Beyond Access:
Open Government Data
& the Right to (Re)use Public Information

7 January 2011

Electronic copy available at: http://ssrn.com/abstract=2586400

This report is published under a Creative Commons License which allows you to use, copy, distribute, and adapt it, provided that you credit the source ("*Beyond Access: Open Government Data and the Right to (Re)use Public Information*" by Access Info Europe and the Open Knowledge Foundation, 7 January 2011) and that you share it in the same way.

Electronic copy available at: http://ssrn.com/abstract=2586400

Contents

Executive Summary – Key Findings

This report reviews the achievements of and challenges facing the access to information and open government data movements in making government information more accessible. It reviews the nature, strategies and composition of the two movements, and analyses the current legal and technical obstacles to achieving full government transparency.

Each section of the report contains a series of recommendations for future actions by civil society, donors, governments, and intergovernmental organisations.

These are the Five Key Findings drawn from the research by Access Info Europe and the Open Knowledge Foundation:

Finding 1 - **There are serious shortcomings in the current international and national standards defining the scope of the right of access to information, resulting in the release of information in formats that cannot be reused.**

This report reveals that the standards for the right of access to information do not yet encompass a right of access to full databases, to raw datasets, and to information in electronic, machine-processable, and non-proprietary formats. These shortcomings are preventing full access to government data, with a number of countries excluding access to databases from the national access to information law and with a practice of releasing documents in formats from which data cannot be extracted for reuse.

When information is obtained in formats which cannot be processed by computers (such as hard copies, images or PDFs) or when the software support used is not of an open format, then members of the public are limited in their possibility of reusing the information. The access to information community therefore urgently needs to engage with open government data activists in tackling these law and standards issues which will otherwise impede further advances by open government data advocates in many countries.

Finding 2 – **Future transparency standards should be anticipated now, both to reduce technical obstacles to releasing the information down the line, and from a policy perspective to harness the full democratic potential of government data.**

If government information is to permit members of the public to produce added-value applications, to participate in decision-making, and to hold governments accountable, it needs to be "discoverable" (easy to find) and simple to access. This means that governments need to ensure that information is collected, ordered, and stored in formats which make it discoverable and easy to release in real time. Disclosure policies need to be anticipated when designing electronic record and document management systems. Records should be tagged and structured with possible disclosure in mind, and publicly accessible search interfaces should be built.In particular, the application of exceptions to large datasets needs to be designed-in to the construction of the databases so that non-sensitive information can be released.

Finding 3 – **There is a lack of clarity about who owns government data. Copyright, database rights and other ownership rights are restricting the right of the public to reuse government data.** In spite of the advances of the right to information movement in securing recognition that the information generated and held by public bodies belongs to the

public, government bodies around the world are still asserting intellectual property rights and ownership rights over the information they produce. This applies in particular to bulk datasets. As a result, this information is either sold or comes with restrictive copyright licences which prevent reuse.

There is an urgent need for review of the legal framework which defines who owns government information, what the intellectual property rights of public bodies actually are, and whether the case can ever be made for selling government information to members of the public. Solutions such as the use of Creative Commons Licences should be explored.

Finding 4 - **There is an unresolved conflict between the right of access to information as an inherent part of the right to freedom of expression and the limitations placed on reuse of government data through copyright licences and charges for commercial reuse.**

Given that international law has established that the right of access to information is directly linked to the right to freedom of expression, any limits on the reuse of information obtained from public bodies would seem to clash with the freedom of expression rights of members of the public. There is a need for more research and standard-setting to resolve this conflict which is currently undermining the social and democratic value of the right of access to information.

Finding 5 - **The access to information and open government data movements are not yet collaborating sufficiently closely and are therefore missing opportunities to advance the transparency agenda. At the same time there are key actors who can make linkages and serve as bridges between the two communities.** This study has found that the transparency agenda could be advanced more effectively if access to information and open government data advocates were to collaborate more closely. The rights-based approach of the access to information movement could complement the arguments about the economic and social benefits of releasing government data employed by open government data advocates.

The research for this report revealed that members of these movements do not talk the same language: open government data experts are not familiar with the law-based approach of access to information advocates, while the technical terminology employed by the civic hackers is baffling for the human rights activists. Further training and networking is needed for these two communities to be able to define common strategies and advocacy goals.

PART I: THE ISSUES

1. Introduction: The Right to Information and Open Government Data

Public bodies[1] produce and collect vast amounts of information as part of the daily functioning of government. There is tremendous social value in this information: it is necessary for the public to hold governments accountable, to participate in decision-making, and to know about and access government services. The use and reuse[2] of this information can contribute to the functioning of both government and of the democratic structures of society as a whole. There is also significant economic potential in this information which can be used by business to create added-value services which they can then market.

There are two main civil society movements which are campaigning for greater openness of the information, documents and datasets held by public bodies. The first is the right to information movement, which promotes a public right of access to information from a human rights perspective (please see Section 4). The second is the open government data movement which uses social and economic arguments to encourage the opening up of government data: putting such information into the public domain benefits society and can stimulate the economy (please see Section5).[3]

There is significant overlap between the aims of both movements, although there are also differences in the approaches and strategies employed by each to argue for more open government. One of the main differences is that right to information advocates place an emphasis on access to qualitative as well as quantitative information, which is often stored in the form of documents, whereas the open government data advocates focus on data that is held in government databases, and they are concerned with both the technical and the legal issues related to the access to and use of these datasets.

---

[1] For the purposes of this report, public or governmental bodies are all branches of state power, including the executive/administrative, judicial, and legislative branches, as well as private bodies performing public functions or operating with public funds. These are referred to as *public bodies* or sometimes *public institutions* or *public authorities*. This is consistent with the definition in the EU's Directive on Re-Use of Public Sector Information which defines a "public sector body" as "*State, regional or local authorities, bodies governed by public law and associations formed by one or several such authorities or one or several such bodies governed by public law*". The Directive also defines a "body governed by public law" as any body which has been "(*a) established for the specific purpose of meeting needs in the general interest, not having an industrial or commercial character; and (b) having legal personality; and* (c) financed, for the most part by the State, or regional or local authorities, or other bodies governed by public law; or subject to management supervision by those bodies; or having an administrative, managerial or supervisory board, more than half of whose members are appointed by the State, regional or local authorities or by other bodies governed by public law." See Directive 2003/98/EC on the Re-Use of Public Sector Information available at: http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive/psi_directive_en.pdf

[2] It has become commonplace to refer to "reuse" of government information. This term carries an implicit assumption that public bodies "use" information and members of the public "reuse" it. This assumption is not entirely consistent with the principle that all the public data is data for use by both public bodies and members of the public alike. Nevertheless, in this report we continue to refer to "reuse" of information, particularly in the context of specific rules on reuse of public sector information.

[3] There is no significant difference in this report between the use of **data** and **information**. Data could be conceived more as small chunks of information held in databases, spreadsheets, etc., but this is also information. We therefore use the terms relatively interchangeably, sometimes preferring information when talking about documents and what they contain.

Another difference between the two movements is that they have arisen from different parts of civil society. Right to information activists generally come from a human rights and democratisation background, and employ rights-based arguments and tools (such as engaging in drafting laws, monitoring, and litigation) to promote and defend the right of access to information. Open government data activists often come from a more technically-oriented background and aim to use new digital technologies to access, process, share and present data held by public bodies. Many of those involved in the emerging open government data movement have the technical skills to build IT applications, often web-based, which permit members of the public to access and make use of publicly-created information in new and innovative ways.

Both movements employ very similar arguments in favour of greater transparency. These include that the public has the right to hold governments accountable both at and between elections. Similarly information is needed to access public services and to participate directly in decision-making. Clearly, both goals are contingent on having access to information. Furthermore, transparency can contribute to the efficiency of public bodies by helping to reveal shortcomings in their functioning and by allowing the public to give feedback on the way in which public bodies function.

