

European Public Sector Information Platform

Topic Report No. 26

(Part 2 of Topic Report 8)

Opening up government data:
making the case

Author: Dr. Pauline Pollard

Published: 26th February 2011

Keywords

Public sector information; PSI portals; open data catalogues; democratic engagement; knowledge economy.

About the Author

Dr. Pauline Pollard is a European Public Sector Information Platform Analyst. Her PhD ‘Coordinating the Sharing of Spatial Data in the UK’ follows the history of the development of the UK Address datasets and provides an institutional analysis for the reasons the dataset is locked out of open access by barriers of cost recovery. She previously worked as a Senior Lecturer for the University of the West of England in the field of Information Systems. Prior to this she worked as a business analyst in local government. She has also contributed to the specialist ICT group within the European Group of Public Administration.

Abstract

This report lays out the benefits of opening up Public Sector Information for re-use using case stories drawn mostly from member states within the EU. It aims to raise awareness of the benefits of open data and support project champions in the member states engaged in raising awareness of the value of cross-sector PSI portals, in particular data catalogues. It discusses risks perceived by PSI custodians when opening up data and provides some brief guidelines on developing data catalogue sites.

Copyright information

© 2010 European PSI Platform - This document and all material therein has been compiled with great care; however, the author, editor and/or publisher and/or any party within the European PSI Platform or its predecessor projects the ePSIplus Network project or ePSINet consortium cannot be held liable in any way for the consequences of using the content of this document and/or any material referenced therein. This report has been published under the auspices of the European Public Sector Information Platform.



The report is released under the Creative Commons Attribution 3.0 License and may be used providing acknowledgement is made to the European Public Sector Information (PSI). The European Public Sector Information (PSI) Platform is funded under the European Commission eContentplus programme.

Content

1. Introduction	4
2. PSI portals and open data catalogues	5
3. Democratic engagement	7
a) Democracy and citizen engagement	7
b) Transparency and the monitoring of government activities	9
c) Better information, data journalism and solving policy problems.....	11
4. Better services.....	13
a) Reduced transactional costs and improved knowledge	13
b) Dialogue in design of services	15
5. The knowledge economy.....	16
a) Developing an information industry	16
b) Supporting business, fostering tourism and cultural activities	19
6. Challenges of opening up data.....	20
Resources	Error! Bookmark not defined.
Data quality	20
Institutional change	Error! Bookmark not defined.
7. Developing an open data catalogue	21
8. Conclusion.....	23

1. Introduction

Information produced by the public sector is the single largest information source in Europe.¹ Its effective use is increasingly perceived not only as having the potential to stimulate economic growth but as at the heart of democratic engagement and as offering the potential for innovative changes in the delivery of public services. Web technologies present the means of re-using this wealth of public sector information (PSI) in new and innovative ways to meet these potentials. As noted in part 1 of this report, a network of open data catalogues is emerging at national, regional and local levels as institutions of governance increasingly make their data available online in standard formats and under licenses intended to permit the free re-use of data.^{2,3}

The emergence of this network is neither an accident nor a purely technological endeavour. It is part of a broader social movement towards open government that calls for transparency and dialogue between state and citizen, within which open data is a pre-condition but not the end-point, and a movement towards open knowledge which recognises the power of digital technologies to rapidly transform and innovate. However, to embrace open data requires a mind shift within the institutions of governance in order to unlock information silos where permission has to be obtained to reuse data.⁴

The underlying difference between much current practice and open data concepts is the difference identified in the groundbreaking research by Shoshana Zuboff between using technology to automate and using technology to ‘informate’, that is to illuminate activities and enable innovation. She argued that technology can be harnessed to either paradigm depending on the institutional values placed upon information technology, but that to informate new organisational approaches are needed.⁵ These new organisational approaches, identified by Taylor and Williams⁶ in the early 1990s, require a shift towards: outward accountability to the citizen, the removal of barriers to enable information to flow seamlessly across organisational boundaries, an integrated service culture and the empowering of the user in (co-)production of applications. Underpinning these principles is the recognition of the need to institutionalise these new organisational approaches into governance practices.

¹ [http://ec.europa.eu/information_society/policy/psi/docs/pdfs/swd_070509/re-usepsi_sec\(2009\).pdf](http://ec.europa.eu/information_society/policy/psi/docs/pdfs/swd_070509/re-usepsi_sec(2009).pdf)

² Rob Davies (2010) *ePSIplatform Topic Report No. 8: PSI Portals: Overview of Progress (Part 1)* available at http://www.epsiplus.net/topic_reports

³ ePSIplatform provides a list of data catalogue sites available at http://www.epsiplus.net/psi_data_catalogues. See also Alexander Schellong and Ekaterina Stepanets's (2011) *Unchartered Waters: the State of Open Data in Europe* This draws on ePSIplatform material to discuss catalogue developments in several EU member states.

⁴ <http://opengovernment.labs.oreilly.com/ch01.html>

⁵ Zuboff (1998) *In the age of the smart machine: The future of work and power* Heinemann

⁶ Taylor J. And Williams H. (1990) *Themes and issues in an information polity* Journal of Information Technology 5, pp. 151-60

There is a challenge, more so in a period of austerity as governments retrench, that the movement could stall or be thrown into reverse. Evidence of this possibility can be seen in countries which led in introducing data catalogues. Thus in September 2010 Ellen Miller⁷ identified a failure in the United States to build on its data catalogue with new and interesting datasets, and she expressed a fear that: ‘the drive for transparency appears stalled’.⁸ And early this year the UK government announced the formation of a ‘public data corporation’ which would provide ‘stability and certainty for businesses’, ‘real value for the taxpayer’ and ‘opportunities for private investment in the corporation.’⁹ This in combination with a shift in responsibility from the Cabinet Office to the Department of Business, Innovation and Skills is being seen as an attempt ‘to co-opt the open data agenda, as a way of shutting it down, smothering it’.¹⁰

Advocates of open data recognise the need to respond to this challenge.¹¹ This report therefore presents good case stories (sections 3-5) in order to defend current gains in the opening up of data and to maintain the impetus towards open data as the best means of securing the economic, social and democratic benefits of public sector information. In addition to good case stories the report aims to address some common concerns that PSI holders have about opening up their data (section 5). Finally it proposes some simple guidelines for developing open data catalogues (section 7). Before making the case for opening up data, the next section briefly considers what is meant by a PSI portal and how this concept might relate to the concept of a data catalogue.

2. PSI portals and open data catalogues

The European Union (EU) sought to ensure the economic benefits of public sector information re-use through the PSI Directive of 2003 which sets minimum rules for PSI re-use and encourages member states to develop portal sites to facilitate access.¹² This raises the question of what a PSI portal site should look like.

Chris Corbin considers that as good practice a PSI portal site should provide all the information relevant to that PSI re-use in the one place so that it acts as a ‘one stop shop for information’. A PSI portal may be international, national or local and may be run by a public sector, private sector or civil society actor. It may be cross-sector or be developed for a particular sector or function and developments be taking place at the EU level (e.g.

