Are benefits being delivered by this or that open data project because the data is open? Or because the data is rich enough, real-time enough, or relevant enough to a particular community? Could impact have plausibly been delivered had the data remained closed, shared only with trusted third parties and under conditions set forth by the data owner?

The social, health, and economic impacts of the *Gluten Free on the Move* app may be hard to quantify, but they are impossible to deny. Like some of the other cases outlined in this report, this impact has been achieved thanks to a specific set of technical conditions—the proliferation among the general UK population of **smartphone**s connected to a nearly always-on internet, and the existence of a highly standardised and rich dataset. Also like other cases, the data in question has been supplied by one entity, **Brandbank**, and transformed by another, **Coeliac UK**, such that it fulfils a purpose not envisaged by those who originally collected it.

And yet the data is not open data. **Coeliac UK** found **Brandbank** and asked their permission to use the data. That permission was granted, and the result is a relationship that after one year is already benefitting the lives of thousands of coeliacs.

I asked Beau Archer whether anyone at **Brandbank** was thinking about how open data approaches might play a role in their business moving forward, and he declined to answer. Beau Archer's rather cryptic description of **Brandbank**'s business model, where brand visibility is achieved through giving access to the data to third parties, appears in many aspects to match what the **Open Data Institute** calls a "cross subsidy model" of open data business[216]. In a cross-subsidy model, data publishers derive value from opening up their data by reaching more customers, or providing enhanced services to existing customers, through wider sharing and use of the data. As such, the "cross-subsidy model" looks like a good fit for **Brandbank**.

Calls to action

FOR OPEN DATA ADVOCATES

- **Brandbank UK may be operating closer to an open data business model than they realise. Their data is rich and standardised and has the potential to benefit many more groups than just Coeliac UK. Governments are obvious targets for open data advocacy because of their public service mission and extensive data collecting activities. Private companies like Brandbank will require a different advocacy approach[AI].**

AI The Open Data Institute has published the first in a series of research outputs on open data and business, see (Open Data Institute, 2015)

Much of this report is good news. In almost every case, the theory of change behind government open data policy is upheld: Private actors have taken government data, and they have transformed it in ways that are useful and valuable to citizens and consumers. Far more time and money has been invested in government data than it is possible to imagine the government ever having done by itself. The impact of this investment, though not always quantifiable, is in most cases tangible and scalable, if not already "at scale".

The cases surface a number of issues around open data impact that are worth discussing in order to broaden our understanding of the challenges re-users of government data face, challenges that may present barriers to impact.

The first issue surfaced by the cases is that there are significant data gaps in the UK that may be acting as barriers to impact. These gaps divide roughly into three categories: data that is collected by government agencies but is not yet released openly; data that is not yet collected by any government agency but could be; and data that is collected by government agencies but is not suitable for release as open data because it contains (often sensitive) personal information.

In terms of data the government has but is not releasing openly, the **HM Land Registry** case reveals gaps in government data releases that are blocking the development of meaningful end-user services. These gaps are mainly around **Ordnance Survey** data, which infuses **HMLR**'s **INSPIRE index polygons** with third-party intellectual property rights that need separate licensing from **OS**, normally for a substantial fee (see box on page 19)[217]. The lack of bulk data on property and land ownership was also cited as an issue by some of the **proptech** developers interviewed for this report, although this may have significant privacy implications where those records relate to individuals.

Policymakers need to make good on their commitment to open data, and release the **core reference data** needed to guarantee that the UK's **national information infrastructure** exists as open data. A history of open data reforms in the UK that has been dominated by the **trading fund** issue, the precedent set by the sale of the **Postcode Address File** during the privatisation of **Royal Mail** (see box on page 7), and a political climate in the UK currently driven by the need to boost the public purse, all indicate that open data advocates will need to adopt a strong position.

The campaign for **beneficial ownership transparency** (detailed in the case study on OpenCorporates) is a campaign to close a data gap by urging governments to collect new categories of data. Jonathan Gray and Tim Davies have called this sort of activity the creation of "**participatory data infrastructures**"[218]. Their argument, that most government open data is an administrative by-product (what the **Shakespeare Review of Public Sector Information** called "**exhaust PSI**"[219]) and may not therefore capture aspects of its subject needed by data re-users such as advocates, is radical and relevant. Developers interviewed for this study who called for **HM Land Registry** to collect different sorts of data (in this case the number of bedrooms and the square footage of a sold property) might also be viewed through this lens, as might **OPSN**'s demands for the government to begin surveying teachers in a standardised way as part of the information it collects on education services[220].

Where data gaps like this exist, policymakers and open data advocates need to be ready to collaborate with other domain-specific stakeholders to create the **participatory data infrastructures** that will close them. Open data advocates should recognise that, in contrast to the first phase of open data reforms in the UK that demanded the release of data that already existed, this second phase will require significantly more investment from government. Greater resistance from policymakers and civil servants should therefore be expected.

The third set of data gaps relate to **personal data**, and come into play in both the **OPSN** study and the **TfL** case. Alongside open data about schools, **OPSN** used the **National Pupil Database**, which is not open and contains **personal data**, to produce their *Lack of Options* report. **OPSN**'s Roger Taylor characterises policy around allowing approved third-party scrutiny of this sort of **personal data** as a "nasty political space"

when compared to the "really straightforward" matter of open data reform. **TfL** is navigating the space between exploiting the **personal data** it holds on its customers through pay systems like **Oyster** and the Santander Cycle Scheme, while still maintaining a strong market able to profit from its open data.

While it remains true that most **personal data** should never be released as open data[AJ], policymakers, open data advocates, and open data developers cannot avoid engaging with the **personal data** agenda anymore. Particularly when it comes to data used to scrutinise and improve public service delivery, all these stakeholders should be ready to work with experts in the fields of data anonymization and re-identification, consent, and data-handling, in order to chart a way through this space. The contentious case of care.data, a scheme to centralise and share medical records previously held by individuals' GPs that collapsed under the weight of public criticism in 2014[221], has shown at a minimum that policymakers should prioritise communicating with the public clearly about how their data will be shared and with whom, and actively seek, rather than assume, the public's consent.

Beyond data gaps, the second issue surfaced by the cases is that organisations transforming government data into social and political impacts need some help to achieve long-term sustainability. In the study of the **Open Public Services Network (OPSN)**, Roger Taylor points to a lack of armchair auditors willing or able to meaningfully interrogate much of the data around public service performance. Business models that allow organisations both to retain the "extremely talented and able people" Taylor says this work needs, and still to maintain a mission focussed on the needs of public service users (as opposed to public service providers), have yet to emerge.

Meanwhile, although **mySociety** has succeeded in keeping parts of **TheyWorkForYou.com** running for over 10 years, a lack of volunteers means it has been unable to update the section of the website that covers the proceedings of the Scottish **Parliament** for some time. This is despite the fact that the UK **Parliament** section of **TheyWorkForYou.com** is delivering significant impact to the third sector through time savings, as well as influencing the behaviour of Parliamentarians in ways that benefit representative democracy.

Funders need to take note here. Although the rise of the internet and computer processing power is driving down the cost of the sorts of independent scrutiny provided by **TheyWorkForYou.com** and **OPSN**, they and organisations like them also need to invest in people with a unique mix of data skills and governance knowledge. Prime Minister David Cameron predicted open data policy would trigger a volunteer "army of effective armchair auditors[222]" who would interrogate government data. In fact, that army has been slow to advance.

The third important issue surfaced by the cases is that even where both the open data and the expertise needed to interpret it exist, in some cases substantial advocacy work beyond the activities of **infomediaries** is needed to achieve impact. OpenCorporates had the data and the skills needed to demonstrate the complexity of corporate ownership structures. But it took a concurrence of factors including political mood, external events, and focussed, concerted efforts from advocacy organisations to move the needle on the **beneficial ownership transparency** issue. This case shows the critical role open data expertise can play in advocacy, but equally how much work that has nothing to do with open data needs to happen in order to achieve successful advocacy outcomes. Echoes of this can be heard in the **OPSN** study too.

This last point is relevant to the entire open data community. In some domains, and the **TfL** case study shows us transport is one, simply releasing data openly can have a huge impact. But in very many others—tackling corruption, improving mental health outcomes, arresting climate change, achieving gender equality—the problems we might aspire to solve are deeply complex, highly contextual, and not always contingent on improved knowledge flows. Expecting open data to fix them all by itself is only inviting disappointment.

Understanding information markets

Another equally important group of issues emerges in this report, which indicates that the open government data community could benefit from a greater understanding of the nascent information market that open government data is helping to shape. OpenCorporates is the largest open database of companies in the world, and Chris Taggart is aware of the power he now holds as its CEO:

We get approached about every 3 or 4 months by VCs wanting to talk to us about investment. And we say no, because we absolutely never want to have an exit. Because an exit means a trade sale to somebody that's going to use that power.[223]

Vernon Everitt, Managing Director for Customer Experience, Marketing and Communications at **TfL** is conscious that the market for re-users of **TfL** data may consolidate:

It's important that we preserve the ability of apps developers to take this stuff and to make products quickly. If this became just a big corporate enterprise, I think that would work against the principles of openness.[224]

The rise of the world wide web as a global communications platform has witnessed rapid market consolidation that many of the internet's early pioneers did not expect. Opening public data should not equal privatising public data, and so far, it has not. Yet we should be conscious of how little we know about the information market created through government open data. It may be the right time to consider what measures should be put in place to ensure the market for government data re-use remains open, diverse, and competitive.