Opening up government information also helps release its potential social and commercial value. Much useful data is generated or compiled by governments and in the digital age data is undoubtedly a key resource for social and commercial activities. Thus by opening up data, government will drive the creation of innovative business and services that deliver social and commercial value. The benefits of doing this are considered further in Section 2 of this report.

1.1 What is Open Government Data?

The term "open government data" has come into prominence relatively recently, becoming popular in 2008 after the publication of a set of open government data principles by advocates in the US (See Box A).[4] The two main elements of open government data can be defined as follows:

- "**Government data**" is any data and information produced or commissioned by public bodies.
- "**Open data**" is defined as material which anyone can use for any purpose.

To qualify as "open", it must be possible for the government data to be freely copied, shared, combined with other material, or republished as part of websites which allow users to explore, analyze, visually represent, or comment on the material, as well as transform it into other formats. Examples of the datasets held by governments which can, potentially, be opened up range from national statistics to budgetary information, from parliamentary records to data about the locations of schools, hospitals, crimes, or post boxes.

Given the vast quantity of information held by public authorities, it is essential that those wishing to use the information can actually find it in the first place. There are various solutions for governments to make information more easily "discoverable". These include the creation of open data catalogues, and the development of "information asset registers". The issue of data discoverability is examined further in Section 2.

[4] See Open Government Data Principles at http://resource.org/8_principles.html, adopted in December 2007

Government data needs to come in a format which allows it to be processed and reused. There are various technical obstacles which can impede both access and use, such as information not being in an electronic format or being in a non-machine-readable format (such as a PDF from which data cannot be extracted), or data being held in proprietary software which cannot be read by open source software. These technical obstacles and the solutions which permit reuse are considered in Section 2.

The definition of "open government data" has particular implications when it comes to the intellectual property rights which govern the use of material obtained from public bodies. The definition does not strictly require the information to be completely free of copyright. Instead, any proprietary licences must be of the type which permits the information to be accessed, redistributed, and reused.[5] These issues are considered in detail in Section 3 of this report.

Box A

The Open Government Data Principles (December 2007)

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

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.

[5] The concept of open government data borrows heavily from and is a subset of the more general notion of "open data", which refers to datasets which anyone can use without legal, technical, or other restrictions. Such data is free from copyright or other restrictions on reuse. Examples of such data include statistics, mapping data (geospatial data), and scientific data which has potential market value but which also has tremendous public value if shared openly. For this reason much academic research, including scholarly publications and data, is now made available free of charge for anyone to reuse and redistribute. For the Open Knowledge Definition see http://www.opendefinition.org/okd/

1.2 How is the Right of Access to Information Relevant?

The right of access to information (ATI) is a right of members of the public to gain access to information held by public bodies. There are two mechanisms by which information can be released to the public in conformity with this right: either reactively in response to access to information requests or proactively at the initiative of public bodies.

The proactive dimension of the right of access to information is especially relevant for open government data activists, particularly as there are now emerging standards on exactly which information public bodies should make available without waiting for requests from the public. Proactive disclosure of information is considered further in Section 4.3.

The right of access to information has been linked by international human rights courts and by national law and jurisprudence to the right to freedom of expression. This has important implications for the right to use public sector information which could be considered as part of the right to freedom of expression. This implies that reuse should be free of intellectual property or other restrictions, although there is, as of yet, relatively little law or jurisprudence to support this. These relevant international standards are set out in Section 4.4 on Key Elements of the right to know.

In principle the right of access to information applies to all information held by public bodies, but in some countries databases are excluded from the scope of the right and in others the law is not clear, while practice varies across countries. Similarly, not all countries establish a right of access to information in electronic format wherever possible, and none of the access to information laws surveyed for this report made reference to machine-readable or open formats. These are areas where more standard-setting is needed, as considered in Section 4.4 on Key Elements of the Right and Section 4.5 on Future Challenges as well as in Section 2 on the Practical and Technical Aspects of Open Government Data.

One area of particular lack of clarity is the question of who owns government information. Many access to information laws presume that public information is to be accessible and in that sense these laws consider the general public as the legitimate owner of public information. However, it is still the case in many countries around the world that public bodies assert intellectual property rights such as copyright and database rights (intellectual property rights generated by compilation of a database even from pre-existing material) over the information they have generated or collected. Even where intellectual property rights are not asserted, public bodies tend to assume that they are the exclusive owners of the information and their economic model sometimes includes selling the information for profit. The implications for the open government data movement of these conflicting principles about who actually owns public data are in Section 3 on the Legal Issues.

Each section of this report looks both at challenges and solutions for increasing access to government data. The analysis is accompanied by recommendations directed at public authorities and at civil society, pointing to ways in which governments can take immediate measures to open up their data as well as enumerating the various ways for civil society organisations to support the drive for greater government transparency.

1.3 Linking open government data and right to information activists

This report demonstrates that the open government data movement shares common goals with the right of access to information community in that both aim to increase transparency of government so that all members of society enjoy the inherent social and economic value of information generated and collected with public funds.

One of the major differences between these two movements is that, while the right to information movement has put an emphasis on the obligation of public bodies to respond to requests for information, the open government data movement emphasises proactive release of large volumes of information in formats and under conditions which permit reuse. These technical and legal considerations are considered further in Sections 2 and 3.

At the same time, as noted in Section 4, access to information advocates have recently put increasing emphasis on the proactive disclosure of information on a large scale. This coincides with the demands of the open government data community for the release of full databases. These developments are coming largely in response to recent technological advances which make it possible for public institutions to release entire datasets at the press of a button. Researchers and campaigners now file access to information requests for entire datasets. Similarly, the journalistic specialisation of computer assisted reporting – also now referred to as "data journalism" – has resulted in a significant increase in demand for direct access to databases. Hence there is an increasing convergence in the demands of the access to information and open government data communities.

One area where the open government data advocates are taking the lead is in calling for release of information in formats which permit it to be reused. Another is in addressing issues of copyright, licensing and charges for reuse of public information. These questions are addressed further in Section 3, which also examines how the open government data and access to information communities can work together to press for fewer limitations on the use of information accessed from public bodies.

It is clear from this survey that although the two communities, OGD and ATI, comprise different actors, there is a strong potential for future collaboration. Access to information advocates have solid legal and policy experience in promoting government transparency, as well as belonging to an extensive international network which has a presence both in developed democracies and in the global south. The strengths of OGD advocates include that they bring new policy arguments for opening up government data in ways that are more comprehensible to the general public. Moreover, the OGD community has the technical expertise to demonstrate with practical examples the economic and social benefits potentially can derive from making use of government-held information.

➢ Recommendations:

It is recommended that the Open Government Data Community:

- Enter into dialogue with public officials at local and national level to encourage release of information which is likely to be of public interest. Strategically, it is recommended to focus on datasets which are easy to release so as to demonstrate the potential benefits of opening government data;
- Document law and policy at the national and international level to help advocates make a more compelling case for opening up government information in their country. To this end, open government data groups and individuals are encouraged to be in touch with

organisations specialising in access to information, academics, and other policy experts, who are likely already to have relevant information on the legal framework;

- Map out datasets that are already available and identify whether they are held in a format and under a licence which facilitate and permit their use;
- Identify key datasets and communicate to government officials and the general public the potential value the applications that could be built using them;
- Collaborate, both globally and country-by-country, with access to information advocates to define strategies that promote the above recommendations.