⁷ The Executive Director of the Sunlight Foundation

⁸ <http://sunlightfoundation.com/blog/2010/09/07/gov2-0-presentation-an-open-government-scorecard/>

⁹ <http://www.cabinetoffice.gov.uk/news/public-data-corporation-free-public-data-and-drive-innovation>

¹⁰ <http://countculture.wordpress.com/>

¹¹ See <http://countculture.wordpress.com/> and <http://sunlightfoundation.com/blog/2010/09/07/gov2-0-presentation-an-open-government-scorecard/>

¹² EU PSI Directive 2003/98/EC (Not all categories of PSI are included within the scope of the Directive)

environmental information,¹³ land registration, business registers and statistical information). Ideally a PSI portal brings the content to users but it may only act as a gateway allowing the user to discover where the content is and to navigate to that site.¹⁴

The PSI Directive recommends open data policies if broader PSI re-use is to be achieved and, as the Directive was enacted in member states, a citizen-led movement urging the opening up of data also emerged leading to the development of data catalogues by public sector bodies, the private sector and citizens. These catalogues not only contain public data but provide collaborative web tools. This ‘bottom up’ movement argues that for data to be considered ‘open’ it should be available according to a set of principles.¹⁵ Thus, whilst the Directive reflects the minimum set of rules required for an internal market the ‘open’ data principles reflect how information might flow in an information society.

As a result of their differing history, these ‘top-down’ and ‘bottom-up’ initiatives use a different terminology in what has been described as two waves of PSI reuse.¹⁶ Whilst this reflects different priorities, there is co-operation within the EU between PSI reusers who made the case for the PSI Directive and the open data network. Open data catalogues can act as PSI portals but reflect broader aspirations than the PSI portals proposed by the Directive. Open data catalogues are the focus of this report because of their potential to be the future gateway for PSI users and re-users.

The ePSIplatform portal provides a rich repository of information related to PSI re-use and a decision was made to utilise this resource to provide evidence of benefits for this report. Case examples of the benefits of opening up data are categorised in this report as follows:¹⁷

¹³ Michael Fanning (2010) *ePSIplatform Topic Report No11 Recognising the road to data.gov.de An assessment of the European and national regulatory framework impacting PSI re-use in Germany* available http://www.epsiplus.net/topic_reports describes Germany’s environmental information platform PortalU

¹⁴ Chris Corbin 25th September 2009 *The European Public Sector Information Platform perspective on PSI portals* available from http://ec.europa.eu/information_society/policy/psi/news_archive/index_en.htm

¹⁵ Government data shall be considered open if it is made public in a way that complies with certain principles:

1. **Complete** All public data is made available. Public data is data that is not subject to valid privacy, security or privilege limitations.
2. **Primary** Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.
3. **Timely** Data is made available as quickly as necessary to preserve the value of the data.
4. **Accessible** Data is available to the widest range of users for the widest range of purposes.
5. **Machine processable** Data is reasonably structured to allow automated processing.
6. **Non-discriminatory** Data is available to anyone, with no requirement of registration.
7. **Non-proprietary** Data is available in a format over which no entity has exclusive control.
8. **License-free** Data is not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed. Compliance must be reviewable.

Available at http://resource.org/8_principles.html

¹⁶ Antti Poikola (2010) *ePSIplatform Topic Report No. 12: Open data in Finland - bottom up and middle out, but not yet from top down* available at http://www.epsiplatform.com/topic_reports/

¹⁷ Much of the categorisation and content of the platform can be attributed to the dedicated work of Chris Corbin and the ePSIplatform team. See <http://www.epsiplus.net/>

- Democratic engagement (section 3);
- Better services (section 4);
- The knowledge economy (section 5).

3. Democratic engagement

Three key benefits of opening up public data are discussed in this section. Firstly, and perhaps the key benefit, is the potential for better governance achieved through citizens being more informed about the workings of government and better able to engage in the political process. Secondly, and closely related to the first benefit, is the potential to provide greater transparency and therefore accountability of government and parliamentary representatives to citizens. Finally there is the potential for greater engagement of citizens, journalists, and others in policy problems. This section provides case examples of these benefits.

a) Democracy and citizen engagement

Access to public data opens up easier means of access to political representatives, online monitoring of their speeches and voting patterns, and enables citizens to comment on what is going on in parliament. Once data is available, the public is better able to educate itself either directly or through other media.

To re-use PSI related to democratic institutions, volunteers ('civic hackers') may collect ('scrape') data from official websites to form a database. Thus UK's [TheyWorkForYou](#) (see Fig: 1) now run by mySociety was launched in 2004 by volunteers to enable people to monitor their representatives and comment on what goes on. In collecting data from online official sources volunteers infringed crown copyright but licensing agreements were later agreed with government. The site rapidly became popular and now attracts up to half a million users each month.¹⁸

¹⁸ http://www.soros.org/initiatives/information/focus/communication/articles_publications/publications/open-data-study-20100519/open-data-study-100519.pdf

Opening up government data: making the case

The screenshot shows the homepage of TheyWorkForYou.com. At the top left is the logo with the tagline 'Keeping tabs on the UK's parliaments & assemblies'. To the right is a search bar with a 'Go' button and a link to 'More options'. Below the search bar is a navigation menu with 'Overview', 'MPs', 'Lords', 'Debates', 'Written Answers and Statements', and 'Bill Committees'. The main content area is divided into two columns. The left column, titled 'Your representative', includes a search for new MPs, MSPs, or MLAs by UK postcode, and a link to read debates. The right column, titled 'Search, create an alert or RSS feed', has a search bar and a list of popular searches: Linda Gilroy, Andrew Miller. Below this are three sections: 'The most recent Commons debates' (listing Mr Speaker in the Chair, Bill Presented, Sex and Relationships Education Bill, Apprenticeships and Skills (Public Procurement Contracts) Bill), 'The most recent Westminster Hall debates' (listing Mr Charles Walker in the Chair – backbench business – Onshore Wind Energy, House of Lords, EU: Agriculture and Fisheries Council), and 'The most recent Written Answers' (listing Transport, Education, Defence, Energy and Climate Change, House of Commons Commission).

Figure 1 TheyWorkForYou

Although intended to provide citizen access to politicians, sites like TheyWorkForYou also enable politicians to have low cost access to constituents.¹⁹ It is therefore not surprising that it was activities such as these which are credited with leading to the emergence in 2009 to 2010 of the open data catalogues data.gov in the USA and data.gov.uk in the UK.²⁰

Volunteers in other EU member states have developed similar tools. In Finland, [kansanmuisti](#) aims to form a 'collective memory' of representatives' speeches, voting behaviour, election funding and initiatives both to monitor if representatives keep their election promises and to enable users to find representatives sympathetic to a particular policy.²¹ In Italy, [Openpolis](#) supports a community of users to find and 'adopt' their local representatives, follow their activities and upload updates on their chosen representative; a sub-site [Open Parlamento](#) provides a real-time monitor of parliament activities, including a [visualisation of voting](#).²²

Tools have also been created to make it easy for citizens to contact their representatives (e.g. the UK's [WriteToThem](#)) and petition their representatives (e.g. the UK Prime Ministers' [Number10.gov](#) (see Fig: 2)) both developed by mySociety. [Number10.gov](#) now has over 5 million unique email addresses (representing around 10% of the population). Private data is held on mySociety's servers, the number of messages No10 can send to petitioners is limited and information is provided on petitions rejected. Written in open source code, the website can be re-used by local councils.²³

¹⁹ Currently data on MPs is not available on data.gov.uk nor are links provided to [TheyWorkForYou](#).