AJ The author sides with those in the open data community who define open data in opposition to personal data. Nonetheless, the author recognises that in some circumstances, for example public interest registries, parliamentary records and company ownership, personal data can and should be published as open data. For an in-depth discussion of open data and privacy, see (Zuiderveen Borgesius, Gray, & van Eechoud, Forthcoming).

For example, funders could help organisations like OpenCorporates to explore appropriate mechanisms to ensure their public interest motivations are embedded in their governance structures. They could also support research that expands on theories put forward by Tim Wu—on the cyclical nature of information empire-building and the role of competition authorities in information industries[225]—to characterise the market for open government data and explore how we might expect it to change over time. Such research would also be a legitimate investment for policymakers.

Open data: The next phase for the movement

The success of the open data movement in achieving mainstream endorsement and adoption is remarkable. A previous study by the author[226] speculated that this success was achieved in the UK due to a collaborative atmosphere among policymakers, civil servants, and "**civic hacker**s". In part, that atmosphere has been enabled by open data's broad appeal as a policy. It's fair to say that often open data's potential to deliver public sector efficiencies and economic growth is put forward to policymakers by advocates who have not economic but transparency and accountability outcomes at the core of their mission.

This collaborative atmosphere may have outlived its usefulness. Anti-privatisation campaigners fear that the current **Conservative** government in the UK intends to sell off **trading fund**s such as **Ordnance Survey**, **Companies House** and **HM Land Registry** in a move to boost the public purse. The sale of the **Postcode Address File** along with **Royal Mail** in 2013 has set a dangerous precedent (see box on page 7). Government has an interest in promoting economic growth and public welfare through open data policies; private companies not so much. Privatisation of public data—and in particular the **core reference data**sets collected and distributed by **trading fund**s—would represent a body blow to the open data movement, and would redefine the future of the information society.

Even if the **trading fund**s remain in public hands, they will still need persuading before they open up their data fully. Tom Steinberg of **mySociety** was closer than most in the community to government thinking on open data until he resigned his position on the UK's Public Sector Transparency Board in the spring of 2012. He believes that parts of government—and particularly **HM Treasury**, which wields significant power over the future of the UK's **trading fund**s—have never been convinced by the economic arguments used to support open data:

Who can blame them when the argument is only ever … speculative? A risk-taking country would probably just say, "Alright, yes, we'll do this. It might work, it might lose us loads of money." That's not exactly the philosophy of the Treasury.[227]

This observation has echoes of the **TfL** case study. Vernon Everitt, the champion of open data policy inside **TfL**, says that if he had been asked to write a traditional transport industry business case for opening **TfL**'s data, he would still be writing it now. Instead, **TfL** took a calculated risk on its open data policy, and was repaid in millions of pounds of monetised time savings delivered to its users by apps based on its data (as well as equally considerable internal cost savings).

In one sense, an eternal business case writing-exercise is exactly what's happening with the UK's **trading fund**s. Open data advocates should be concerned that, even while UK policymakers drag their heels releasing key datasets (and even consider privatising them), the country's top position on two separate open data indices allows UK leaders to tour the world vaunting their open data—and open government—credentials[AK].

For his part, Steinberg believes it is time for open data advocates to adopt the "struggle mentality" that has been conspicuously absent from the story so far, and incorporate demands for open data into a rights-based agenda, similar to the movements to establish Freedom of Information laws across the world. Other advocates I spoke to echo his feelings. Chris Taggart of OpenCorporates says he finds difficult a narrative around open data that emphasises showing the evidence base:

It's a bit like [asking for] the evidence basis for a justice system. We need a justice system because that's the society we want, one with a justice system, with the rule of law …. Let's stop trying to argue for open data, and make [opponents] argue for closed data, right? Let's stop trying to look for unicorns and start to have quiet, dirty, unsexy work as being a measure of success.[228]

While the author hopes that this report has contributed to the evidence base for open data impact, she shares these advocates' views that open data fits just as well into a rights-based agenda, and that—so long as the privacy of individuals is respected—the public has a right to access and interrogate public data, and to share what it learns in so doing. Open data is not just an economic policy, it is a democratic imperative. It is time for the movement to stop playing nice.

AK It's also worth noting the launch, in July 2015, of a cross-party review into the Freedom of Information Act "likely to be viewed as an attempt to curb public access to government documents" (Syal, Freedom of Information Act review 'may curb access to government papers' 2015). For a discussion on how open data may have contributed to a clouding of our ideas of what an open government is, see (Yu & Robinson, 2012)

The author recognises that the UK-focus and often technical nature of this report requires a little orientation for non-specialist or non-UK readers. This glossary of terms should help. Those requiring a grounding in the UK's political institutions, how they function, and their relationships to one another other, may find the following two resources helpful:

- **How government works**
  https://www.gov.uk/government/how-government-works
- **Categories of public bodies**
  https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/80075/Categories_of_public_bodies_Dec12.pdf

.csv
A common data format. Short for "comma separated values", Open Knowledge calls it "probably the simplest possible structured format for data". For more, see http://data.okfn.org/doc/csv

.xlsx
A data format developed by Microsoft as part of their Open Office XML suite.

allergen
A substance that can produce abnormal reactions from the immune system. The EU specifies 14 allergens, including all cereals containing **gluten**, peanuts, soy, eggs, and milk, which must be labelled on the packaging of any food that contains them.

Android
An operating system for **smartphone**s developed by Google.

API
A programming approach/tool that allows computer programs to share data resources. Crucially, it allows its owner to control how much of the data and the software that organises it is shared. Short for Application Programming Interface.

API key
An identifying credential provided by someone running an **API** to someone who wants to use it.

Autumn Statement
One of two official statements the **Chancellor** makes to **Parliament** every year about the UK economy.

backbencher
A Parliamentarian who holds no government office (or, if she is a member of the opposition, is not an official spokesperson on any issue for the party). Often more likely to **rebel** against party policy.

beneficial ownership transparency
A campaign to collect and publish information about corporate structures that would make it easier to identify the true beneficiaries of the companies' activities. Relevant to cases of financial corruption involving anonymous shell companies.

Big Bang
Common term for a series of financial sector reforms of the 1980s, including the introduction of screen-based trading on the London Stock Exchange.

Brandbank
A UK company that provides full back of package information on thousands of food products to online retailers.

Cabinet Office
A department of the UK government that provides support to the Prime Minister and the government, in a coordinating role. It is the department of government that has championed the open data agenda, particularly under the stewardship of Francis Maude **MP**, who served as its lead minister between 2010 and 2015.

calls
(on an **API**)—requests for data.

CC-BY licence
A type of Creative Commons licence—an off-the-peg copyright licence that permits the re-use of content so long as its original creator is attributed.

Chancellor
Short for Chancellor of the Exchequer—the government minister in charge of **HM Treasury** and the UK's economic and financial affairs. Understood to be the most powerful office in government after the Prime Minister.

civic hacker
Common term for an individual who uses data and computer programming tools to advance public goods.

civil service
The bureaucracy/secretariat that supports the UK government.

coalition government
A government formed of at least two different political parties, usually after no one party has won an overall majority in a General Election. The UK had its first coalition government in several decades between 2010 and 2015, made up of the centre-right **Conservative** party and the social-liberal **Liberal Democrat**s.

CodePoint Open
The **Ordnance Survey**'s postcode product, released under an **Open Government Licence**.

coeliac disease
An auto-immune disease that causes the body to attack itself in the presence of **gluten** damaging the lining of the gut. The only cure is a strict **gluten**-free diet for life.

Coeliac UK
A patient self-help society for sufferers of **coeliac disease** and their carers. It has over 60,000 members.

Companies House
A government agency responsible for maintaining the register of companies in the UK.

Congestion Charge
A scheme to alleviate congestion in central London. It uses a system of cameras that recognise number plates to charge motorists driving through central London on a weekday £11.50.

Conservative
A centre-right political party in the UK.

constituency
An electoral district.

conveyancing
The transfer of legal **title** of property from one person or entity to another.

core reference data
A term for the most valuable data held by the government, commonly understood to be data that helps refer to things in the real world, such as geospatial data, data about legal entities such as companies and data about land and property ownership. Its use as a term came to prominence after the **Shakespeare Review of Public Sector Information** recommended the UK government establish a **national information infrastructure** based on its most valuable data sets.

Crossrail
A major rail infrastructure project crossing London east-west, that is expected to open in 2019.

data aggregator
A service that collects and packages data from different publishers.

Data Protection Act
Legislation in the UK that is designed to promote privacy by giving people some power over how organisations can process their data.

data wholesaler
A commercial **data aggregator**.

data.gov.uk
The UK government's open data portal.

Digital Culture, Services Platforms and Data Board
A new board overseen by the UK's Department for Business, Innovation and Skills that took over the responsibilities of the **Public Data Group** in April 2015.

DirectGov
The precursor to **gov.uk**—the UK government's digital service.

Docklands Light Railway
A public transport service that serves the redeveloped Docklands area in East London.

estate agent
A business/person who acts on behalf of a property owner during the sale of a property. Roughly equivalent to a "realtor" in the US.

exhaust PSI
A term used in the **Shakespeare Review of Public Sector Information** denoting data "generated through the performance of regular activities that are not data collection specific".