It is recommended that governments and individual public bodies:

- Ensure that within government those responsible for open government data and those already working on access to information, including at the ministerial level, and bodies such as information commissioners, collaborate closely. The governance of new open government data is an important issue to integrate with existing transparency rules, policies and mechanisms.
- Take the initiative to organise events and competitions around the release of data sets in order to stimulate the user community and demonstrate to a wider public the value of making government data open;
- Attend consultations with civil society organisations and prospective users of government information so as to clarify their information needs, in order to identify which datasets should be released as a matter of priority.
- Anticipate disclosure policies while designing electronic record and document management systems.

It is recommended that funders:

- Support citizen-driven projects to use government data that is currently available, providing seed funding for small initiatives. These could include competitions and information-design projects focused on promoting democratic engagement by the general public;
- Support targeted campaigns and pilot projects which press for access to and the right to use information from specific public bodies or in a particular sector of government activity. For example, campaigns could press for access to datasets from legislative and judicial bodies, which in many countries has not traditionally been made available in bulk;
- Support collaborative projects between open government data advocates and access to information experts to target the legal and policy framework that currently impedes the release of government datasets;
- Support the production of guides and toolkits and the organisation of workshops so that access to information specialists and other interested groups, such as journalists, can become acquainted with the technical aspects of the right of access to information in digital formats in order to be able to conduct effective advocacy. The workshops should include practical demonstrations of how databases work, what happens on a hack day, and how competitions are organised.
- Commission further research into the law relating to access to information, access to databases, copyright and reuse of public sector information in order to have a clearer

picture of national legal frameworks and to permit advocates to develop targeted law reform and litigation strategies;

- Commission further research into the economic aspects of release of government data, in particular the evidence that releasing information has a value to stimulating the economy, so as to equip advocates with arguments in favour of releasing information free of charge.

2. Practical and Technical Aspects of Open Government Data

For government data to be fully "open" in the sense called for by advocates of both the right of access to information and of open government data, there are a number of criteria which need to be met. These include that the information be structured technically in a way that makes it easy to find or discoverable, that the information be stored according to technical specifications that make it reusable, that information be available for downloading in bulk, and that information be available in formats which do not restrict reuse.

This section examines the practical and technical aspects of open government data, and maps the extent to which they are currently recognised in the law and practice of the countries surveyed in the research for this report. This section also examines the issues which will need to be addressed by both governments and transparency campaigners, and from both the ATI and OGD movements in the coming few years, in order to achieve true openness of government data.

This section also identifies where further research is needed in order to build an evidence base that can be used in advocacy for changes in law and practice.

The issues considered in this section are:

- **Discoverability**: making sure that information can be found;
- **Reusability**: technical aspects related to the reuse of information, including the right of access to information in electronic formats, machine-readable formats, and open file formats;
- **Enhancing Accessibility**: how to make provision for disaggregated, up-to-date and linked data;
- **Cost considerations**: what are the costs and cost benefits of opening up government data?

2.1 Discoverability

If the public is to make full use of open government data, it needs to know what information exists in order to be able to access it or to request it. Even the mere fact that data has been put online on the website of a public body is not sufficient if it is not easy to find; so data has to be findable. This is the issue of the "discoverability" of government data and in this section we will examine various options for improvement.

There are a number of solutions which make information more "discoverable". These include:

- government registers of the information they hold (such as Information Asset Registers), official open government data catalogues, and community-driven data catalogues;
- search engine optimization, proper indexing and metadata management, microformats, semantic web and linked open data techniques.

Ensuring that members of the public can find the government information they are interested in is a challenge, no matter whether information is published in printed or electronic format. Some of the longer-standing ATI regimes have addressed this problem by requiring public bodies to publish registers of the documents that they hold. This is a well-developed practice in Sweden and Denmark for example. It is also a good practice which is recommended (although not mandatory) under the Council of Europe Convention on Access to Official Documents. In addition to requirements in access to information laws, many governments have built single

portals for accessing proactively published information. Such portals exist, for example, in Mexico and Hungary.[6]

For individual public bodies which are putting information on-line, they will know what information is being released and should therefore be able to present an index to the public. The challenge becomes more complex when creating an index of all government data across multiple departments. Technology now permits automated indexing,[7] but it is still hard to map out and to organise indices in a way which will make sense for someone searching for the information. The common solutions to this are Information Asset Registers and Data Catalogues, which are discussed in the following sections. In addition, use of appropriate tagging techniques helps make the information findable.

2.1.1 Information Asset Registers

Information Asset Registers (IARs) are registers specifically set up to capture and organise meta-data about the vast quantities of information held by government departments and agencies. A comprehensive IAR would include databases, old sets of files, recent electronic files, collections of statistics, research, etc.

For decades, governments have maintained lists of public information assets, from card catalogues to paper registers to online web portals. In the 1990s the US government began working on a project to improve the discoverability of official information, as part of a broader programme to encourage innovation in the information sector and to improve the "National Information Infrastructure".[8]

As a result, the Government Information Locator Service (GILS) was developed to help users identify, locate, and access publicly available US Federal information resources, including electronic information resources. The vision for GILS was one of a distributed system of access points to information held by different government agencies that would give the user comprehensive coverage of Federal government resources. During this period in the mid-1990s, the Australian and Canadian governments also began work on implementing GILS systems based on the same technology.

In the late 1990s, the UK government began work on a similar system to guide information-seekers through the maze of official information and materials by providing them with a single entry-point. The UK Information Asset Register (IAR) was announced in the White Paper on the Future Management of Crown Copyright in 1999. The service "inforoute" was created at the Office of Public Sector Information (OPSI). Inforoute provides direct access to the Government's Information Asset Register (IAR). The IAR lists information resources held by

[6] The Mexican Transparency Portal, run by the Information Commissioners' Office aims to ensure compliance with the proactive disclosure rules of the Federal Law on Transparency and Access to Information (2002) which requires that 17 classes of information be published proactively. Launched in early 2007, the Transparency Portal now holds millions of registers (including 1.3 million contracts and 1.6 million entries in the register for concessions, permissions, and authorisations). The Mexican Transparency Portal, in Spanish, can be found at http://portaltransparencia.gob.mx/pot/. In 2008 there were almost 14 million consultations (counting documents which were actually opened, rather than just hits on the website). The most popular information is the directory of public servants, the details of the salaries of public servants, the register of concessions and the register of contracts.

[7] There has been a dramatic change of technology since 1993 when Sir Tim Berners-Lee maintained the first systematic index of the internet in the form of a manually compiled index also known as the W3Catalog.

[8] See http://archive.ifla.org/documents/libraries/cataloging/metadata/gils.txt

the UK Government, concentrating on unpublished resources. In doing so it enables users to identify, from one single source, the information held in a wide variety of government departments, agencies and other organisations. Inforoute is a key part of the Government's agenda for freeing up access to information. Over a decade later asset registers are still an important part of the UK government's digital strategy although the focus has now changed to putting on line datasets that the public may wish to use.[9]

The EU's 2003 Directive on Reuse of Public Sector Information[10] also recognises the importance of asset registers for prospective re-users of public information, and requires that member states provide lists, portals, or something similar. It states:

> *Tools that help potential re-users to find documents available for re-use and the conditions for re-use can facilitate considerably the cross-border use of public sector documents. Member States should therefore ensure that practical arrangements are in place that help re-users in their search for documents available for reuse. Assets lists, accessible preferably online, of main documents (documents that are extensively re-used or that have the potential to be extensively re-used), and portal sites that are linked to decentralised assets lists are examples of such practical arrangements.*[11]

Information Asset Registers (IARs) can be developed in different ways. Government departments can develop their own IARs (an example is the UK's Department of Health[12]) and these can be linked to a national IARs (The UK's national IAR[13]). IARs can include information which is held by public bodies but which has not yet been – and maybe will not be – proactively published. Hence they allow members of the public to identify information which exists and which can be requested.