²⁰ http://www.soros.org/initiatives/information/focus/communication/articles_publications/publications/open-data-study-20100519/open-data-study-100519.pdf

²¹ <http://www.kansanmuisti.fi/>

²² <http://owni.fr/2010/10/05/a-journey-through-tech-for-transparency-projects/>

²³ <http://www.mysociety.org/projects/no10-petitions-website/>

The screenshot shows the 'Number 10.gov.uk BETA' website. The header includes the site name and 'The official site of the Prime Minister's Office'. The navigation menu has 'Communicate' selected. A search bar is present with the text 'Enter search terms' and a 'Go' button. Below the navigation is a large 'E-Petitions' heading. On the left, there is a list of links: 'Petitions home', 'View petitions', 'About e-petitions', 'Step-by-Step Guide', 'FAQs', 'Terms and Conditions', and 'Privacy Policy'. The main content area features a graphic with a clipboard and a blue pen, with the text 'View petitions'. To the right of this graphic is a text block: 'Petitions have long been sent to the Prime Minister by post or delivered to the Number 10 door in person. You can now both create and sign petitions on this website too, giving you the opportunity to reach a potentially wider audience and to deliver your petition directly to Downing Street.' Below this text is a search bar labeled 'Search petitions:' and a 'Go' button. On the far right, there is a 'Newsletter' section with a dark blue background and white text: 'Sign up to our newsletter to keep updated with the latest information from Number 10' and a 'Click here to subscribe' link.

Figure 2: Petitions site within the official site of the Prime Minister's Office

b) Transparency and the monitoring of government activities

Transparency implies openness and accountability, and the opening up data makes it easier for citizens to monitor their government's activities including the services they provide, how efficiently taxes are spent, lobbyist activities, electoral activities and corruption.²⁴

In 2010, to provide transparency, the UK government made COINS available (see Fig: 3).²⁵ One of the world's largest government databases²⁶ it provides a detailed record of UK public spending.^{27,28} Its publication provided news stories. One story raised the question of who 'monitors the monitors' when it emerged that the body in charge of monitoring government spending spent over £84 million on office refurbishment including £2.33 million on furniture and £20 million on temporary accommodation. Another story highlighted past policy failures when it emerged the Department for Energy and Climate Change spent over half its £3 billion budget on nuclear waste legacy, leaving other programmes with insignificant budgets.²⁹

²⁴ <http://razor.occams.info/pubdocs/opendataciviccapital.html>

²⁵ <http://poynder.blogspot.com/2010/06/free-our-data-for-democracy-sake.html> The release of COINS followed a parliamentary expenses scandal which demonstrated the limitations of the Freedom of Information Act.

²⁶ <http://www.freeourdata.org.uk/blog/2010/06/> advises that it contains 24 million spending items in a CSV file

²⁷ http://www.hm-treasury.gov.uk/psr_coins_data.htm

²⁸ Criticisms have been made that the data is historical rather than current (the year 2010-2011)

²⁹ <http://www.guardian.co.uk/politics/2010/jun/09/spending-watchdog-costs-80m-offices>

The screenshot shows the data.gov.uk website interface. At the top, there is a search bar with the text "What are you looking for?" and a "Search" button. Below the search bar is a navigation menu with tabs for "Data", "Apps", "Ideas", "Forum", "Wiki", "Blogs", "Linked Data", and "Resources". The main content area features a dark header for "Combined Online Information System" with a post date of "23/11/2010", "87 comments", and a "Download" button. The description states: "The Combined Online Information System (COINS) is a database of UK Government expenditure provided by government departments. The data is used to produce reports for Parliament and the public including: expenditure data in the Budget and Pre-Budget reports; Supply Estimates; Public Expenditure Statistical Analyses (PESA); the monthly Public Sector Finance Releases. It is also used by the ONS for statistical purposes." Below the description are links for "Overview" and "Resources (download)". A "Download" button is also present, with a sub-link for "In-year pending data Q1 - 2010-11". To the right, there is a "Share this" section with social media icons and a "Related apps" section featuring "Where Does My Money..." by Jonathan Gray, which includes a visualization of data points.

Figure 3: COINS database available on data.gov.uk presented with a visualisation of the data by WhereDoesMyMoneyGo

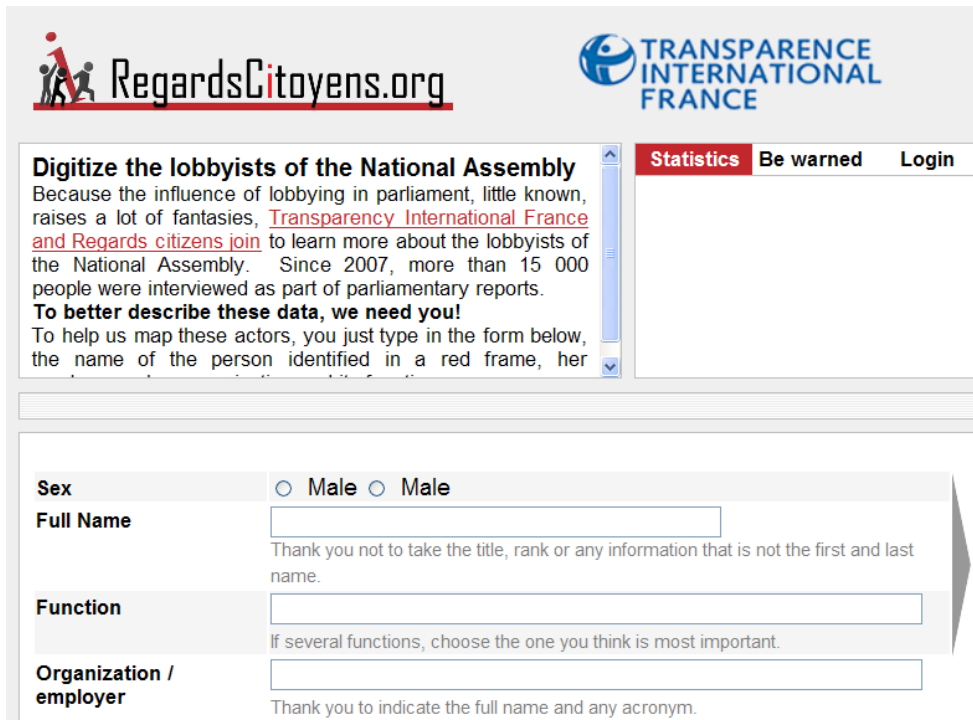
Initiatives to scrutinise government spending can be found in other EU countries. For example, [Offener Haushalt](#) is a German civil society project aiming to scrutinise government budgets by ‘scraping’ data from the finance ministry’s website (and attributing sources), and [GovData](#) collects procurement information from about 70 Swedish authorities that amounts to about 90% of central government spending.³⁰ The EU budget and how it is spent is also the subject of scrutiny. The 2005 European Transparency Initiative was intended to open up 55 billion Euros to public scrutiny. However, [followthemoney](#) reports weakness in the legal framework and ‘bureaucratic obfuscation’ by member states. The scale of potential savings through transparency is illustrated by a Canadian citizen who analysed revenue agency data and exposed \$3.2 billion fraudulent claims of charitable donations to evade tax.³¹ Although this data was not available on-line it demonstrates the benefits of making such data available.