GCSE
The most common type of exam taken by school children in England at age 16. Short for General Certificate of Secondary Education.

GLA
see **Greater London Authority**.

gluten
A protein found in wheat, barley, and rye.

glutened
A term used by coeliacs to mean eating **gluten** inadvertently.

Google Now
Personal assistant software developed by Google that proactively delivers information to its users based on their personal locational, browsing, and search histories.

gov.uk
UK government information website.

Greater London Authority (GLA)
The strategic regional authority for London.

GS1
An international nonprofit responsible for maintaining the barcode standard.

GTIN
Part of a barcode—a unique identifier between 8 and 14 digits long. Short for Global Trade Item Number.

Hansard
The official record of the proceedings of **Parliament**.

Health and Safety notice
A formal enforcement document issued by the Health and Safety Executive (a public body in the UK that regulates health and safety conditions in the work place). A public register of Health and Safety Enforcement notices is made available here: http://www.hse.gov.uk/notices/.

HM Land Registry (HMLR)
A non-ministerial department of the UK government that registers the ownership of property and land in the UK.

HM Treasury
The UK government's economic and finance department.

HMLR
see **HM Land Registry**.

House of Commons
The lower house of **Parliament**, the supreme legislative body of the UK. It contains 650 elected members.

House of Lords
The upper house of **Parliament**, the supreme legislative body of the UK. It currently has 790 members of which 92 have inherited the office, 26 are bishops and the remainder are appointed for life by the Prime Minister.

HS2 rail project
A proposed new rail link between London and Birmingham.

ICO
see **Information Commissioner's Office**.

iGoogle
A content aggregator or personalised home page launched by Google in 2005 and popular in the late 2000s. It was discontinued in 2013.

infomediary
Neologism denoting an individual or organisation that transforms data—usually sourced from a third party—to serve the needs of a particular target audience.

Information Commissioner's Office (ICO)
The regulator in charge of enforcing (among others) the **Data Protection Act** and the Freedom of Information Act in the United Kingdom.

INSPIRE index polygons
Part of the EU's INSPIRE scheme (short for Infrastructure for Spatial Information in the European Community), a computational model of the land mass of the UK, where each polygon is the legal boundary of a parcel of registered land, on an extract of the **Ordnance Survey** map with a unique identification (ID) number (see http://blog.landregistry.gov.uk/inspire-index-polygons).

iOS
An **smartphone**- and tablet-focussed operating system developed by Apple.

Labour
A centre-left political party in the UK.

Land Register
The register of land and property ownership in England & Wales.

league tables
Data about examination results published by the schools regulator **Ofsted**.

Liberal Democrat
A social-liberal political party in the UK.

linked data
Structured data published in a way that allows it and its various properties to be queried and linked to.

London Datastore
The open data portal of the **GLA**.

London Underground
London's public rapid transit system, much of which runs underneath the capital. Also known as the **Tube**.

Lough Erne
Venue for the 39th G8 summit, held in 2013. It is a resort on the southern shore of a large lake system in Northern Ireland.

Member of Parliament (MP)
An elected member of the **House of Commons**.

modern languages
Common term used in the education system of England & Wales to denote the study of, e.g., French, Spanish, Italian, Chinese, Russian, as a second language.

MP
see **Member of Parliament**.

mySociety
UK NGO whose aim is to build websites that empower citizens.

national information infrastructure
Term coined in the **Shakespeare Review of Public Sector Information** for a collection of high-value datasets, including but not limited to the nation's **core reference data**. The work to establish the UK's national information infrastructure is ongoing.

National Pupil Database
One of the richest education datasets in the world, this database contains pupil-level information about test results, prior attainment, and progression for all state-run schools in England. It began collecting data in 2002.

NetVibes
A content aggregator or personalised home page. Originally similar to **iGoogle**, and used in the same way in the late 2000s, it has since reinvented itself as a business intelligence platform.

non-elected representatives
A term for members of the **House of Lords**.

Northern Ireland Assembly
The devolved legislature of Northern Ireland, it has the power to legislate in areas including education, health, and culture.

ODI
see **Open Data Institute**.

Office for National Statistics
A non-ministerial department charged with the collection and publication of statistics.

Office for Public Sector Information (OPSI)
Created in 2005, **OPSI** advises on the regulation of public sector information and its re-use. It is part of the National Archives, an executive agency of the government.

Ofsted
The regulator of schools in England, short for the Office of Standards in Education, Children's Services and Skills.

OGP
see **Open Government Partnership**.

ONS
see **Office for National Statistics**.

open data
Data that can be freely used, modified, and shared by anyone for any purpose.

open definition
A precise definition of open developed by the Open Knowledge community. As applied to data, its main stipulations are:

- that data must be published under an open licence with no additional terms that contradict the licence;
- that data is available to download in bulk (for free or for no more than a one-time reproduction charge); and
- that data is available in an open format "such that there are no unnecessary technological obstacles to the performance of licensed rights".

The full definition is available here:
http://opendefinition.org/od/

Open Data Institute (ODI)
Independent, nonprofit, organisation founded by Sir Tim Berners Lee and Sir Nigel Shadbolt to "catalyse the evolution of open data culture". Its initial funding came from the UK government.

open data value chain
The variety of organisations and individuals who create, transform, and consume open data, and the people and issues affected by those activities.

Open Government Licence
An off-the-peg copyright licence for data and documents published by the UK government, first released in 2010.

Open Government Partnership
Launched in 2011, the Open Government Partnership is "a multilateral initiative that aims to secure concrete commitments from governments to promote transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance". See http://www.opengovpartnership.org

Open Public Services Network (OPSN)
A programme based at the **Royal Society for the Arts** that uses data to encourage debate about the quality of public services in ways that engage and empower service users.

OPSI
see **Office for Public Sector Information**.

OPSN
see Open Public Services Network.

Ordnance Survey (OS)
The national mapping agency of Great Britain.

OS
see **Ordnance Survey**.

Overground
London's suburban rail network.

Oyster
London's smartcard ticketing system.

Parliament
The main legislative body of the UK.

participatory data infrastructures
Phrase coined by Jonathan Gray and Tim Davies. It refers to the planning of new data collection activities where that planning involves civic actors, and highlights the role of public information systems in shaping collective life.

PDC
see **Public Data Corporation**.

PDG
see **Public Data Group**.

personal data
Information that relates to an identified or identifiable individual, as defined by the **Data Protection Act**.

Police.uk
Official information and data portal of the police force of England, Wales, and Northern Ireland.

Postcode Address File
Database maintained by **Royal Mail** that holds a record of all known delivery points in the UK, including domestic addresses but also commercial premises.

Price Paid Dataset
Dataset administered by **HM Land Registry**, it contains over 20 million transactions for residential properties sold for full market value going back to 1995.

Private Eye
The UK's best-selling current affairs magazine.

proptech
Businesses using data and technology to innovate in the property sector.

Public Administration Committee
A Parliamentary Select Committee that scrutinises the administrative efforts of the **civil service**.

Public Data Corporation (PDC)
Short-lived government-established company that the Department for Business briefly intended to have a role in the UK's government data strategy. It was superseded by the **Public Data Group** and the Data Strategy board, following consultation with stakeholders.

Public Data Group (PDG)
A grouping of the UK's data-handling **trading fund**s, chaired by Claudia Arney. It was formed in 2011 to improve the quality and quantity of data being released from its members: **Companies House**; **HM Land Registry**, **The Met Office**, and **Ordnance Survey**. It dissolved in 2015.

rebel/rebelling
[as a **Member of Parliament**] voting against the party line.

Royal Mail
The UK's postal service, it was part-privatised in 2013.

Royal Society for the Arts (RSA)
An organisation established in 1754 to promote the arts, manufacture, and commerce. Today, its mission is to "create the conditions for the enlightened thinking and collaborative action needed to address today's most pressing social challenges".

RSA
see **Royal Society for the Arts**.

Santander Cycle Hire Scheme
A public bike hire scheme in London, previously known as "Barclay's Bikes" and colloquially referred to as "Boris bikes" after the Mayor of London, Boris Johnson.

scrape/scraping/scraper
Scraping is harvesting or extracting structured data from a resource, like a webpage, designed to be read by a human. A scraper is a software tool designed to fulfil this goal.

Shakespeare Review of Public Sector Information
Independent Review of Public Sector Information conducted by the founder of polling company YouGov, Stephan Shakespeare, between 2012 and 2013, at the request of the **Cabinet Office** and the Department for Business. The review commissioned Deloitte's "Market Assessment of PSI" report. Its conclusions includes recommendations to establish a **National Data infrastructure**.

Shoreditch
Area of London close to the financial district known for technology innovation and trendy haircuts. It is home to Tech City (a technology startup cluster) and the **ODI**.

smartphone
Mobile phones with superior computing power.

Student Loans Company
A government-owned company set up in 1989 to provide loans to students.

TfL
see **Transport for London**.

The Met Office
The UK's national weather service.

TheyWorkForYou.com
Website designed and maintained by **mySociety** that provides an accessible, searchable version of the official record of the proceedings of **Parliament**.

title
Evidence of land and property ownership.

title plan
Map showing extent of a piece of land or property.

trading fund
A type of government agency that has the authority (under the Government Trading Funds Act 1973) to meet its outgoing costs from the money it charges for services.