For open government data and access to information activists, it is important that any registers of information held be as complete as possible in order to be able to have confidence that documents can be found. The lack of completeness of some registers (such as the EU's documents register) is a significant problem as it creates a degree of unreliability which may discourage some from using the registers to search for information.

It is essential that the metadata (See Section 2.1.3) in the IARs be comprehensive so that search engines can function effectively. In the spirit of open government data, public bodies should make available their IARs to the general public as raw data under an open licence so that civic hackers can make use of the data, for example by building search engines and user interfaces.

➢ Recommendations:

---

[9] The role of Information Asset Registers, 10 September 2008 John Sheridan, Office of Public Sector Information, UK. http://www.eps

[10] EU PSI Directive: 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive/psi_directive_en.pdf

[11] EU PSI Directive, ibid

[12] The Department of Health Information Asset Register lists the information resources it holds with a focus on unpublished resources. It can be viewed at: http://www.info.doh.gov.uk/doh/iar.nsf?open

[13] Maintained by the Office of Public Sector Information, this IAR lists information resources held by the UK Government, concentrating on unpublished resources. According to OPSI, the IAR "enables users to identify, from one single source, the information held in a wide variety of departments, agencies and other organisations." See http://www.opsi.gov.uk/iar/index.

Governments and IGOs should take steps to make information more discoverable, namely they should:

- Consider creating Information Asset Registers where these do not already exist as a tool to help members of the public identify information they might want to access;
- Ensure that, where multiple IARs exist, a central portal or search function should guide users to these information resources;
- Make available the raw data from Information Asset Registers under an open licence so that others can reuse it.

2.1.2 Data catalogues

Open Government Data catalogues are web portals which offer access to government data sets. The aim is to make it easy for the public to find relevant government information in a single register instead of having to search various sites or use multiple search engines.

Data catalogues have existed in other sectors of society for many years; for example, in scientific communities, where there are many comprehensive data catalogues. The first government data catalogue was launched in October 2008 by the District of Columbia Government in the US. The positive experience of this first open government data catalogue is explained by Vivek Kundra, now President Obama's Federal Chief Information Officer:

> *When we first opened the doors to government data, people were quick to respond. Individuals and organizations are [now] not only viewing our government data, but are actually improving upon our work by analysing and repurposing the information in useful ways.*

An example of an initiative which made use of the DC Data Catalogue was EveryBlock.com which put the data into an online community news forum. Visitors can type in their zip code and find and exchange information about things of interest in their neighbourhoods such as reviews of local businesses, real estate listings, crimes, road construction, city service requests, community meetings, and more.[14]

The launch in May 2009 of the US Federal Government's data catalogue, data.gov brought data catalogues to mainstream public attention.[15] Its stated purpose is to "increase public access to high value, machine readable datasets generated by the Executive Branch of the Federal Government". The catalogue was launched with 47 datasets, and now holds over 270,000. It lists over 236 applications and services which have been developed for citizens, based on these released datasets.

The UK followed suit in January 2010 with the launch of data.gov.uk, which contained almost 2000 datasets.[16] From the start, the data.gov.uk team worked closely with the semantic web community to help ensure that official datasets were linked to a growing 'web of data'. The website is powered by CKAN (See Section 2.1.3 below) an open source data registry which is

[14] EveryBlock was originally funded by a two-year grant from the Knight Foundation (http://www.knightfdn.org/) through its Knight News Challenge program (http://www.newschallenge.org/). It is now wholly owned by msnbc.com (msnbc.com).

[15] Both data catalogues were built under the supervision of Vivek Kundra who was Chief Technology Officer in DC and then became Federal Chief Information Officer under Obama.

[16] See http://data.gov.uk/. This initiative was led by Tim Berners-Lee and involved close consultation with civil society, including representatives of MySociety and the Open Knowledge Foundation. It now forms a key part of the new governments Transparency Strategy, to be found at http://www.cabinetoffice.gov.uk/newsroom/statements/transparency/pm-letter.aspx

currently being used to develop over a dozen citizen driven data catalogues around the world. There is a growing community around the site, which currently features over 60 applications and services for citizens.

Data catalogues have now become the preferred format for presenting proactively disclosed data sets to the general public. The number of data catalogues is growing rapidly, and today there are around 30 data catalogues worldwide.[17]

Data catalogues can be created at different levels of government, ranging from the federal government level (for example, data.gov.au in Australia) to local authorities (for example the City of Edmonton in Canada has its own Open Data Catalogue, data.edmonton.ca). Inter-governmental bodies have also stared to launch their own data catalogues, including the UN[18] and the World Bank[19]

The number of data catalogues is not an issue, but it is essential that there be full "interoperability" between the data catalogues so that information can be exchanged or channelled through to a central portal if necessary. Similarly, data catalogues should be well structured so that datasets can be found easily, and the metadata should be comprehensive to facilitate searching. Ensuring the discoverability of information contained in data catalogues is essential as over them they will contain increasing numbers of datasets.

➢ Recommendations:

In order to make data discoverable, governments and IGOs should:

- Set up data catalogues by central, regional or local government linked to policies on the progressive release of datasets;
- Encourage and permit individual public bodies wishing to create their own data catalogues in countries where there is no central portal to do so;
- Ensure that data catalogues contain comprehensive information about the documents and datasets that the public can freely use.
- Ensure that the raw data contained in data catalogues is published under open licences so that people can use it.

2.1.3 Citizen-driven catalogues

In addition to data catalogues created by government, there are a growing number of citizen-driven catalogues. Often these are created by open government data advocates to map out what information is available and how open it is (both legally and technically).

Good examples of citizen-driven catalogues include:

- Datadotgc.ca[20] in Canada which includes nearly 700 datasets added by Canadian open government data advocates;

- The Offene Daten[21] project in Germany which contains over 300 datasets added by the Open Data Network, a German open government data organisation;

[17] Fact sheet on Open Government Data Catalogues: http://spreadsheets0.google.com/ccc?key=t71YHp9k6km8hqep3dUsbhA&hl=de&ui=2#gid=0
[18] See UN Data at http://data.un.org/
[19] See World Bank Data Catalogue at http://data.worldbank.org/data-catalog
[20] See http://www.datadotgc.ca/

- National Data Catalog[22] in the US being developed by the non-governmental organisation the Sunlight Foundation.

There are several other independent catalogue projects around the world, as listed in Annex 3.

An advantage of citizen-driven catalogues is that, while governments are generally in a much better position than citizens to **publish** official datasets, members of the public may have more innovative ideas on how to **present** these datasets to users. For example, public bodies may not want citizens to edit, comment on, or add to official registries, but web services with these features may be more desirable for user communities, especially given current efforts to link government data to other sources of information (e.g. from international sources, research bodies, and so on).

An example of the flexibility of citizen initiatives is the Comprehensive Knowledge Archive Network (CKAN), developed by the Open Knowledge Foundation. The CKAN software, developed through a large-scale collaborative open source software project, was originally developed to create a register where it is possible for anyone to find data that is free to use in order to, for example, build new kinds of applications and services. The CKAN software is now being used to run data.gov.uk as well as citizen catalogues such as Offene Daten in Germany.

In addition, the CKAN register itself contains information from governments as well as "open knowledge" sets from other sources such as science and the arts. Anyone can add to or edit information in the CKAN register. This means, for example, that users can download datasets, convert them into easy-to-use formats, and reload them to the register. Work is currently underway on enhancing functionality, which will allow users to work with datasets in increasingly sophisticated ways such as automating certain tasks and improving the tools for the analysis and visual representation of datasets.

➢ Recommendations:

- Members of the public are encouraged to set up open data catalogues, particularly where there is no suitable government catalogue or where they have fresh ideas about how to present the contents of the catalogue;
- Citizen-driven catalogues can undertake projects which governments are unlikely to do, such as linking to data catalogues in other countries, providing translation of the metadata where necessary.