Where data is not readily available citizen groups are finding alternative ways to create data in order to monitor. In 2010 [Regards.Citoyens](#), a transparency project in France that monitors parliamentary representatives partnered with Transparency International France to monitor lobbyists of the National Assembly. 1000 volunteers digitised publicly available

³⁰ Sand, Fredrick (2010) *Topic Report No. 9: PSI in Sweden: from infringement to enforcement?* Available at http://www.epsiplus.net/topic_reports

³¹ <http://www.thestar.com/news/investigations/charities/article/287682--charity-rules-beefed-up>

parliamentary reports (containing over 16,000 names)³² in less than two weeks to create a ‘crowdsourced’ database and work is now in progress to analyse the data (see Fig: 4).³³



The screenshot shows the website 'RegardsCitoyens.org' with the 'TRANSPARENCY INTERNATIONAL FRANCE' logo. The main heading is 'Digitize the lobbyists of the National Assembly'. Below this, there is a text block explaining the project: 'Because the influence of lobbying in parliament, little known, raises a lot of fantasies, Transparency International France and Regards citoyens join to learn more about the lobbyists of the National Assembly. Since 2007, more than 15 000 people were interviewed as part of parliamentary reports. To better describe these data, we need you! To help us map these actors, you just type in the form below, the name of the person identified in a red frame, her'.

Navigation links include 'Statistics', 'Be warned', and 'Login'. The form fields are:

- Sex:** Radio buttons for 'Male' and 'Male'.
- Full Name:** A text input field with a note: 'Thank you not to take the title, rank or any information that is not the first and last name.'
- Function:** A text input field with a note: 'If several functions, choose the one you think is most important.'
- Organization / employer:** A text input field with a note: 'Thank you to indicate the full name and any acronym.'

Figure 4: Crowdsourcing the database to analyse lobbying activities from open parliamentary reports

c) Better information, data journalism and solving policy problems

Whilst citizens want transparent government some data is difficult to understand. A recent German survey found that although two-thirds of citizens want transparent government only one-third want to see the data.³⁴ This creates a role for others to make the raw data more usable. In particular, it paves the way for ‘data journalism’ in which journalists not only to scrutinise policy but also make it available for others to use so as to verify or refute the case, or to build on the case.³⁵ The power of data journalism is illustrated by *The Guardian* data blogs.³⁶ Not only may data journalism lead to a higher quality of reporting, it also means a great potential for improved analysis and visualization in which citizens can participate.³⁷

³² <http://www.regardscitoyens.org/numerisons-les-lobbyistes-de-lassemblee-avec-transparence-international-france/>

³³ http://www.epsiplus.net/news/news/french_crowd_sourcing_project_completed

³⁴ <http://www.zeit.de/digital/internet/2010-08/umfrage-open-data>

³⁵ Datasets can be provided via links or be prepared by the journalist, for example, in a spreadsheet.

³⁶ <http://www.guardian.co.uk/news/datablog/2010/nov/08/housing-benefit-reform-impact-area>

³⁷ Dariusz Glazewski (2010) *Data driven journalism - what a refreshing perspective* available at http://www.epsiplatform.com/guest_blogs/data_driven_journalism_what_a_refreshing_perspective

The availability of data in the public domain can also support think tanks, researchers and political organisations to make an independent analysis of the implications of policy and new tools are enabling analysis. Where data is available, the impact of government policies can be rapidly assessed with responses emerging within hours of a policy announcement. Those opposed to a policy can also propose solutions to problems based on the data available. The Open Knowledge Foundation, for example, launched [energy.publicdata](#) to coincide with a European Council meeting discussing energy policy as a core topic. The application aims to help to put European energy policy into context and also provides a visualisation to support policy-makers and citizens.³⁸

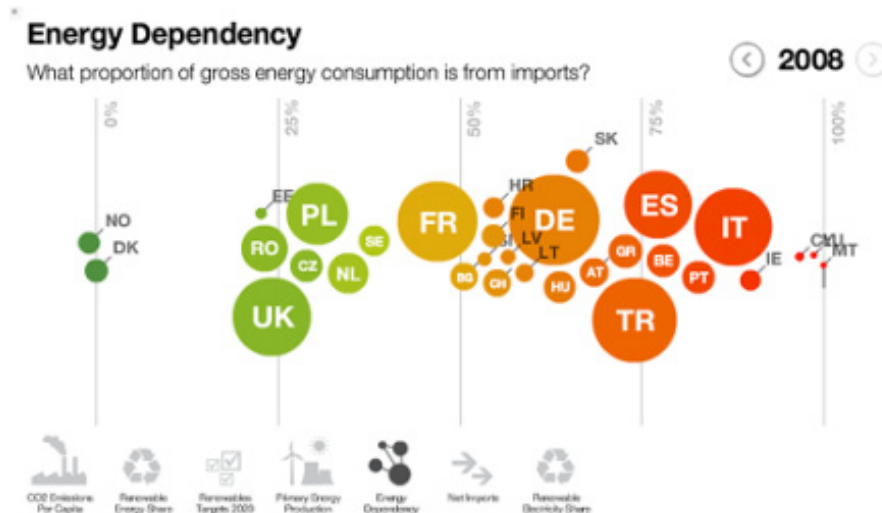


Figure 5: Open Knowledge Foundation's visualisation of energy dependency

Of particular significance is the potential to solve complex policy problems.³⁹ David Eaves sees the provision of open data as a means to involve professionals and volunteers outside of government who may provide an 'interesting analysis or a different perspective that can dramatically enhance a debate.' He cites an architect company who used Vancouver's open data to investigate how global warming might transform the city's shorelines as sea-levels rise and storm surges increase not only to highlight the need to reduce CO₂ emissions but for better planning.⁴⁰ The ability to access data online reduces costs enabling speedy download and report creation helps to advance debate.⁴¹

Within the EU environmental data is being developed within the framework of the INSPIRE Directive and the European Space Agency's collection of satellite data. Open access to this

³⁸ <http://blog.okfn.org/2011/02/04/europes-energy-a-new-mini-app-to-put-the-european-energy-targets-into-context/>

³⁹ <http://datadrivenjournalism.net/>

⁴⁰ <http://www.straight.com/article-298192/vancouver/get-ready-rising-sea>

⁴¹ Eaves, D. (2010) Case Study: How Open data saved Canada \$3.2 Billion [online] <http://eaves.ca/2010/04/14/case-study-open-data-and-the-public-purse/> 14th April 2010

data has the potential to support the use of other datasets in addressing many of the world's complex policy problems.