Tramlink
A tram system that serves parts of South London.

Transport for London (TfL)
The local government body responsible for implementing transport strategy and managing transport services across London.

triple science
Commonly-used shorthand to denote the individual study and certification of Biology, Physics, and Chemistry at **GCSE** level. This is in contrast, for example, to studying these subjects as "dual award science"—where the three subjects are combined into two exams.

Tube
see **London Underground**.

Whitehall
Term used to refer to the **civil service**, after the address in London where a large part of it is based.

widget
A code snippet designed to be embedded in a third-party website.

Notes

1 (Hogge 2011)
2 (Gray and Davies, Fighting Phantom Firms in the UK: From Opening Up Datasets to Reshaping Data Infrastructures? 2015)
3 (Prime Minister's Office 2010)
4 (Hogge 2011)
5 (Atz, Heath and Fawcett 2015)
6 (Robinson, et al. 2009)
7 (World Wide Web Foundation 2015)
8 (Open Knowledge Foundation n.d.)
9 (Royal Mail n.d.) "Postcode Address File"
10 (Hope 2013)
11 (*BBC News* 2014)
12 (Jee 2015)
13 (Caplan, et al. 2014)
14 (Craveiro, et al. 2014) as cited in (Davies, Open Data in Developing Countries: Emerging Insights from Phase I 2014)
15 (Caplan, et al. 2014)
16 (Thwaites 2012)
17 (Williams 2015)
18 (McGee and Gaventa 2010)
19 (Keserű 2015)
20 Rebecca Rumbul, personal communication
21 (Atz, Heath and Fawcett 2015)
22 (Deloitte 2013)
23 (Lateral Economics 2014)
24 (Lateral Economics 2014)
25 (*BBC News* 2015)
26 (Transport for London n.d.) "What we do"
27 (Transport for London n.d.) "What we do"
28 Interview: Vernon Everitt
29 (Transport for London n.d., Transport for London 2015) "How we are funded"
30 (Topham 2014)
31 (Reed 2015).
32 (Transport for London n.d.) "Transport Data Service terms and conditions"
33 Interview, Phil Young
34 (Siddle 2014)
35 (Mirani 2014)
36 Interview, Phil Young
37 Interview, Vernon Everitt
38 Private correspondence, Phil Young and Vernon Everitt
39 Interview, Vernon Everitt
40 Interview, Paul Clarke
41 Interview, Emer Coleman
42 Interview Phil Young
43 Interview, Emer Coleman
44 (Everitt 2014)
45 (Reed 2015)
46 (Reed 2015)
47 (Deloitte 2013)
48 Interview, Vernon Everitt
49 (Transport for London 2015)
50 (Baxevanis 2015)
51 Interview, Vernon Everitt
52 Interview, Vernon Everitt
53 Interview, Phil Young
54 (Deloitte 2013)
55 (Reed 2015)
56 Interview, Phil Young
57 Interview, Vernon Everitt
58 Interview, Vernon Everitt
59 Interview, Paul Clarke
60 Interview, Phil Young
61 Interview, Vernon Everitt
62 Interview, Vernon Everitt
63 (Land Registry 2006)
64 (Land Registry 2006)
65 (KPMG 2011)
66 (Hogge 2011)
67 (Arthur, Met Office and Ordnance Survey to be part of 'public data corporation' 2010)
68 (Department for Business, Innovation & Skills and Cabinet Office 2011)
69 (Cabinet Office 2011)
70 (Department for Business, Innovation & Skills and Cabinet Office 2011)
71 (Syal, Land Registry privatisation plans abandoned by ministers 2014)
72 (Quinn 2015)
73 (We Own It n.d.)
74 (Public Data Group n.d.)
75 Interview, Lynne Nicholson
76 (Land Registry 2012)
77 (Land Registry 2013)
78 http://archive.is/pFEzX#selection-1665.0-1665.251
79 (Private Eye 2015)
80 (Land Registry 2014)
81 (legislation.gov.uk n.d.) "Copyright, Designs and Patents Act 1988"
82 (Department for Business, Innovation & Skills and Cabinet Office 2011)(22 of the 116 respondent were identified by the government as "market participants")
83 Interview, Lynne Nicholson
84 (Department for Communities and Local Government 2010) "Property sales based on Land Registry data"
85 (Department for Communities and Local Government 2010) "Median house price"
86 (HM Treasury 2012)
87 Interview, Lynne Nicholson
88 Interview, Lynne Nicholson
89 (Cabinet Office 2011)
90 Interview, Lynne Nicholson
91 (Land Registry 2013)
92 (Mapping London 2015)
93 (Ramsay 2014)
94 (Timita 2014)
95 (Ramsay 2014)
96 Interview, Adrian Black
97 Interview, Adrian Black
98 Interview, Adrian Black
99 Interview, Henry Pryor
100 (Assetti 2015)
101 (Fowler 2015)
102 Interview, Peter Thum-Bonanno
103 Interview, Peter Thum Bonanno
104 http://developer.zoopla.com/
105 Interview, Vasanth Subrahmanian
106 Interview, Vasanth Subrahmanian
107 Interview, Henry Pryor
108 (HMRC 2015)
109 (OECD n.d.)
110 (HMRC 2015)
111 Interview, Lynne Nicholson
112 (Land Registry 2013)
113 Interview, Lynne Nicholson
114 Interview, Peter Thum-Bonanno
115 Interview, Claudia Arney
116 (Aston 2015)
117 (Hope 2013)
118 (Arthur, MPs and open-data advocates slam postcode selloff 2014)
119 (van de Does de Willebois, et al. 2011)
120 (OECD 2001)
121 (OpenCorporates 2010)
122 (ODI n.d.) "Case study: OpenCorporates"
123 (Cabinet Office 2013)
124 Interview, Robert Palmer
125 Confidential source, Cabinet Office
126 Confidential source, Cabinet Office
127 Interview, Robert Palmer
128 (G8 2013)
129 (Cameron 2013) (my emphasis)
130 Interview, Chris Taggart
131 Interview, Robert Palmer
132 Interview, Chris Taggart
133 Interview, David McNair
134 Private communication, Robert Palmer
135 (Transparency International 2014)
136 Interview, Chris Taggart
137 (Gray and Davies, Fighting Phantom Firms in the UK: From Opening Up Datasets to Reshaping Data Infrastructures? 2015)
138 (Gray and Davies, Fighting Phantom Firms in the UK: From Opening Up Datasets to Reshaping Data Infrastructures? 2015)
139 (Gray and Davies, Fighting Phantom Firms in the UK: From Opening Up Datasets to Reshaping Data Infrastructures? 2015)
140 (Prime Minister's Office 2010)
141 (digitalhealth.net 2007)
142 (Open Public Services Network 2013)
143 (Open Public Services Network 2013)
144 (Open Public Services Network 2015)
145 (Department for Education 2013)
146 (Department for Education 2013)
147 (Open Public Services Network 2013)
148 Interview, Charlotte Alldritt
149 (Hansen, Joshi and Dex 2010)
150 Interview, Charlotte Alldritt
151 (Adams 2013)
152 (Open Public Services Network 2015)
153 (Cabinet Office 2013)
154 (World Bank n.d.)
155 (Coughlan 2015)
156 (Times Educational Supplement 2015)
157 (Nottingham Post 2015)
158 (Bradford Telegraph and Argus 2015)
159 (Sampson 2015)
160 (Harding 2015)
161 (Skidmore 2015)
162 (Gray, Bounegru and Chambers, The Data Journalism Handbook: How Journalists Can Use Data to Improve the News 2012)
163 Interview, Roger Taylor
164 (Prime Minister's Office 2010)
165 (Thwaites 2012)
166 (Wheeler 2012)
167 (Freeguard, Munro and Andrews 2015)
168 Interview, Roger Taylor
169 Interview, Roger Taylor
170 Interview, Roger Taylor
171 Interview, Roger Taylor
172 Interview, Roger Taylor
173 Interview, Roger Taylor
174 (TheyWorkForYou.com n.d.) "About TheyWorkForYou"
175 (mySociety n.d.)
176 (Hogge 2011)
177 Email-based interview, Matthew Somerville
178 Email-based interview, Matthew Somerville
179 (Nixon 2014)
180 (Hadi 2014)
181 Interview, Tom Steinberg
182 Email interview, Matthew Somerville
183 (*BBC News* 2013)
184 (Escher 2011)
185 Interview, Tom Steinberg
186 (Hurst 2006)
187 (Prentice 2008)
188 (The Economist 2012)
189 (TheyWorkForYou.com n.d.) "Frequently Asked Questions"
190 Private communication, Myf Nixon, mySociety marketing and communications manager
191 (Escher 2011)
192 Interview, Matthew Somerville
193 Email interview, Matthew Somerville
194 (Prime Minister's Office 2010)
195 (Coeliac UK n.d.)
196 (Wikipedia n.d.)
197 (Howdle 2010)
198 (Brandbank 2015); (Mills & Reeve 2015)
199 (Coeliac UK 2014)
200 Interview, Beau Archer
201 Interview, Beau Archer
202 (Brandbank n.d.) "FAQs" (See question "Why do I have to pay an annual fee?")
203 Interview, Beau Archer
204 (Shopper Goggles n.d.)
205 Interview, Beau Archer
206 (Coeliac UK n.d.) "History of Coeliac UK"
207 Interview, Beau Archer
208 (TalkTalk Digital Heroes n.d.)
209 Private communication, Kathryn Miller
210 (Coeliac UK 2014)(see comment stream underneath post)
211 (Hall, Rubin and Charnock 2013)
212 (Hall, Rubin and Charnock 2013)
213 Private communication, Kathryn Miller
214 Interview, Kathryn Miller
215 (Booth 2014)
216 (ODI n.d.) "How to make a business case for open data"
217 (Land Registry 2014)
218 (Gray and Davies, Fighting Phantom Firms in the UK: From Opening Up Datasets to Reshaping Data Infrastructures? 2015)
219 (Deloitte 2013)
220 (Open Public Services Network 2013)
221 (Triggle 2014)
222 (Prime Minister's Office 2010)
223 Interview, Chris Taggart
224 Interview, Vernon Everitt
225 (Wu 2010)
226 (Hogge 2011)
227 Interview, Tom Steinberg
228 Interview, Chris Taggart

Adams, Richard. "School database lets parents compare GCSE results by subject." *The Guardian*. 11 September 2013. http://www.theguardian.com/education/2013/sep/11/school-database-gcse-results-subject.