2.1.4 Technical Dimensions of Discoverability

In addition to having catalogues or registers of government data, the formats that information is published in can make a difference to how easy it is to find, or how "discoverable" it is.

Paying attention to the technical issues related to discoverability can ensure that information can be found easily and that it remains easy to find in the long term. There are best practice standards which have been developed to guide public bodies on how to do this. The guidance developed by the W3C eGov Interest Group was developed "to encourage publication of government data, allowing the public to use this data in new and innovative ways" and it suggests best practices which public authorities are encouraged to follow.[23]

---

[21] See http://offenedaten.de/

[22] See http://nationaldatacatalog.com/

[23] These guidelines have been in development since 2004 with the latest version dating from 2009; they are still formally in draft form. http://www.w3.org/TR/gov-data/

One challenge for governments is to ensure that information published on the web remains accessible years after it is been published. Broken and outdated links can frustrate and disillusion the public. One study found that in 2006 only 48% of the links from the official parliamentary record of the UK, called Hansard, to other government websites were working.[24] There is a need to invest in long-term storage and effective stewardship of links, making full use of technical solutions such as permanent URLs and Uniform Resource Identifiers (URIs) to maintain access to information over time.

There are a number of format and tagging issues which make information easier to process with machines and hence make it easier both to locate and to reuse. One way to make data easier to locate is the use of "**metadata**", which is often defined as "data about data". Metadata standards exist for many areas of activity, including archiving, the arts, biology, education, geographic data, government, social sciences, linguistics, libraries, media, and science. The commonly agreed standards for these fields allow data to be classified in a way that makes them easier to describe, locate, retrieve and manage.[25]

An example of the benefits of linking metadata and search engines to make government data more "discoverable" comes from the UK where in 2004 the government launched "DirectGov", a single portal for citizens to access government services. The operational budget was £4.4 million. In 2005, open government data activists created "Directionlessgov.com", which was a simple web page that linked to the Google search engine and produced comparable search results to the DirectGov search engine. This alternative was built in less than an hour, making the point that the government might have been better to invest in the optimisation of the text and titles on its websites and on the capacity of its search engines.

Another way of making information more findable is through the use of "**microformats**" embedded within the web coding (HTML code[26]) so that information intended for end-users (such as contact information, geographic coordinates, calendar events, and the like) can be automatically processed by software. As a result, software applications that collect data about on-line resources, such as web crawlers, or desktop applications such as e-mail clients or scheduling software, can reference the data. Microformats can also be used to facilitate "mash ups" such as exporting all of the geographical locations on a web page into Google Maps, to visualize them spatially.

➢ Recommendations:

It is recommended that governments and IGOs should

- Take measures to ensure that information which is published on the internet remains accessible over the long term;
- Publish data using metadata and microformats to ensure that they are indexable by search engines.

[24] See UK Government Web Continuity: Persisting Access through Aligning Infrastructures by Amanda Spencer, John Sheridan, David Thomas, published by the National Archives (UK), to be found at: http://www.ijdc.net/index.php/ijdc/article/view/106/81

[25] See Wikipedia on Metadata: http://en.wikipedia.org/wiki/Metadata_standards

[26] This is a web-based approach to semantic mark-up that re-uses existing HTML/XHTML tags to convey metadata and other attributes, in web pages and other contexts that support (X)HTML, such as RSS.

2.2 Reusability: Technical Aspects of Openness

For information to have maximum value in terms of reusability, documents and datasets should be available in technical formats which can be read, processed and manipulated using computers. This section addresses the technical issues relating to the release and reusability of government data.

2.2.1 Electronic formats and access to databases

Government information was traditionally held as paper documents but is increasingly held as electronic documents and in databases. Some information will be recorded in other formats, such as audio-visual formats as photographs, film footage, microfiche, etc.

This means that the information can be made available in a variety of ways. If the information is only available in hard copy, then receiving paper copies is one option, although it is now possible to scan such documents into PDF, JPG or other electronic formats. For the majority of current information which is held in digital format, it can be made available to the public by posting it on government websites or sending it by e-mail to those who request it.

The availability of information in electronic format usually permits users to analyse, process and re-present the information with significantly greater ease than if the information is available only in hard copy (See Section 2.2.2 for analysis of which electronic formats best facilitate processing of electronic material).

As noted in Section , the right of access to information should apply to all information and hence should apply to databases. The definition of "information" in most access to information laws confirms this, as it typically refers to "all information recorded in any format" which should include databases. However, there is often not an explicit reference to a right of access to databases, except for in the laws of Finland and Norway, which do expressly permit access to databases. On the other hand, in Sweden such access is provided, but only in printed format, while in the Netherlands and Denmark databases are specifically excluded from the scope of the law. This is a problem predominantly with older access to information laws. In the majority of countries where there is not specific exclusion for databases, access to information and open government data advocates can use the wording of the national access to information law to argue that the right applies to databases. This is an area where the two communities need to work together to achieve greater legal clarity.

It is also not yet completely clear whether or not the right of access to information includes a right of access to information in electronic format, although the Council of Europe Convention on Access to Official Documents does establish a right of access to information in any form or format where this would be reasonable:

> *When access to an official document is granted, the applicant has the right to choose whether to inspect the original or a copy, or to receive a copy of it in any available form or format of his or her choice unless the preference expressed is unreasonable.(Article 6.1)*

In many countries, the law provides this right of access to documents in electronic format, as the Table 1 shows.

Table 1: Electronic Formats and the Right of Access to Databases

Country	Access in electronic format	Access to databases
Albania	Yes, format option	Not mentioned in the FOIA
Armenia	Yes	Not mentioned in the FOIA
Australia	Yes, format option	Only with relation to data protection
Austria	Not mentioned in the FOIA	Not mentioned in the FOIA
Belgium	Not mentioned in the FOIA	Not mentioned in the FOIA
Bosnia	Yes, format option	Not mentioned in the FOIA
Bulgaria	Yes	Not mentioned in the FOIA
Canada	Not mentioned in the FOIA	Not mentioned in the FOIA◎
Croatia	Not mentioned in the FOIA	Not mentioned in the FOIA
Denmark	Not mentioned in the FOIA	Specifically excluded
Estonia	Yes	Not mentioned in the FOIA
EU	Yes	Not mentioned in the FOIA
Finland	Yes	Yes
France	Yes	Not mentioned in the FOIA
Georgia	Yes, format option	Not mentioned in the FOIA
Germany	Yes	Not mentioned in the FOIA
Greece	Not mentioned in the FOIA	Not mentioned in the FOIA
Hungary	Yes	Only with relation to data protection
Ireland	Yes	Yes
Italy	Not mentioned in the FOIA	Not mentioned in the FOIA
Kosovo	Yes	Not mentioned in the FOIA
Latvia	Yes, format option	Not mentioned in the FOIA
Macedonia	Yes	Not mentioned in the FOIA
Moldova	Yes	Not mentioned in the FOIA
Montenegro	Yes	Not mentioned in the FOIA
Netherland	Yes	Specifically excluded
New Zealand	Yes	Yes
Norway	Yes	[not clear]
Poland	Yes, format option	Not mentioned in the FOIA
Portugal	Yes	Not mentioned in the FOIA
Romania	Not mentioned in the FOIA	Not mentioned in the FOIA
Serbia	Yes	Not mentioned in the FOIA
Slovakia	Yes	Not mentioned in the FOIA
Slovenia	Yes	Not mentioned in the FOIA
Switzerland	Yes	Not mentioned in the FOIA
Sweden	Yes	Only access in printed format
UK	Yes	Not mentioned in the FOIA
US	Yes	Not mentioned in the FOIA

Legend

◎ Not mentioned in the FOIA but national experts report that in databases can be accessed in practice

Yes = access to electronic format specifically mentioned

Yes, format option = access to electronic format is not specifically mentioned but you have the option of choosing how you want to access the information.