4. Better services

The second category of benefit in opening up data is the potential for better more effective services. The opening up of public sector information ties in with strategic ambitions of e-government providing the ability to make comparisons of quality and cost between public sector bodies, and to deliver citizen-oriented services online.⁴² Web technology increases the potential for citizens to engage with the design of services and gives rise to the need for government to engage in a dialogue with citizens rather than to make an assumption of how a service, and its interface with the public, is designed. This section discusses the benefits of open data in the provision of services to the citizen.

a) Reduced transactional costs and improved knowledge

Exposure to public scrutiny can improve service standards or reduce costs as public sector bodies are required to account for discrepancies. Government spends a lot of time answering queries. Providing data for the citizen online in a searchable format can lead to a drop in the cost of servicing customers for the PSI holder whilst increasing transparency. When Bristol City Council introduced its open data catalogue it identified benefits in reduced transaction costs. It claimed that the cost of a typical transaction was up to 15 times more expensive if answered in person (cost £15) or telephone (cost £12) than if answered over the internet (cost £1).⁴³ Furthermore, costs from servicing Freedom of Information Act requests could be reduced.^{44,45}

Data can enable citizens to compare services. For example, it is possible to find statistics on the success of routine hospital operations.⁴⁶ Businesses and civil society organisations can also take available PSI and present it to users in a more readily available form. In Sweden **Omvård** uses official statistics to compare services in hospitals and health care organisations to make it easier for patients to make informed choices in health care, **Lagen** a non-profit volunteer-run web site provides access to legal information with advanced linking and

⁴² Open data is a policy objective of the Malmö Ministerial Declaration on eGovernment and the European Digital Agenda

⁴³ <http://www.connectingbristol.org/2010/06/07/b-open/>

⁴⁴ <http://countculture.wordpress.com/2010/10/20/opening-up-council-accounts%E2%80%A6-and-open-procurement/>

⁴⁵ It was estimated in 2007 that the annual cost of FOI requests in the UK was £26 million for central government and that local government costs are similar

⁴⁶ <http://data.gov.uk/blog/my-top-ten-datagovuk-datasets-guest-post-simon-rogers>

[Jobbkartan](#) allows searches for vacant jobs using post codes from the Employment Service's web listings.⁴⁷

Public bodies are also able to use open data made available to complete their tasks more efficiently. In the Netherlands, open data is being used by the Ministry for Education to answer citizen's queries, by the Cultural Heritage department working with historical societies and the Wikimedia Foundation to improve data sets, and by the Amsterdam fire brigade to get intelligence on hazardous materials in buildings, water levels in rivers, etc.⁴⁸ In the UK [LG Group Inform](#) is an online service intended to achieve benefits for public sector bodies working with open data. Local government members will be able to upload their own data and access other local authority data and national sources (e.g. national statistics). It will allow councils to compare data to make informed decisions, reduce costs and improve services. The service will require that all data sources are both open and linked.

Sharing data can also enable services to be improved. MySociety's [FixMyStreet](#) (shown in Fig: 6) allows individuals to report local problems (like broken paving slabs or street lighting) which it then sends to the responsible body. The site tracks statistics on problems and repairs done, and lets users browse the database or set up an email alert to be told of problems reported within a local area. This 'crowdsourcing' approach reduces lead times in dealing with problems and enables both the responsible body and citizens to get a better view of what is happening in their area. FixMyStreet is built on open source code so that people can easily launch versions in other countries. [Verbeterdebuurt.nl](#), for example, is a Netherlands initiative led by [CreativeCrowds](#).⁴⁹

⁴⁷ Sand, Fredrick (2010) *Topic Report No. 9: PSI in Sweden: from infringement to enforcement?* Available at http://www.epsiplus.net/topic_reports

⁴⁸ Ton Zijlstra (2010) *ePSIplatform Topic Report no. 17 State of Play: PSI in the Netherlands* available at http://www.epsiplus.net/topic_reports

⁴⁹ <http://www.openinnovators.net/category/corporate-crowdsourcing-co-creation/page/2/>

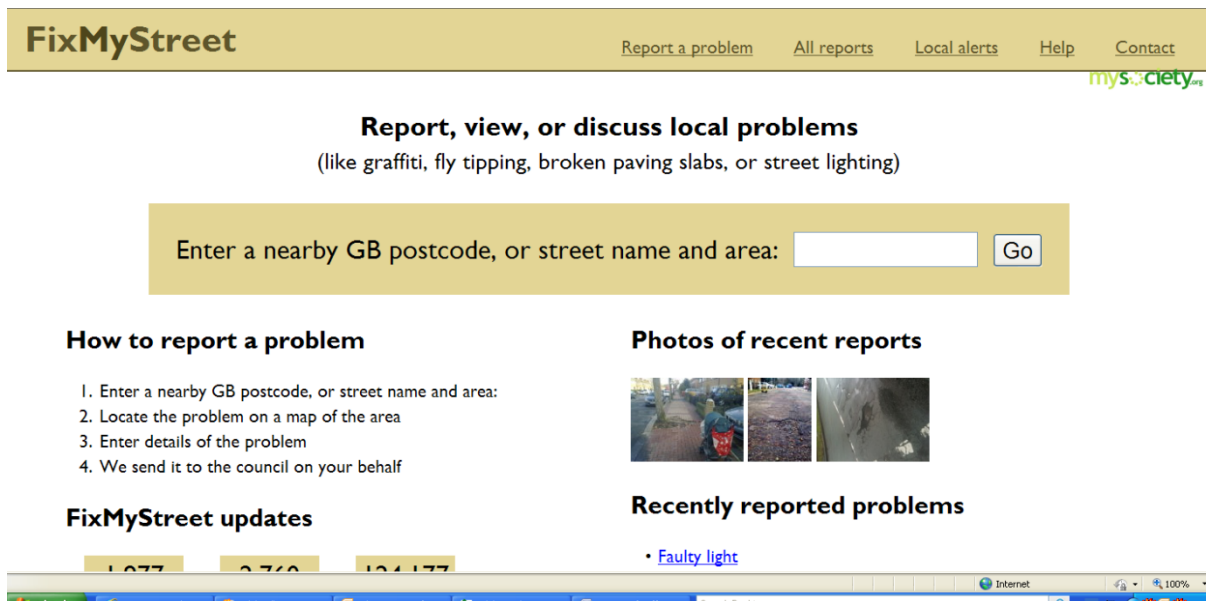


Figure 6: FixMyStreet

It is not possible for public data originators to identify what innovations can be achieved by individuals using data and combining it with other datasets. Open data advocates encourage ‘hack’ days and competitions to explore the potential applications that can be derived from operational data. Websites like <http://www.ilive.at/> (a winner in Washington, D.C.’s 2008 Apps for Democracy contest) has transformed city data into useful tools providing information on neighbourhood life for someone moving to the area.

b) Dialogue in design of services

Once data is on-line and available it can be interpreted directly by the citizen or by citizen groups. This enables citizens to engage directly in the design of services through access to relevant data. [Frankfurt Gestalten](#) aims to bring news and local policy decisions down to the district level. It tracks issues to a neighbourhood or street, tags documents with key words for access, emails citizens about changes in neighbourhood (e.g. planning applications), provides internet discussions (e.g. plans for a speeding camera), and enables citizens to bring change ideas or find neighbours with similar interests.^{50;51} Whilst it not easy to obtain city data,⁵² the ambition of the project (and projects like it in the USA)⁵³ shows that open data is not the

⁵⁰ <http://blog.zeit.de/open-data/2010/10/27/frankfurt-gestalten/>

⁵¹ The Frankfurt journal: http://www.journal-frankfurt.de/?src=journal_news_einzel&rubrik=10&id=10253 supports the project

⁵² <http://blog.zeit.de/open-data/2010/10/27/frankfurt-gestalten/>

⁵³ This project is similar to a much larger USA project [everyblock](#) which the New York Times describes as: ‘One of the most ambitious hyperlocal sites.’

endgame, but a precondition for something more radical: ‘The project will not only make the public information available to the city, it wants a dialogue.’