Arthur, Charles. "Met Office and Ordnance Survey to be part of 'public data corporation'." *The Guardian*. 19 November 2010. http://www.theguardian.com/technology/2010/nov/19/government-public-data-corporation.

—. "MPs and open-data advocates slam postcode selloff." *The Guardian*. 17 March 2014. http://www.theguardian.com/technology/2014/mar/17/mps-and-open-data-advocates-slam-postcode-selloff.

Assetti. "The Rise of the New Property Technology (Prop Tech)." *assetti.pro*. 14 June 2015. http://assetti.pro/en/2015/06/14/the-rise-of-the-new-property-technology-prop-tech/.

Aston, S. "Graham Farrant to replace Ed Lester as Land Registry chief." *Civil Service World*. 13 March 2015. https://www.civilserviceworld.com/articles/news/graham-farrant-replace-ed-lester-land-registry-chief.

Atz, Ulrich, Tom Heath, and Jamie Fawcett. "Benchmarking open data automatically." *ODI*. 2015. https://theodi.org/guides/benchmarking-data-automatically.

Baxevanis, Alexander. *FOI request: Santander Cycles app*. 13 May 2015. https://www.whatdotheyknow.com/request/santander_cycles_app.

BBC News. "London's population hits 8.6m record high." *BBC News*. 02 February 2015. http://www.bbc.co.uk/news/uk-england-london-31082941.

—. "Ministers criticised for sale of postcode database." *BBC News*. 17 March 2014. http://www.bbc.co.uk/news/business-26605375.

—. "Three quarters of people 'cannot name their local MP'." *BBC News*. 16 May 2013. http://www.bbc.co.uk/news/uk-politics-22555659.

Booth, Susanna. "The growth of gluten-free food: 10 of the best products ." *The Guardian*. 14 May 2014. http://www.theguardian.com/lifeandstyle/2014/may/14/growth-gluten-free-10-best-products.

Bowden, Thomas. "FOI request: Overseas company properties title numbers." *WhatDoTheyKnow.com*. 8 September 2015. https://www.whatdotheyknow.com/request/overseas_company_properties_titl.

Bradford Telegraph and Argus. "Fears over job prospects of Bradford pupils who fail to study languages at GCSE." *Bradford Telegraph and Argus*. 11 February 2015. http://www.thetelegraphandargus.co.uk/news/localbrad/11786686.Fears_over_job_prospects_of_Bradford_pupils_who_fail_to_study_languages_at_GCSE/.

Brandbank. "FAQs." *Brandbank*. https://www.brandbank.com/about/faqs/ (accessed September 30, 2015).

—. "Nielsen acquires Brandbank." *Brandbank*. 22 January 2015. https://www.brandbank.com/blog/brandbank-news/nielsen-acquires-brandbank/; .

Cabinet Office. "Case study: Guardian Good GCSE Guide." *gov.uk*. 24 October 2013. https://www.gov.uk/government/case-studies/open-data-guardian-good-gcse-guide .

—. *Making Open Data Real*. 04 August 2011. https://www.gov.uk/government/consultations/making-open-data-real.

—. *Policy paper: Open data measures in the Autumn Statement* 2011. 29 November 2011. https://www.gov.uk/government/publications/open-data-measures-in-the-autumn-statement-2011.

—. "Prime Minister David Cameron's speech to the World Economic Forum in Davos." 24 January 2013. https://www.gov.uk/government/speeches/prime-minister-david-camerons-speech-to-the-world-economic-forum-in-davos.

Cameron, D. "HC Deb, 19 June 2013, c913." *TheyWorkForYou.com*. 19 June 2013. http://www.theyworkforyou.com/debates/?id=2013-06-19a.893.0#g913.4.

Campbell-Dollaghan, Kelsey. "Check Out These Totally Not Shocking Maps of Corporate Tax Avoidance." *Gizmodo*. 17 July 2013. http://gizmodo.com/check-out-these-totally-not-shocking-maps-of-corporate-818076424.

Caplan, Robin, Tim Davies, Asiya Wadud, Stefaan Verhulst, Jose M. Alonso, and Hania Farhan. "Towards Common Methods for Assessing Open Data." May 2014. opendataresearch.org/sites/default/files/posts/Common%20Assessment%20Workshop%20Report.pdf.

CASA blog network. "Category arhive: TfL." *CASA blog network*. http://blogs.casa.ucl.ac.uk/category/TfL/ (accessed September 30, 2015).

Coeliac UK. "About coeliac disease and dermatitis herpetiformis." *Coeliac UK*. https://www.coeliac.org.uk/coeliac-disease/about-coeliac-disease-and-dermatitis-herpetiformis/ (accessed September 30, 2015).

—. "Coeliac UK's smartphone app available on iPhone and Android devices now." *Coeliac UK*. 27 June 2014. https://www.coeliac.org.uk/about-us/news/coeliac-uks-smartphone-app-available-on-iphone-and-android/.

—. "History of *Coeliac UK*." Coeliac UK. https://www.coeliac.org.uk/about-us/our-organisation/history-of-coeliac-uk/ (accessed September 30, 2015).

—. "Post announcing launch of new version of *Gluten Free on the Move*." *Facebook*. 29 September 2014. https://www.facebook.com/CoeliacUK/posts/811396572214734 .

Computer Weekly. "OpenCorporates makes company data public." Computer Weekly. 11 July 2013. http://www.computerweekly.com/news/2240187869/OpenCorporates-makes-company-data-public.

Coughlan, Sean. "Pupils in some areas are not offered 'vital' GCSEs." *BBC News*. 11 February 2015. http://www.bbc.co.uk/news/education-30983083 .

Craveiro, G.S., J.A.S Machado, A.M.R Martano, and T.J Souza. "Exploring the impacts of web publishing budgetary information at the sub-national level in Brazil." 2014.

Davies, Tim. *Open Data in Developing Countries: Emerging Insights from Phase I*. ODDC, Worldwide Web Foundation, 2014.

—. "UK Open Data Timeline (to Feb 2014)." *OKFN Timemapper*. http://timemapper.okfnlabs.org/timdavies/uk-open-data-timeline (accessed September 30, 2015).

Deloitte. "Market Assessment of Public Sector Information." *Department for Business, Innovation and Skills*. May 2013. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198905/bis-13-743-market-assessment-of-public-sector-information.pdf.

Department for Business, Innovation & Skills and Cabinet Office. *A consultation on data policy for a Public Data Corporation: government response*. 04 August 2011. https://www.gov.uk/government/consultations/data-policy-for-a-public-data-corporation.

Department for Communities and Local Government. "Median house price." *data.gov.uk*. 09 September 2010. https://data.gov.uk/dataset/median_house_price.

—. "Property sales based on Land Registry data." *data.gov.uk*. 09 February 2010. https://data.gov.uk/dataset/property_sales_based_on_land_registry_data.

Department for Education. "The national pupil database: User guide." 2013. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/261189/NPD_User_Guide.pdf.

digitalhealth.net. "DH blasted for 'back room deal' with Dr Foster." *digitalhealth.net*. 18 July 2007. http://www.digitalhealth.net/news/22819/.

Escher, Tobias. "Analysis of users and usage for UK Citizens' Online Deomcracy." May 2011. https://www.mysociety.org/files/2011/06/TheyWorkForYou_research_report-2011-Tobias-Escher1.pdf.

Bibliography

Everitt, Vernon. "Delivering better customer information through free open data ." *PTI*, 2014: 8-11.

Ford, S, R Howard, and J Oyebode. "Psychosocial aspects of coeliac disease, a cross-sectional survey of a UK population." *British Journal of Health Psychology* 17, no. 4 (November 2012): 743-57.

Fowler, Ryan. "Looking at the future of #proptech." Tech City News. 23 June 2015. http://techcitynews.com/2015/06/23/looking-at-the-future-of-proptech/ .

Freeguard, Gavin, Robyn Munro, and Emily Andrews. "Whitehall Monitor: Deep Impact?" *Institute for Government*. April 2015. http://www.instituteforgovernment.org.uk/publications/whitehall-monitor-deep-impact.