It is positive that the majority of access to information laws surveyed for this report made specific reference to the right of requestors to state a preference for access to information in electronic format, which is in line with the standard set by the Council of Europe Convention on Access to Official Documents. Both the international and national standards, however, limit this right of access to pre-existing electronic formats and place no obligation on public authorities to digitalise information.

While it is often possible to get access to information in electronic format, and this is becoming increasingly common in practice even if not anticipated in all access to information laws, there are instances when there has been resistance to disclosing information in this way and activists have had to resort to other means to compile the information, or have had to lobby to get the full data sets released electronically, for example:

- **Parliamentary Lobbying in Sweden**. When television journalists in Sweden requested copies of the parliamentary records of who visited the parliament each day, they were first given the information in electronic spreadsheets. When it became apparent that this information was being used to track lobbying activity, officials switched to providing paper records (which is all they have to do under Swedish law). They also informed the journalist that the information would henceforth have to be requested on a daily basis. In spite of this, the journalist continued to file regular requests for the information, later converting it into a database.[27] The issue of a right of access to documents in electronic format therefore remains unresolved in Sweden.

- **UK government spending data**. In February 2010 members of the *Where Does My Money Go?* project requested information about the financial spending database of the UK government's Treasury office. The "COINS" database contains fine-grained information about government financial transactions and is used as the basis for many important financial reports which the civil society activists used in order to enable members of the public to explore and visually represent UK public spending. In March 2010 the official response to the request for this financial database was a 345 page hard copy document containing 14,000 lines of information from the database. After a significant advocacy campaign – and an election which brought to power a government which had committed to the "right to data", the information was eventually released in an electronic, machine-readable format on 4 June 2010.[28]

➢ Recommendations:

In order to ensure maximum reusability of information, governments and IGOs should:

- Ensure that information is made available in electronic formats whenever and wherever it already exists in those formats;
- Incorporate a right of access to information in electronic format wherever possible into all access to information laws;

[27] Presentation by Helena Bengtsson at the Legal Leaks project meeting, Berlin, 18 February 2010. See www.legalleaks.info

[28] Some users of the data have commented on technical issues of access and the difficulty of doing something with the data without expert knowledge pointing to a wider problem of what happens once complex data sets are released.

- Progressively convert key information which they hold but which is not yet digitised into a digital format.

2.2.2 Machine readable formats

Information which is "machine readable" broadly means that it can be processed by a computer (or other automatic device or machine in the case of older, non-digital formats such as punch cards). For the open government data community the term "machine readable" is a narrower definition which refers to electronic formats which allow the data they contain to be automatically processed and transferred from one software programme to another.

An example would be ensuring that budget data is released in a format which can be imported into a database for accounting analysis. Releasing such information in PDF format, for example, would not permit this.

To maximise the benefits of opening up government data, it is essential that government information (whether text and/or numerical data) be released in formats which can be processed automatically by software programmes without requiring hours of manual work by those who wish to reuse the information.

There are a number of electronic formats which limit such automatic processing because the data cannot be extracted or cannot easily be extracted from these documents for further processing.

At one end of the scale are file formats from which it is not possible to extract any textual or numerical material. These include some information contained in PDF (Portable Document Format), as well as some other scanned formats such as the photographic formats JPG and TIF. This category also includes the use of Flash formats for information on websites as these are not picked up by search engines and text often cannot be extracted from them.

With respect to PDF documents, there are essentially two kinds of PDF documents made available by governments in response to access to information requests or published on websites:

1. **PDF Documents containing an image which is not machine readable**. These documents are created in two ways:
    - The PDF document is a scanned version of a document which was not originally machine readable, such as hand-written documents or those prepared on typewriters. These are generally older documents which were never in electronic format and have been scanned. In this case it is usually not be possible to use a computer to search through the information looking for keywords or data. (The same applies to older documents which have been scanned into other formats such as JPG and TIF). One solution is to re-enter the data into a computer when it is important that it be in a machine-readable format; another solution is to use optical character recognition technologies which, although imperfect, are rapidly improving and can make it possible at least to extract text from the scanned documents.
    - PDF documents which have been produced from documents that were originally prepared in electronic format but have then been printed and scanned. This is often done, for example, when it is necessary to add a signature to the document. It is relatively common to receive answers to information requests from the European Union in a scanned format (either PDF or TIF). The use of

such formats has proved a significant burden for information reusers; sometimes it is not even possible to cut and paste text from the documents or to click on links. The need to add signatures to some documents in order to make them have legal value is understandable, but given how much correspondence takes place by e-mail these days, the absolute necessity of printing and rescanning of some letters to information-seekers should be given careful consideration.[29]

2. **Electronic Documents Converted to PDFs**: Documents which have been converted into PDF or another format directly by the computer. The former are often not machine-readable and the latter often allow only for limited use: it may be possible to search for key words but cutting and pasting material to put it into another software programme may lose much of the formatting and significantly reduce the usability of the information.

   For example**:** EU Cohesion Fund data is released in many countries in a PDF format which means that the budgetary information cannot be processed on computers using database or spreadsheet software. Hence it is not possible to aggregate the different data sources to provide an overview of how the money was spent – unless, that is, the data from the PDF documents is manually re-entered or computer programmers construct elaborate scripts to extract the data. Open data advocates who identified this problem at the May 2009 EU Open Data Summit in Brussels then submitted several access to information requests in order to obtain the information in machine-readable form. The responses included that the PDF files were already published on their websites: many ATI laws specifically exclude from the right of access information which has already been published, requiring only that the requestor is referred to the public source of the information. One document from the Hungarian government contained 10,000 pages of budgetary information in PDF format. Programmers had to spend considerable time extracting the data from this file.

Civic society developers and others seeking to reuse government information have developed various methods to extract structured data (or information in a database-like form that a computer can read) from sources which are more or less unstructured (such as government websites, PDF documents, and scanned documents). This involves identifying patterns in the unstructured sources (such as columns and rows in a budget document) and writing a computer program to reconstruct the underlying data sets on the basis of these patterns. This process, known as 'screen scraping' can be time-consuming and may often require a degree of technical ingenuity.