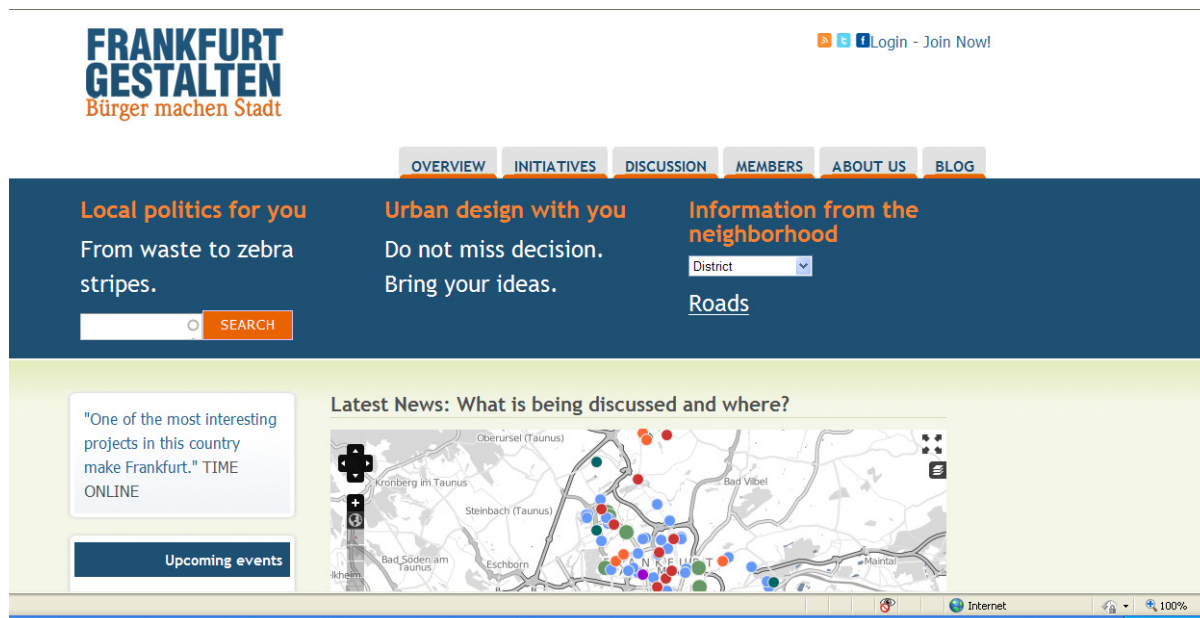


Figure 7: Frankfurt Gestalten

5. The knowledge economy

Finally, there are identifiable benefits for private innovation and the economy in re-using public sector information. However, whilst there is agreement that public sector information reuse benefits the knowledge economy, private innovation and growth, how best to realise these economic benefits of PSI is strongly contested. The potential economic benefits and the nature of the contestation are discussed in this section. It is also claimed that opening up data can encourage businesses to an area and develop tourism. These benefits are also discussed.

a) Developing an information industry

Many EU member states seek to recover costs of data production where it has a recognised commercial value (e.g. mapping and weather data). Whilst advocates perceive this as reducing the cost to taxpayers, re-users are concerned about the high cost of data which may be over-specified for their needs and copyright restrictions which limit sharing. In addition, governments have permitted the public sector to engage in provision of value-added services

Opening up government data: making the case

in competition with the private sector which has resulted in reusers concern that there is a lack of a level playing field when competing with public sector providers. In the USA a different view of PSI has prevailed in which data is perceived as a public good. This has led to information society policies in which data once produced is either free to citizens and reusers or sold at the marginal cost of distribution (considered to be zero when distributed over the internet). Various studies demonstrate how open government data benefits a national economy in comparison to the cost-recovery model prevailing in Europe.

When the European Commission commissioned a report to assess the economic case for open data, Pira International (2000)⁵⁴ found that the USA's open access policies had led to an information content industry five times larger than Europe. Characterising cost recovery policies as a barrier to economic growth, it claimed that selling PSI products at a lower cost would lead to higher employment and increased tax revenue. It concluded that the EU's information market would not have to double before tax receipts would exceed losses from charging and that the difference in size of the USA and EU information content industries represented the potential for economic growth from open access policies (see table 1):

	EU	USA
Investment in PSI	9.5 billion Euro pa	19 billion Euro pa
Size of information content industry	68 billion Euro pa	750 billion Euro pa

Table 1: Size of information content industry compared to investment in PSI (Pira International, 2000).

The weather information industry illustrates the potential. It is 50 times larger in the USA than Europe and is estimated to be worth \$1.5 billion to \$2 billion a year. Weiss (2002),⁵⁵ a weather information industry expert, attributes this difference in size to open data. He argues that charging at the marginal cost of dissemination for PSI will lead to optimal economic growth and this outweighs the benefits of cost recovery approaches.

However, there is a strong preference within finance ministries for the economic benefits of a reliable income stream. When the UK Treasury (2000)⁵⁶ reviewed open data policies, it accepted that there was a risk that PSI holders might charge high prices to a low volume market of captive users thereby recovering costs without developing the information market. However, it retained commercial operations. The Office of Fair Trading (2006) estimates that

⁵⁴ PIRA International (2000) Commercial exploitation of Europe's public sector information report available at: http://ec.europa.eu/information_society/policy/psi/docs/pdfs/pira_study/commercial_final_report.pdf

⁵⁵ Weiss, P. (2002) *Borders in cyberspace: conflicting public sector information policies and their economic impacts* Available from: <http://www.primet.org/documents/Weiss%20-%20Borders%20in%20Cyberspace.htm> [Accessed on: 18th January 2005]

⁵⁶ Treasury (2000) *Spending Review 2000* Available from: http://www.hm-treasury.gov.uk/spending_review/spending_review_2000/associated_documents/spend_sr00_ad_ccrappc.cfm

Opening up government data: making the case

there is £500 million of untapped economic value in the UK PSI market on top of the £590 million currently generated through cost recovery.⁵⁷

There are very few examples of releasing data from cost recovery policies. However, the example of the 2002 Danish agreement to make address data free of charge highlights the potential benefits of open data.⁵⁸ It resulted in improved data quality because of increased use and cross-sector benefits from using the same reference when exchanging information that contains an address reference. It has impacted the consumer market: 46% of families (about 1.3 million) have a GPS navigation system each with a copy of address data.⁵⁹