G8. "G8 Action Plan Principles to prevent the misuse of companies and legal arrangements." *gov.uk*. 2013. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207532/G8-Action-Plan-principles-to-prevent-the-misuse-of-companies-and-legal-arrangements.pdf.

Gray, Jonathan. "Towards a Genealogy of Open Data." *General Conference of the European Consortium for Political Research*. 2014. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2605828.

Gray, Jonathan, and Tim Davies. "Fighting Phantom Firms in the UK: From Opening Up Datasets to Reshaping Data Infrastructures?" *Working Paper for the Open Data Research Symposium, Ottawa, Canada, May 2015*. May 2015. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2610937.

Gray, Jonathan, Lilian Bounegru, and Lucy Chambers. *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. O'Reilly, 2012.

GS1. "Global Trade Item Number (GTIN)." GS1. http://www.gs1.org/gtin (accessed September 30, 2015).

Hadi, Zeid. "Welcome to data.parliament – The Blog ." *data.parliament.uk*. 2 May 2014. http://blog.data.parliament.uk/2014/05/hello-world-2/ .

Hall, Nicola, Gregory Rubin, and Anne Charnock. "Intentional and inadvertent non-adherence in adult coeliac disease: a cross-sectional survey." *Appetite* 68, no. 1 (2013): 56-62.

Hansen, Kirstine, Heather Joshi, and Shirley Dex. *Children of the 21st century: The first five years*. Policy Press, 2010.

Harding, Eleanor. "Schools 'stop poorer GCSE pupils taking hard subjects': Ploy to boost league rankings by denying access to exams including sciences ." *Daily Mail*. 10 February 2015. http://www.dailymail.co.uk/news/article-2948393/Schools-stop-poorer-GCSE-pupils-taking-hard-subjects-boost-league-rankings-denying-access-exams-including-sciences.html .

HM Treasury. *Public Data Group Business Case*. 13 02 2012. https://www.gov.uk/government/publications/open-data-public-data-group-business-case.

HMRC. "UK Property Transaction Statistics." *gov.uk*. 22 September 2015. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/461354/UK_Tables_Sep_2015__cir_.pdf.

Hogge, Becky. "Open Data Study: New Technologies." *Transparency and Accountability Initiative*. May 2011. http://transparencyinitiative.theideabureau.netdna-cdn.com/wp-content/uploads/2011/05/open_data_study_final1.pdf.

Hope, Chirstopher. "Everyone's postcodes to be privatised in Royal Mail flotation, despite objections from Sir Tim Berners-Lee." *The Daily Telegraph*. 19 April 2013. http://www.telegraph.co.uk/news/uknews/royal-mail/9994741/Everyones-postcodes-to-be-privatised-in-Royal-Mail-flotation-despite-objections-from-Sir-Tim-Berners-Lee.html.

Howdle, P. *Your Guide to Coeliac Disease*. Hodder Arnold, 2010.

Hurst, G. "The MPs who can't stop talking." *Times Online*. 27 February 2006 http://web.archive.org/web/20070216112003/http://www.timesonline.co.uk/tol/news/politics/article735429.ece

Jee, Charlotte. "Open data in the UK: a job half done." *Computerworld UK*. 12 August 2015. http://www.computerworlduk.com/data/open-data-in-uk-job-half-done-3622607/.

Kastrenakes, Jacob. "Trace the convoluted worldwide structures that your bank uses to dodge taxes ." *The Verge*. 12 July 2013. http://www.theverge.com/2013/7/12/4517178/opencorporates-visualizes-convoluted-global-banking-structures.

Keserű, Júlia. "Help us build an evidence base on the social impact of open data!" *Sunlight Foundation*. 23 January 2015. http://sunlightfoundation.com/blog/2015/01/23/help-us-build-an-evidence-base-on-the-social-impact-of-open-data/.

Kiln. "OpenCorporates Maps." *Kiln*. http://www.kiln.it/projects/opencorporates (accessed September 30, 2015).

KPMG. "HMLR Feasibility Study." 2011.

Land Registry. "INSPIRE Index Polygons spatial data > conditions of use." *gov.uk*. 14 August 2014. https://www.gov.uk/guidance/inspire-index-polygons-spatial-data#conditions-of-use.

—. *Internet Archive copy of webpage: Welcome to Land Registry's 10-year strategic plan*. April 2006. http://web.archive.org/web/20070602044143/http://www.landregistry.gov.uk/strategy/10yearplan/.

—. "Privacy Impact Assessment Report: Making price paid data available through publication in a machine readable and reusable format." *Land Registry*. 2012. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/343616/ppd_pia.pdf.

—. "Privacy Impact Assessment Review: Price paid data, transaction data and historical price paid data." Land Registry. 2013. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/343604/PIA_Report_6_13.pdf.

Lateral Economics. "Open for Business: How open data can help achieve the G20 growth target." *Omidyar Network*. June 2014. https://www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FNL.pdf.

legislation.gov.uk. "Copyright, Designs and Patents Act 1988." *legislation.gov.uk*. http://www.legislation.gov.uk/ukpga/1988/48/section/30 (accessed September 30, 2015).

Mapping London. *House Prices: A Borough Cartogram*. 19 March 2015. http://mappinglondon.co.uk/2015/house-prices-a-borough-cartogram/.

Maybin, Jo. "Experience-based policymaking." *Institute for Government*. 16 April 2013. http://www.instituteforgovernment.org.uk/blog/5671/experience-based-policymaking/.

McGee, Rosemary, and John Gaventa. "Review of Impact and Effectiveness of Transparency and Accountability Initiatives: Synthesis Report." *Institute of Development Studies*. October 2010. https://www.ids.ac.uk/files/dmfile/IETASynthesisReportMcGeeGaventaFinal28Oct2010.pdf.

McKenzie, Jessica. "Open Corporate Data For Everyone, Everywhere." *TechPresident*. 16 October 2013. http://techpresident.com/news/wegov/24436/open-corporate-data-everyone-everywhere.

McKinsey Global Institute. "Open data: Unlocking innovation and performance with liquid information." *McKinsey & Company*. October 2013. http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information.

Meyer, D. "Want to follow the money? OpenCorporates uses open data to expose corporate structures." *GigaOm*. 12 July 2013. https://gigaom.com/2013/07/12/want-to-follow-the-money-opencorporates-uses-open-data-to-expose-corporate-structures/.

Mills & Reeve. "Mills & Reeve advise on multi-million dollar sale." *Mills & Reeve*. 09 February 2015. http://www.mills-reeve.com/mills--reeve-advise-on-multi-million-dollar-sale-02-09-2015/ .

Mirani, Leo. "London's bike-share program unwittingly revealed its cyclists' movements for the world to see." *Quartz*. 16 April 2014. http://qz.com/199209/londons-bike-share-program-unwittingly-revealed-its-cyclists-movements-for-the-world-to-see/.

Moon, James C, Andrew S Flett, Brian B Godman, Anthony M Grosso, and Anthony S Wierzbicki. "Getting better value from the NHS budget." *British Medical Journal*. 17 December 2010. http://www.bmj.com/content/341/bmj.c6449.

Morozov, Evgeny. *To Save Everything, Click Here: Technology, solutionism and the urge to fix problems that don't exist*. London: Allen Lane, 2013.

mySociety. "Homepage." *mySociety*. https://www.mysociety.org (accessed September 30, 2015).

—. "Twitter conversation between mySociety, Matthew Somerville and TheyWorkForYou user Tom Ward." *Twitter*. 17 December 2013. https://twitter.com/mysociety/status/412957434706403328.

Nixon, Myfanwy. "Open Data day resources." *mySociety*. 20 February 2014. https://www.mysociety.org/2014/02/20/open-data-day-resources/ .

Nottingham Post. " Nottinghamshire pupils 'miss out on GCSE subjects'." *Nottingham Post*. 12 February 2015. http://www.nottinghampost.com/Pupils-miss-GCSE-subjects/story-26013138-detail/story.html .

ODI. "Case study: OpenCorporates." ODI. https://theodi.org/case-studies/opencorporates-case-study (accessed September 30, 2015).

—. "How to make a business case for open data." *ODI*. https://theodi.org/guides/how-make-business-case-open-data (accessed September 30, 2015).

OECD. " Behind the Corporate Veil: Using Corporate Entities for Illicit Purposes." 2001. http://www.oecd.org/daf/ca/behindthecorporateveilusingcorporateentitiesforillicitpurposes.htm.

—. "Focus on house prices." *OECD*. http://www.oecd.org/eco/outlook/focusonhouseprices.htm (accessed September 30, 2015).

Open Data Institute. "Open Data means business: UK innovation across sectors and regions." 2015. http://theodi.org/open-data-means-business-uk-innovation-sectors-regions.

Open Knowledge Foundation. "Global Open Data Index." http://index.okfn.org/.

Open Public Services Network. "Empowering Parents, Improving Accountability." September 2013. https://www.thersa.org/discover/publications-and-articles/reports/empowering-parents-improving-accountability/.

—. "Lack of options: how a pupil's academic choices are affected by where they live." February 2015. https://www.thersa.org/discover/publications-and-articles/reports/lack-of-options-how-a-pupils-academic-choices-are-affected-by-where-they-live/.