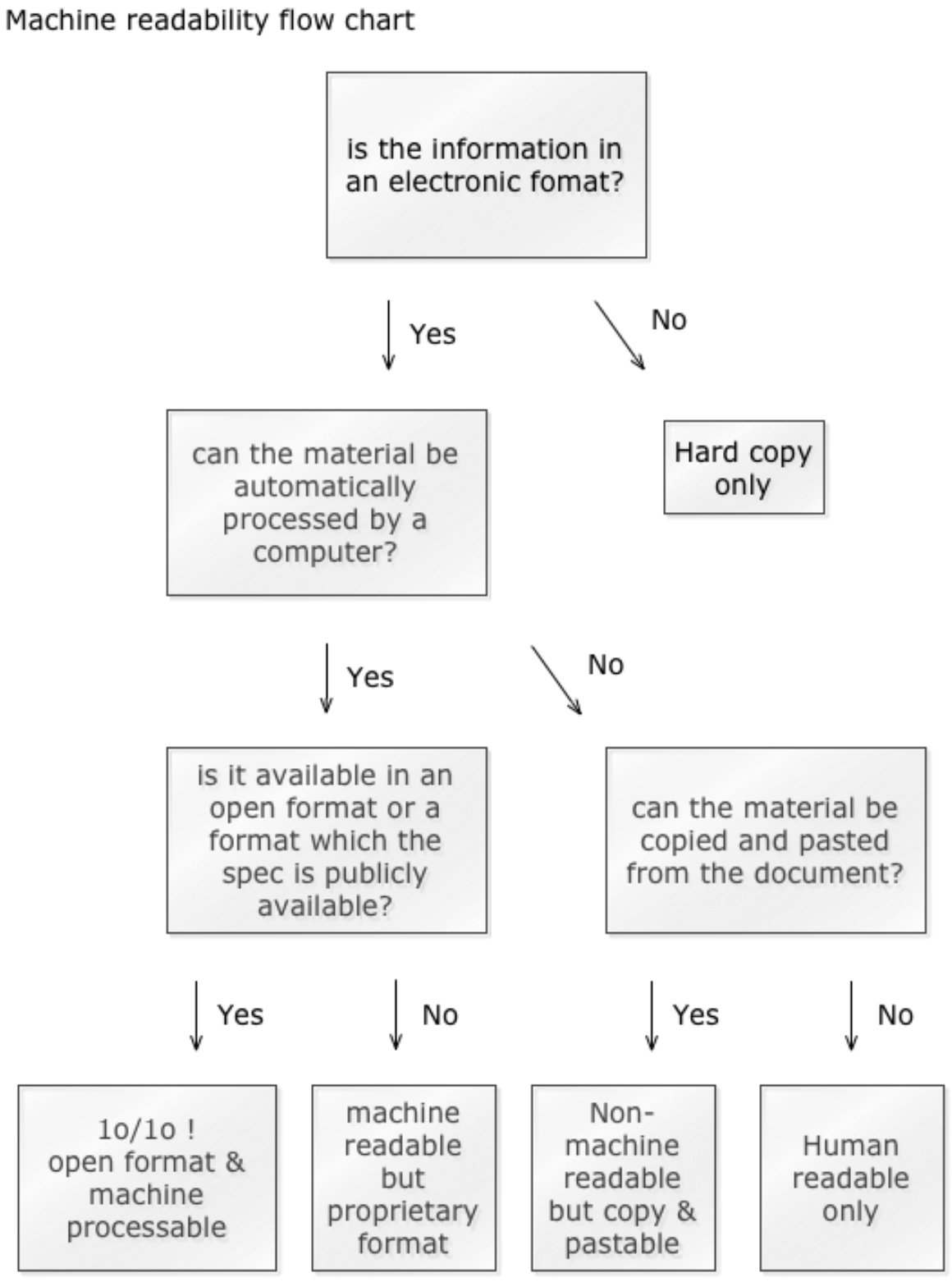
There are several projects that aim to make screen scraping more accessible to people without a technical background. An example is the ScraperWiki project.

- **ScraperWiki** is a web platform for collecting and publishing public data. It was started in 2008 by civic web developer Julian Todd, who has worked on various projects aiming to open up the activities of official bodies, including They Work for You, Public Whip, and UNdemocracy. ScraperWiki allows users to extract useful structured data from a wide

[29] Law and policy should balance the need to deliver documents that contain a signature and the interests of the user in having information that can be reused. Such solutions might include use of electronic signatures, use of scanned signatures embedded in files as an image, or discarding the signature requirement for most information sent by e-mail (unless a signed document is specifically requested). If a signature is absolutely necessary, a practical solution is to send two copies of documents, one with a signature that can be used for any legal processes (such as an appeal against a refusal), and one in fully machine-readable version without the signature from which information can be extracted.

> variety of different kinds of online sources, from non-machine-readable PDF files, web pages, and web interfaces to large databases. ScraperWiki aims to make it easier for users to collaborate on the creation and maintenance of screen-scraping scripts, thus helping to ensure that clever pieces of code are shared and that scraped data is accurate and up-to-date.

While projects such as ScraperWiki are extremely valuable, in most situations it would be far preferable for official bodies to release the data underlying documents, websites, and other non-machine-readable sources, rather than leaving the public to attempt to reconstruct this using a process which can often be laborious and fallible.



There are different reasons why public officials might be reluctant to make information available in non-PDF formats. One reason is that there are some advantages to publishing information in PDF format.[30] These include making available and searchable pre-electronic documents, as well as the possibility of including digital signatures in documents that need to be used in official and legal processes. Another reason is suggested by Jack Thurston, from the Farmsubsidy.org project, who has fought various battles over the right of access to information in machine-readable formats with various EU governments. Thurston conjectures that in benign cases government representatives consider PDF files to be a bit like paper: they are not easy to change. Hence the data cannot be tampered with or modified and, to a certain degree, the data is fixed. He suggests that in more malicious cases some officials used PDF files in order to technically obstruct third parties from analysing and representing the data contained within them.

In spite of these concerns, a right to access information in machine-readable format wherever possible is not anticipated in any of the access to information laws reviewed for this report. This oversight can perhaps best be explained by the fact that the laws were mostly written before there was a clear understanding of these issues and the distinction between different types of electronic access. This issue has not yet been addressed by the ATI community but there has been advocacy by the OGD community, for example in the US by the Sunlight Foundation:

- **Sunlight Foundation Campaign** In the United States many draft laws are published as PDFs and some government websites present other material in PDF and Flash formats. In response to this, the Sunlight Foundation called in December 2009 for a change of practice: *"Here at Sunlight we want the government to STOP publishing bills and data in PDFs and Flash and start publishing them in open, machine readable formats like XML and XSLT. What's most frustrating is, Government seems to transform documents that are in XML into PDF to release them to the public, thinking that that's a good thing for citizens. Government: We can turn XML into PDFs. We can't turn PDFs into XML."*[31]

Responding to this demand, governments have adopted policies explicitly calling for public bodies to make information available in machine-readable formats. For example in December 2009 the UK government published a set of "public data principles" as part of the "Smarter Government" initiative.[32] One of these principles states: "Public data will be published in reusable, machine-readable form". Similarly, the US Open Government Directive, also published in December 2009, explicitly states that Executive Departments and Agencies should take steps to make information available in machine-readable formats. The new New Zealand policy launched in July 2010 specifically requires public bodies to "release information in formats which make the data easy to use, taking into account the wishes of likely users."

➢ Recommendations:

It is recommended that governments and IGOs should:

- Make information available in machine-readable formats wherever possible and in all the cases in which it already exists in these formats.

[30] For some of the arguments see, for example, *The good and the bad of PDFs* http://wiki.opengovdata.org/index.php?title=The_good_and_the_bad_of_PDFs

[31] Adobe is Bad for Open Government http://sunlightlabs.com/blog/2009/adobe-bad-open-government/

[32] See: http://blog.okfn.org/2009/12/07/uk-government-announces-lots-of-new-open-data/

- Avoid exclusive use of PDFs to deliver government information and develop policies to that effect. A good practice which is increasingly common is to publish documents on websites in multiple formats, including HTML, Open Office/Word and PDF.
- Make clear that public institutions have an obligation to facilitate access to information in machine-readable formats whenever possible. This can be achieved through access to information laws, regulations, and/or government policy documents;
- Adopt policies and practical solutions which make it possible to deliver documents in fully machine readable formats by, for example, eliminating the need for hand written signatures in the majority of cases of correspondence with members of the public.

2.2.3 Open File Formats

Even if information is provided in electronic, machine-readable format, and in detail, there may be issues relating to the format of the file itself.

The formats in which information is published – in other words, the digital base in which the information is stored - can either be "open" or "closed". An open format is one where the specifications for the software are available to anyone, free of charge, so that anyone can use these specifications in their own software without any limitations on reuse imposed by intellectual property rights.[33]

If a file format is "closed", this may be either because the file format is proprietary and the specification is not publicly available, or because the file format is proprietary and even though the specification has been made public, reuse is limited. If information is released in a closed file format, this can cause significant obstacles to reusing the information encoded in it, forcing those who wish to use the information to buy the necessary software.