A 2010 a study provided evidence of significant financial benefits from the agreement. The report shows that whilst the total cost of the agreement up to 2009 was around 2 million Euros the benefits to society from 2005 to 2009 amounted to around 62 million Euros. These benefits will increase as costs decline - in 2010 the social benefits from the agreement are anticipated to be about 14 million Euros while costs will total about 0.2 million. Whilst most of these benefits accrued to the private sector (approximately 70%) the public sector also benefited significantly (approximately 30% of the total benefits). This success is in sharp contrast with the UK government policy. Its decision to leave it to the 'market' has led to the spectacle of two public sector data holders competing with each and government subsidies to one whilst proclaiming the other in its e-government strategy.⁶⁰

It is not possible for the originators of public data to identify what new knowledge and innovation can be achieved by others using data and combining it with other datasets. Advocates of open data therefore encourage public bodies to make data available irrespective of its commercial value to allow for the possibility of innovation by citizens or the private sector.⁶¹ For example, the cartographic map once digitised lends itself to new ways of manipulating it, leading to new services and products. Benefits to the knowledge economy are diverse as noted in a report on PSI reuse in Denmark:

‘The advent of Web 2.0 has led to the emergence of new markets for digital content and new ways to create digital solutions involving citizens. In order for Denmark to secure the future of digital products and services, and maintain its position as a lead country in digitisation of society, it is necessary to create a framework where everyone can be both consumers and creators of digital products and services.’

Two reports provide further evidence of the potential re-use market. In 2009 Denmark's Open Data Innovation Strategy commissioned report was commissioned which estimated the business potential of the reuse of public data in Denmark could be more than 80 million

⁵⁷ Office of Fair Trading (2006) The commercial use of public information (CUPI) available at http://www.offt.gov.uk/shared_offt/reports/consumer_protection/oft861.pdf

⁵⁸ Where financial charges are made these are for the cost of distribution

⁵⁹ http://www.epsplus.net/news/news/value_of_danish_address_data

⁶⁰ Pollard (2006) Co-ordinating the sharing of spatial data in the UK PhD

⁶¹ <http://www.bcbusinessonline.ca/bcb/business-sense/2010/10/08/difference-data-makes#ixzz15CkuhJm8>

Euros a year.⁶² The UK government's decision to release its financial database COINS (see Section 3) was made on the basis of estimates that it could stimulate an industry to analyse and create online services from it worth up to £6bn a year.^{63;64}

b) Supporting business, fostering tourism and cultural activities

Businesses have been analysing government data for decades, for example to decide where to open a new office or shop. Open data reduces the transaction costs of getting this and other information making it possible just to visit the website and download what is needed. Those areas that most enable this kind of data availability are considered to be more likely to attract SMEs and inward investment.⁶⁵

There is a growing awareness of public sector information being used to encourage tourism. This can be by providing generalist information, for example, a tourist map of the Basque country brings together several datasets hosted at Open Data Euskadi or ⁶⁶ can provide information on tourism specific to an area, e.g. caving data by the Piedmont region in Italy.

Tourism can be fostered whilst developing cultural facilities for the residents of a locality. This is demonstrated by Marseille-Provence a French region of 2 million people. In 2013, as European Capital of Culture, Partners of the [Association Marseille Provence 2013](#) (local, state, universities, etc) will open public digital data so as to widely disseminate images and sounds around the works of artists for the arts and culture programme. It has also launched a digital territorial study to expanding the fields of tourism, transport and community life.^{67;68}

Google's Art Project demonstrates the potential to combine tourism and culture. Seventeen museums ⁶⁹ have contributed to bringing its street view technology to art works - providing virtual tours and close ups. It is believed that the project will benefit those who can't travel

⁶² Catherine Lippert (2010) *ePSIplatform Topic Report No: 20 Public Sector Information Reuse in Denmark* available from http://www.epsiplus.net/topic_reports/public_sector_information_reuse_in_denmark

⁶³ <http://www.titticimmino.com/2010/06/04/coins-opendata-and-transparency-lesson-1-for-data-gov-it/>

⁶⁴ The COINS database contains millions of rows of complex data in large files and some competence in manipulating large volumes of data is required. It is likely these data will be most easily used by organisations that have the relevant expertise and then presented in a way that is more accessible to the public.

⁶⁵ <http://data.london.gov.uk/>

⁶⁶ <http://opendata.blog.euskadi.net/blog-en/good-practice/tourism-map-of-the-basque-country/>

⁶⁷ http://www.epsiplus.net/news/news/marseille_provence_european_capital_of_culture_2013_digital_project_on_release_of_government_data

⁶⁸ <http://www.reseaufing.org/pg/blog/slyan/read/25900/marseilleprovence-2013-sengage-dans-la-libration-des-donnees-publiques-en-tant-que-capitale-europenne-de-la-culture>

⁶⁹ Participating museums: Alte Nationalgalerie, Berlin; Freer Gallery of Art, Smithsonian, Washington DC; The Frick Collection, NYC; Gemäldegalerie, Berlin; The Metropolitan Museum of Art, NYC; The Museum of Modern Art, NYC; Museo Reina Sofia, Madrid; Museo Thyssen, Bornemisza, Madrid; Museum Kampa, Prague; National Gallery, London; Palace of Versailles, Versailles; Rijksmuseum, Amsterdam; The State Hermitage Museum, St Petersburg; State Tretyakov Gallery, Moscow; Tate Britain, London; Uffizi Gallery, Florence; Van Gogh Museum, Amsterdam.

and provide an inspiration to travel to the galleries. It is also thought that creative ‘hackers’ may draw on the newly available art from the project in unforeseen ways.⁷⁰

6. Challenges of opening up data

Whilst there are benefits in opening up public sector information for re-use custodians of public data also perceive risks that can hamper the opening up of PSI. In particular, there are issues relating to resources, data quality and institutional change. However, where there is a commitment to opening public data there is evidence that these risks can be managed.⁷¹ This section briefly discusses issues raised in order to demonstrate this.

‘We don’t have the resources’

Many countries are currently experiencing budgetary constraints leading to reluctance to take on more tasks particularly where the public sector body itself may not benefit. However, data publishing need not be complicated or costly if a step-by-step approach is taken and effort can be made to identify agency benefit from investment in data publishing, for example, transaction costs in responding to queries or Freedom of Information requests can be reduced. The possibility of an increased workload due to enquiries from the public or media about published data can be managed by providing information about the data and the level of support that can be expected. In addition, contact with the public can lead to an increased awareness of the importance of data accuracy which can benefit the custodian.

‘We are worried about data accuracy and data privacy’

Publishing public data can reveal inaccuracies that embarrass the PSI custodian or cause harm but such anxieties can be minimised by a gradual approach, where stakeholders are consulted and the reliability of the data explained. When the UK government published crime data journalist criticised inaccuracies (in part caused by the recording of some crimes at the police station itself) and reported fears that the information could impact negatively on house prices. However, even in this high profile case with the site receiving 18 million hits per hour, it was possible to manage expectations. Journalists argued that the data will improve through exposure, that not all stakeholders had been consulted and that crime should not be swept ‘under the carpet’ simply because of fears about house prices.⁷²

Public bodies also worry that they may release personal data (or confidential business data) and believe that determining whether or not data can safely be published requires expert legal advice. However, it is usually the case that datasets which contain personal data have already been identified and marked with privacy flags to meet data protection legislation.