OpenCorporates. "Introducing OpenCorporates: The open database of the corporate world." *OpenCorporates*. 20 December 2010. http://blog.opencorporates.com/2010/12/20/press-release-opencorporates-launches/.

Pollock, Rufus. "Open Data, Technology and Government 2.0 – What Should We, And Should We Not Expect." *Open Knowledge*. 13 September 2012. Open Data, Technology and Government 2.0 – What Should We, And Should We Not Expect.

Prentice, Gordan. "HC Deb, 23 October 2008, c151WH." *TheyWorkForYou.com*. 23 October 2008. http://www.theyworkforyou.com/whall/?id=2008-10-23b.139.0#g151.0.

Prime Minister's Office. "PM's podcast on transparency." *gov.uk*. 29 May 2010. https://www.gov.uk/government/news/pms-podcast-on-transparency.

Private Eye. *In the back: Selling England by the offshore pound*. 2015. http://www.private-eye.co.uk/registry.

Public Data Group. *Public Data Group*. https://www.gov.uk/government/groups/public-data-group.

Quinn, James. "George Osborne to privatise £23bn of taxpayer assets." *The Telegraph*. 20 May 2015. http://www.telegraph.co.uk/finance/newsbysector/banksandfinance/11618980/George-Osborne-to-privatise-23bn-of-taxpayer-assets.html.

Ramsay, Frank. *Telling stories with open data and maps*. 03 July 2014. http://blog.landregistry.gov.uk/using-open-data-map-future/.

Reed, Simon. "Open Data and Bus Apps." Transport for London Customer Group , 2015.

Robinson, David G, Harlan Yu, William P. Zeller, and Edward W. Felten. "Government Data and the Invisible Hand." *Yale Journal of Law and Technology*. 2009. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1138083.

Royal Mail. "Postcode Address File." *Royal Mail*. http://www.royalmail.com/business/services/marketing/data-optimisation/paf (accessed September 30, 2015).

Sampson, Lindsay. "Middlesbrough a 'subject desert': Pupils unlikely to take exams that could be vital to job prospects." *Gazette Live*. 11 February 2015. http://www.gazettelive.co.uk/news/teesside-news/middlesbrough-subject-desert-pupils-not-8626289 .

Shopper Goggles. " Terms & Conditions and Legal Stuff ." *Shopper Goggles*. http://shoppergoggles.com/terms-conditions-legal-stuff/ (accessed September 30, 2015).

Siddle, James. *I Know Where You Were Last Summer: London's public bike data is telling everyone where you've been*. 10 April 2014. http://vartree.blogspot.co.uk/2014/04/i-know-where-you-were-last-summer.html.

Skidmore, Chris. "HC Deb, 24 March 2015, c1324." *TheyWorkForYou.com*. 24 March 2015. http://www.theyworkforyou.com/debates/?id=2015-03-24a.1324.2 .

Solon, Olivia. "Open data platform reveals complex corporate structures of banks." *Wired.co.uk*. 11 July 2013. http://www.wired.co.uk/news/archive/2013-07/11/opencorporates.

Syal, Rajeev. "Freedom of Information Act review 'may curb access to government papers'." *The Guardian*. 17 July 2015. http://www.theguardian.com/politics/2015/jul/17/ministers-to-review-freedom-of-information-act.

—. "Land Registry privatisation plans abandoned by ministers." *The Guardian*, 14 July 2014.

Taggart, Chris. "Announcing open corporate network data: not just good, but better." *OpenCorporates*. 11 July 2013. http://blog.opencorporates.com/2013/07/11/open-corporate-network-data-not-just-good-but-better/.

TalkTalk Digital Heroes. "2014 Winners." *TalkTalk*. http://digitalheroes.talktalk.co.uk/past-winners/2014-winners/ (accessed September 30, 2015).

The Economist. "Unleashed: Why backbench MPs are becoming so much harder to control." *The Economist*. 27 October 2012. http://www.economist.com/news/britain/21565220-why-backbench-mps-are-becoming-so-much-harder-control-unleashed .

TheyWorkForYou.com. "About TheyWorkForYou." *TheyWorkForYou.com*. http://www.theyworkforyou.com/about/ (accessed September 30, 2015).

—. "Frequently Asked Questions." *TheyWorkForYou.com*. http://www.theyworkforyou.com/help/ (accessed September 30, 2015).

Thwaites, Emma. *Prescription Savings Worth Millions Identified by ODI* incubated company. 06 December 2012. https://theodi.org/news/prescription-savings-worth-millions-identified-odi-incubated-company.

Times Educational Supplement. "Pupils in poor areas denied chance to study science and foreign languages, says study." *Times Educational Supplement*. 11 February 2015. https://www.tes.co.uk/news/school-news/breaking-news/pupils-poor-areas-denied-chance-study-science-and-foreign-languages .

Timita, Manuel. *London house prices: evolution over 13 years*. 27 February 2014. http://illustreets.co.uk/blog/maps-and-apps/london-house-prices-evolution-over-13-years/.

Topham, Gwyn. "Open data and driverless buses: how London transport heads to the future." *The Guardian*. 13 August 2014. http://www.theguardian.com/uk-news/2014/aug/13/transport-london-tube-bus-oyster-data.

Transparency International. "Six things to know: new G20 Beneficial Ownership Principles." *Transparency International*. 17 November 2014. http://www.transparency.org/news/feature/six_things_to_know_new_g20_beneficial_ownership_principles.

Transport for London. *How we are funded*. https://TfL.gov.uk/corporate/about-TfL/how-we-work/how-we-are-funded? (accessed September 30, 2015).

—. *Revolutionary new Santander Cycles App launched*. 11 May 2015. https://TfL.gov.uk/info-for/media/press-releases/2015/may/revolutionary-new-santander-cycles-app-launched.

—. *Transport Data Service terms and conditions*. https://TfL.gov.uk/corporate/terms-and-conditions/transport-data-service (accessed September 30, 2015).

—. *What we do*. https://TfL.gov.uk/corporate/about-TfL/what-we-do (accessed September 30, 2015).

Triggle, N. "Care.data: How did it go so wrong?" *BBC News*. 19 February 2014. http://www.bbc.co.uk/news/health-26259101.

van de Does de Willebois, E, E.M Halter, R.A. Harrison, J.W. Park, and J.C Sharman. "The Puppet Masters: How the Corrupt Use Legal Structures to Hide Stolen Assets and What to do about it." *World Bank* . 2011. https://star.worldbank.org/star/sites/star/files/puppetmastersv1.pdf.

We Own It. *Don't sell of our top trumps*. http://weownit.org.uk/take-action/dont-sell-our-top-trumps (accessed September 30, 2015).

Wheeler, Brian. "Government online data ignored by 'armchair auditors'." *BBC News*. 9 November 2012. http://www.bbc.co.uk/news/uk-politics-20221398.

Wikipedia. "Coeliac UK." *Wikipedia*. https://en.wikipedia.org/wiki/Coeliac_UK (accessed September 30, 2015).

Williams, Oscar. "Open data could save the NHS hundreds of millions, says top UK scientist." *The Guardian*. 29 May 2015. http://www.theguardian.com/media-network/2015/may/29/open-data-nhs-healthcare-nigel-shadbolt.

World Bank. "Education Sector uses of open data." *World Bank*. https://finances.worldbank.org/Reference/Education-Sector-uses-of-open-data/grcz-ymf2 (accessed September 30, 2015).

Worldwide Web Foundation. "Open Data Barometer: Second Edition." January 2015. http://www.opendatabarometer.org/assets/downloads/Open%20Data%20Barometer%20-%20Global%20Report%20-%202nd%20Edition%20-%20PRINT.pdf.

Wu, Tim. *The Master Switch: The Rise and Fall of Information Empires*. New York: Alfred A Knopf, 2010.

Yu, Harlan, and David G. Robinson. "The New Ambiguity of Open Government." *UCLA Law Review*, 2012.

Zuiderveen Borgesius, F, J Gray, and M van Eechoud. "Open Data, Privacy, and Fair Information Principles: Towards a balancing framework." Forthcoming.

Methodology

This report, consisting of six case studies about the economic, social, political, or cultural impact of the UK government's open data policy, was commissioned by Omidyar Network in March 2015. As part of the terms of reference for this work, the researcher was asked to limit case study selection to cases that demonstrated past and present impact, rather than those that projected future value, in order to answer the question, "has open data already had measurable impact"? The researcher was asked to focus on cases that demonstrated "tangible impact at scale".

The researcher began by conducting a literature review of open data impact using a cascading search methodology, and interviewing open data researchers and key players in the field. Questions used to guide this research are provided below. Case study selection was driven by the researcher's desire to reflect different lessons about open data and the study of its impact, and was approved by Omidyar Network in May 2015.

The researcher then conducted a literature review of material relevant to each case, again using a cascading search approach. For each case, the researcher attempted to contact for interview data publishers, data re-users, and the users of the products that were the outcome of this data re-use. Questions used to guide these interviews are also provided below, and a list of interviewees is given in Appendix 2.

The calls to action at the end of each case study are driven by the observations made and evidence collected in the course of the author's research, and may not reflect the policies of Omidyar Network.

Interview framework 1: General interview

1. **Can you name your top three examples of where you think open data has had a positive impact?**
2. **Are there any particular sectors or spheres where you think open data is more likely to have a positive impact (use Rufus' graph to illustrate this point)?**
3. **Is open data more likely to have positive economic impacts than it is political, social, or environmental? Or are economic impacts simply easier to quantify? Or both?**
4. **What approaches have you seen to thinking and talking about open data impact that you think are valuable?**
5. **How conscious do we have to be of the potentially negative impacts and unforeseen consequences of releasing open data sets or of open data policies in general?**
6. **Is it too soon to get a good idea about impact? Has open data been given enough of a chance to prove its worth?**
7. **Where around the world (and not just US/UK) do we see open datasets that pre-date the recent OGP commitments, and where we might therefore expect to see measurable impact?**

8. **Do you subscribe to the theory that open data growth will be subject to network effects?**
9. **What are your favourite examples of measuring social/political/environmental impacts from other fields? Can they be applied to open data?**
10. **Given one of the cited advantages of open data is that it allows many possible benefits, including unanticipated benefits, and given that research designed to evidence impact generally needs to be specific to a target or goal, is it harder to evaluate the impact of open data policies than it might be for other types of policies?**

Interview framework 2: Case study

1. **What was the data set that you used to create the impact?**
2. **Does it conform to the open definition?**
3. **Is it regularly updated?**
4. **Is it subject to adequate privacy protections?**
5. **When was this dataset released?**
6. **Who, if anyone, was calling for the data to be released before it was open?**
7. **Had the data been obtained through other means (e.g., existing access to paper-based records/RTI laws/scraping activity/leaks) before it was released as open data?**
8. **Were there any stakeholders against its release as open data? What concerns did they have and how were those concerns answered or met?**
9. **Once the data had been released as open data, what happened next? Who accessed the data? Who transformed it? Who were they and what were their motivations? E.g.,:**
    a. **Existing or new business wanting to develop new products and services?**
    b. **Civil servants seeking to improve their work?**
    c. **Civic hackers messing around with data/acting out of civic duty/demonstrating why more data should be released?**
    d. **Policy advocates seeking specific policy change?**
10. **Who was the target audience for these infomediaries? E.g.,:**
    a. **Citizens seeking to hold their government to account?**
    b. **Consumers seeking to make better consumption choices?**
    c. **Civil servants?**
    d. **Journalists?**
    e. **Funders/Venture capitalists?**
11. **What happened next? What was the outcome of this activity and what was its impact on:**
    a. **The original stakeholders in the data?**
    b. **The people who accessed and transformed it?**
    c. **The target audience?**
    d. **General social/economic/political/environmental well-being?**
12. **What is the status of the dataset in 2015? Is it still released openly, and does it remain up-to-date?**
13. **Where do you see this story headed in the future?**

List of interviewees

- Charlotte Alldritt, Director, Open Public Services Network
- Beau Archer, Technical Pre-Sales Consultant, Brandbank
- Claudia Arney, Chair, Public Data Group 2012-2015
- Adrian Black, Founder and Managing Director, YOUhome
- Paul Clarke, Independent contractor
- Emer Coleman, Director (Business Development), Transport **API**; Director of Digital Projects, Greater London Authority 2009-2011
- Tim Davies, Worldwide Web Foundation
- Vernon Everitt, Managing Director in charge of Customer Experience, Marketing and Communications, TfL
- Jonathan Gray, Director of Policy and Research, Open Knowledge
- Júlia Keserű, International Policy Manager, Sunlight Foundation
- David McNair, Director (Transparency and Accountability) ONE Campaign; Deputy Director of Policy and Research, Save the Children 2012-2014
- Kathryn Miller, Food Policy Lead, Coaliac UK
- Lynne Nicholson, Head of Data Products and Services, HM Land Registry
- Robert Palmer, Head of Banks and Corruption campaign, Global Witness
- Rufus Pollock, Economist, co-founder, Open Knowledge
- Henry Pryor, Independent Residential Property Expert
- Katelyn Rogers, Project Manager, Open Knowledge
- Rebecca Rumbul, Head of research, mySociety
- Matthew Somerville, Lead Developer, TheyWorkForYou, mySociety
- Tom Steinberg Director of mySociety 2004-2015
- Andrew Stott, Former UK Government Director of Transparency & Digital Engagement
- Vasanth Subramanian, Co-founder and CTO, Splittable
- Chris Taggart, CEO, OpenCorporates
- Roger Taylor, Chair, Open Public Services Network
- Peter Thum-Bonanno, Co-founder and CTO, GetAgent
- Phil Young, Head of Online, TfL

About the author

Becky Hogge is a journalist, freelance writer, and researcher. Her writing on information politics, human rights, and technology appeared regularly in the UK political magazine the *New Statesman* in the first decade of this century, and she has also been published in *Index on Censorship*, *The Guardian*, and *Prospect*. Her first book, Barefoot into Cyberspace: Adventures in search of techno-Utopia, was flash-published in the Summer of 2011: The book was well-received and subsequently highlighted by *The Guardian* as their non-fiction choice. She published her second book, *A Guide to the Internet for Human Rights Defenders*, in December 2014.

Beginning in 2007, Becky spent two years as Executive Director of the Open Rights Group, a grassroots campaigning organisation that exists to protect the rights of citizens and consumers in the digital age. Becky has produced research papers on internet governance, online campaigning, and online markets for various international NGOs, has been on several national awards panels, and has been invited by the UK and other European governments to consult on information policy. Her opinions on internet governance and internet censorship have been broadcast around the world.

In recent years, Becky's work has focussed on helping grant-giving organisations with communications and strategy work. In 2010, at the request of the Transparency and Accountability Initiative, a funder collaborative including Omidyar Network and Open Society Foundations, she wrote an influential paper, *Open Data Study*, which charted the rise of open data reforms in the UK and US.

Becky was a board member of UK open data pioneers, the Open Knowledge Foundation for five years until 2013, and she sits on the Advisory Council of that organisation, as well as for the Foundation for Information Policy Research and the Open Rights Group. She is a member of the UK telecommunications regulator OfCom's Advisory Committee for England.

Image credits

Figure 1: Timeline of TfL data releases (page 11) Source: (Everitt 2014)

Figure 2: Visualisation of TfL Tube data (page 13) reproduced with kind permission of Will Gallia. Original video available to view at https://vimeo.com/121462010.

Figure 3: Cartogram of median house prices in London (page 20) Source: Uncredited image published under OGL at http://data.london.gov.uk/dataset/average-house-prices-borough

Figure 4: Screenshot of video timeline of rise in London house prices (page 20) reproduced with kind permission of Manuel Timita and Illustreets. Original video available here: http://illustreets.co.uk/wp-content/uploads/miniapps/london-house-prices/#11/51.4921/-0.0962

Figure 5: Total UK Residential Property Transactions (page 23) Taken from (HMRC 2015)

Figure 6: Detail from OpenCorporates and Kiln's visualisation of complex corporate structures (page 27) Copyright OpenCorporates. Reproduced with kind permission of OpenCorporates.

Figure 7: Access to triple sciences across English schools, taken from Lack of Options report. (page 31) Copyright OPSN. Reproduced with kind permission of OPSN. Original available at https://www.thersa.org/discover/publications-and-articles/reports/lack-of-options-how-a-pupils-academic-choices-are-affected-by-where-they-live/Download

Figure 8: Number of Science GCSEs per pupil in LEAs coloured by deprivation quintile, taken from Lack of Options report (page 32). Copyright OPSN. Reproduced with kind permission of OPSN. Original available at https://www.thersa.org/discover/publications-and-articles/reports/lack-of-options-how-a-pupils-academic-choices-are-affected-by-where-they-live/Download

Figure 9: Screenshot of front page of TheyWorkForYou.com (page 36). Copyright mySociety. Reproduced with kind permission of mySociety.

Figure 10: Screenshot from TheyWorkForYou.com (page 36) Copyright mySociety. Reproduced with kind permission of mySociety.

Figure 11: TheyWorkForYou.com - visitor statistics (page 38). Copyright Tobias Escher. Source: (Escher 2011) Reproduced under the terms of CC-BY-NC 2.0 licence.

Acknowledgements

The author wishes to thank Laura Bacon at Omidyar Network for the tremendous support and guidance she has provided throughout this project. Jonathan Gray, Tim Davies, and Tom Steinberg all shared their time and expertise in ways that were invaluable to the author in helping to frame this report, and for this I am deeply grateful. Thanks also to Rebecca Rumbul at mySociety, to Júlia Keserű at Sunlight Foundation and to Rufus Pollock at Open Knowledge for their insight into methodological approaches to studying impact.

To all the interviewees who took time out of their busy day jobs to contribute to the report, heartfelt thanks. Thanks also to Martin Tisné at Omidyar Network, Tom Heath at the Open Data Institute, Janet Haven at Open Society Foundations, Fabrizio Scrollini at Data Uruguay, Stefaan Verhulst at GovLab, and Naser Faruqui at IDRC for their time, ideas, and input. Many thanks to the organisers of the International Open Data Conference in Ottawa, Canada in May 2015 for giving the author an early platform to discuss this work.

Thank you to James Casbon for his help with the data work, particularly in establishing the scope of the **Price Paid Dataset**, and for his personal support throughout this project. Continuing on this personal note, the author would like to thank Ginnie Hogge and Sue Casbon for their faith, encouragement, and support.

Any errors remain the responsibility of the author.