⁷⁰ <http://www.creativereview.co.uk/cr-blog/2011/february/google-art-project>

⁷¹ Catherine Lippert (2010) *ePSIplatform Topic Report No: 20 Public Sector Information Reuse in Denmark* available from http://www.epsiplus.net/topic_reports/public_sector_information_reuse_in_denmark

⁷² <http://www.guardian.co.uk/uk/2011/feb/01/online-crime-maps-power-hands-people>

‘It is our data – we decide how it is used’

Cultural barriers are difficult to address because they are often unacknowledged. There is a perception that to publish data is to relinquish the power associated with the data to a wider audience, described as a fear of loss of ‘interpretational sovereignty’.⁷³ However, if other barriers are removed PSI holders can be persuaded of the benefits.

If there is a lack of supportive policy measures - often determined by higher level political action -this can make it harder to publish data. Even so, where there is a lack of policy there may be sufficient interest amongst public administrators to take some steps forward. In the Netherlands, there is an active community of civil servants, (Ambtenaar 2.0) which discusses issues around digitisation and PSI re-use.⁷⁴ Many public administrators want data to flow for the purposes of e-government if not PSI re-use.

Where public data is subject to cost recovery policies resistant to releasing data can become more entrenched because of the potential loss of a secure income stream. In many cases, the decision cannot be made by the PSI holder acting alone. This can impact on the realisation of the democratic, social and economic benefits and may require higher level political engagement. The approach of the open data movement has been to seek open access to some datasets even where cost recovery remains in place. As result, even PSI holders who charge for data have been willing to open up access to some datasets whilst maintaining a charging policy.

7. Generating good practice in developing open data catalogue

Across the EU, there is a lack of uniform practice when it comes to access to public data. Many potential reusers of public data do not know that data exist and many public authorities are not aware of the potentials of public data reuse. To resolve these issues it is not only beneficial to raise the level of public sector awareness of the potentials of data reuse but to ensure that the process of opening data is not made overly challenging. This section provides some simple guidelines of on what is important to develop of a functional, sustainable, usable interactive PSI portal drawing on material available at ePSIplatform and other sources.

- **Determine an information policy appropriate to the context:** a policy is more likely to be achievable if initially it is planned to keep the project simple. A civil society site run by volunteers may have fewer resources than a public authority and may seek only to make available a list of current datasets.

⁷³http://assets1.csc.com/de/downloads/CSC_policy_paper_series_01_2011_unchartered_waters_state_of_open_data_europe_English_2.pdf

⁷⁴ Ton Zijlstra (2010) *ePSIplatform Topic Report no. 17 State of Play: PSI in the Netherlands* available at http://www.epsiplus.net/topic_reports

- **Plan for sustainability:** design of a website is relatively straightforward, the challenge is to keep the site fresh and interesting so that users revisit, to ensure the data and links to data are maintained and to meet users needs. This requires consideration of the organisational and social aspects and decisions about whether to provide competitions, hack days, etc.
- **Learn from existing sites:** it is worth evaluating existing sites, and talking to developers, custodians and users of existing sites to discover what works and what doesn't.⁷⁵
- **Design with re-users and consult on what data to release:** data is not just for developers but for people who want to look up a fact, analyse the data or write a report. Interact with the community (citizens, developers, businesses) and be aware that there may be different interests and needs e.g. a commercial re-user requires a clear regulatory procedure. However, the focus should remain on a simple reliable accessible infrastructure rather than struggling to meet all needs.⁷⁶
- **Use participative design tools:** web 2.0 technologies support participative design and help to keep the site interesting. If the resources are available plan to provide blogs, wikis, forums, showcasing of apps and opportunities to request new datasets, etc.
- **Provide information about the data:** re-users need to know information such as the owner of the dataset, its provenance, its quality and the frequency of updates. This also reassures PSI custodians that their data is understood and supports the release of data.
- **Provide simple, digital, standard licences:** if rights statements are confusing or missing, re-use of PSI may be sub-optimal. Tools, such as the Creative Commons licenses used internationally, support the sharing of PSI - communicating information use rights and permissions in advance. These licenses grant broad access and re-use for persons wishing to use PSI while ensuring credit is given to public sector bodies providing the PSI.⁷⁷
- **Publish raw data in reusable formats where possible:** re-users prefer raw data in reusable formats (e.g. xml, cvs, rdf) that enable automated processing but publish in other formats if that is the only format available.
- **Make your site known and accessible:** reusers may be seeking to collate data from more than one site and it is important to find information about catalogues and PSI portals available. Also ensure that the data catalogue is inclusive in its message and available to a wide range of users and not just developers.
- **Consider alternative ways of delivering PSI for re-use:** it might be possible to approach civil society initiatives or repositories (e.g. [Freebase](#)) to make PSI data available alongside existing data.

⁷⁵ See Timothy Vollmer (2011) Topic Report No: 25: State of Play: Public Sector Information in the United States for a discussion on [data.gov](#)

⁷⁶ Robinson, David, Harlan Yu, William P. Zeller, and Edward W. Felten, *Government Data and the Invisible Hand*, 11 Yale Journal of Law & Technology, 2009, at 161. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1138083

⁷⁷ Timothy Vollmer and Diane Peters (2011) Creative Commons and Public Sector Information: Flexible tools to support PSI creators and re-users available from http://www.epsiplus.net/topic_reports

Data catalogues are about more than data delivery. They are a part of a process of using technology to develop the knowledge economy, to improve services and to enable deeper democratic engagement.⁷⁸ In the longer-term there is a need to integrate the opening up of data into operational information systems and to be interoperable with data held in other PSI portals if the full potential of PSI re-use is to be realised. However, it is necessary to take simple steps towards publishing in order to develop a network of catalogues.

8. Concluding remarks

This report has provided case stories to raise awareness of the democratic, social and economic potentials of open data. It is intended to provide support for open data champions. As PSI re-use increases these examples will be added to with richer stories of benefits achieved.

In identifying these benefits it is not intended to suggest that there are not significant institutional challenges in opening up public sector information. There are. The move towards open data requires a paradigm shift within the institutions of governance from one of automating to one of informing. This requires an awareness by politicians and public administrators of the way in which information technology illuminates are activities and enables us to innovate in ways not previously anticipated.

The report endeavours to demonstrate the simplicity of innovations that use open source software to do things in new ways outside of an automate and control paradigm because if these benefits are understood then resistance by public sector bodies to change to a new way of working can be managed. The greatest resistance will almost certainly continue to be in the area of cost recovery because of the economic benefits of an assured income stream to the PSI holder. However, many public administrators wishing to share information for the purposes of e-government are already aware of the benefits of re-use without restrictions of cost and copyright and these administrators are potential allies for PSI re-users.

The benefits will increase as a network of data catalogues and PSI portals is established but whilst there has been waves of interest in the opening up of data, it is an incremental process and the report has attempted to provide some guidelines to highlight the need to act to make step changes but to keep the design of data catalogues simple, reliable and sustainable. It is these simple actions and the re-use of data from these catalogues that have the potential to develop the open data ecosystem - making it stronger and more likely to embed a new ways of doing things.

⁷⁸ <http://opengovernment.labs.oreilly.com/ch01.html>