[33] See http://en.wikipedia.org/wiki/Open_format. There is in fact no single definition of an open format but the definition used here captures the core elements. Other aspects include that the process of developing the standard should have been open and the standard will be maintained by a non-governmental body committed to updating it over time. The level of detail in the documentation of the specification is also an important criteria, as is the fact that the documentation is available free of charge.

Table 2: The "openness" of commonly used formats

File format	Information /Text can be extracted?	Fully machine processable = "machine readable"	Specification available?	Open Format?
Plain Text (.txt)	✔	✔	✔	✔
Comma Separated Value (.csv/.txt)	✔	✔	✔	✔
Hyper Text Markup Language (.html/.htm)	✔	✔	✔	✔
Extensible Markup Language (.xml)	✔	✔	✔	✔
Resource Description Framework (.rdf)	✔	✔	✔	✔
Open Document Format (.odt, .ods, etc)	✔	✘	✔	✔
Microsoft Word (.doc/.docx)	✔	✘	✔	✘
Microsoft Excel (.xls/.xlsx)	✔	✔	✔	✘
Portable Document Format (.pdf)	✘	✘	✔	✔
Image files (.jpg, .tif)	✘	✘	✔	✔

The benefit of open file formats is that they permit developers to produce multiple software packages and services using these formats. This then minimises the obstacles to reusing the information they contain.

Using proprietary file formats for which the specification is not publicly available can create dependence on third-party software or file format license holders. In worst-case scenarios this can mean that information can only be read using certain software packages, which can be prohibitively expensive, or which may become obsolete.

The preference from the open government data perspective therefore is that information be released in **open file formats which are machine-readable**.

> **Example**: **UK traffic data**. Andrew Nicolson is a software developer who was involved in an (ultimately successful) campaign against the construction of a new road, the Westbury Eastern bypass, in the UK. Andrew was interested in accessing and using the road traffic data that was being used to justify the proposals. He managed to obtain some of the relevant data via freedom of information requests, but the local government provided the data in a proprietary format which can only be read using software produced by a company called Saturn, who specialise in traffic modelling and forecasting. There is no provision for a "read only" version of the software, so Andrew's group had no choice but to purchase a software license, eventually paying £500 (€600)

> when making use of an educational discount. The main software packages on the April 2010 price list from Saturn start at £13,000 (over €15,000), a price which is most likely beyond the reach of most ordinary citizens.[34]

Although no access to information law gives a right of access to information in open formats, open government data initiatives are starting to be accompanied by policy documents which stipulate that official information must be made available in open file formats. Setting the gold standard has been the Obama Administration, with the Open Government Directive issued in December 2009, which says that:

> *To the extent practicable and subject to valid restrictions, agencies should publish information online in an open format that can be retrieved, downloaded, indexed, and searched by commonly used web search applications. An open format is one that is platform independent, machine readable, and made available to the public without restrictions that would impede the re-use of that information.*[35]

➢ Recommendations:

It is recommended that governments and IGOs should:

- Require that wherever possible and on a progressive basis, information released under access to information laws should be in an open file format or at least in a format for which the specification is publicly available;

- Make provisions for progressively publishing information which is currently stored in non-machine-readable or proprietary formats in machine-readable file formats which are open, or at least in a format for which the specification is publicly available.

2.3 Enhancing Accessibility: Complete, Raw, Timely Data

There are a number of practical steps that governments can take to increase the openness and the utility of the data that is released. These include ensuring that data can be downloaded in bulk, as well as releasing it as rapidly and in the maximum level of detail possible following its collection. Government data will also be more meaningful and useful for members of the public if it is published in formats which add value to it, such as by being able to link it to other data. This section examines how public bodies can ensure the release of complete, detailed, timely, and meaningful information.

2.3.1 Raw Data Now!

"Raw data now" has become the rallying cry for this call for a shift in the way information is released to the public.[36] The open government data movement calls for information to be

---

[34] See http://www.saturnsoftware.co.uk/, price list on file with the authors.

[35] US Open Government Directive, 8 December 2009, http://www.whitehouse.gov/omb/assets/memoranda_2010/m10-06.pdf

[36] In November 2007 OKF Director Dr. Rufus Pollock wrote a blog post titled "Give Us the Data Raw, and Give it to Us Now". The blog entry concludes "We want the data raw, and we want the data now." In February 2009 Sir Tim Berners-Lee addressed technology conference TED 2009 calling on public bodies to open up raw data sets for others to use and asking the crowd to chant with him "Raw Data Now!", alluding to Rufus Pollock's 2007 meme. See http://www.ted.com/index.php/talks/tim_berners_lee_on_the_next_web.html. The "raw data now" demand has subsequently gained traction in various communities of open data advocates, technologists, journalists and others, and was widely covered in the technology press throughout the world.

made available in the most disaggregated form available. In practice this means providing access to the same raw data as public institutions have at their disposal.

The Open Knowledge Foundation has been campaigning for public bodies to move away from the tendency to present information with a "shiny interface" and to give direct access to the underlying "raw" data sets. An example is the set of indicators used to measure progress towards the Millennium Development Goals, which was presented in a way that looked good on the website but which meant that data could not be download in bulk because the web interface only allowed users to query the data but not to actually access the whole database in order to analyse it and represent it for themselves. Civil society activists "scraped" this data to create a more usable data set, which resulted in discussions with the UN about improving the way in which it released the data. [37]

The Open Knowledge Foundation argues that fancy front ends to official databases can often restrict what users can do with the data, can be expensive, tend to date rapidly, focus on the interface rather than the underlying data, and can delay access to the data while they are being built.

Permitting bulk downloads requires public authorities to recognise that, increasingly, information-seekers are interested in accessing not just one or two pieces of government data but entire databases. This is something which is at the heart of the open government data movement because there are programmers keen to develop applications which make the most of the publicly generated data, and who are requesting full datasets.

This is not new, however, and not even a feature unique to the digital era: civil society organisations and journalists working to monitor government performance and track public spending have often wanted to get access to complete sets of information, such as full budget sending data.

Nevertheless, the significant change with digitisation is that an ever-increasing proportion of government data is held in databases as opposed to being stored in discrete documents. This means that the traditional bureaucratic concept of access to completed "documents" no longer holds much validity. This is not a significant problem in countries whose laws are "access to information" laws, but does pose some challenges in countries where an "access to documents" paradigm still prevails and hence where databases are not defined as documents, or even explicitly excluded from the right of access, such as is the case in the Netherlands and Denmark.

As noted in the Section on the right of access to information, members of the public should, in principle, be provided with entire databases when they request it, providing that none of the content falls under a legitimate exception. If some of the data falls under an exception, then partial access should be provided to the remainder of the database.

One problem which has arisen in practice is that public bodies post information online in its entirety, but not in a format that enables the entire database to be downloaded. Rather, a common practice is to let users search for a limited number of records. Although the idea of a searchable database interface is often intended to assist members of the public in searching for information, such systems in fact limit rapid access to larger quantities of information. Examples of the challenges faced by reusers of the information include:

[37] See the UN website at http://unstats.un.org/unsd/mdg/Default.aspx and the revised presentation by the Open Economics project, at http://www.openeconomics.net/mdg/.

- **EU Farm subsidy data.** When trying to collect information from national government sources, the Farm Subsidy project found that in many cases information was only available via a web interface, and was not available to download in bulk. For example the French, German, and Dutch websites only allowed users to query the data on their websites. This made it impossible to get a "bigger picture" of where the money goes, for example by establishing which large international companies receive in total across different EU member states. Some websites, such as Ireland's data portal, were particularly difficult to extract data from. Members of the project said that representatives from the Dutch government admitted that extracting data in bulk was deliberately difficult as they did not want people to have access to the full database.[38]

- **NHS Choices data.** Civic web developer Sym Roe was interested in creating a web application that would help people to locate their closest accident and emergency department. He decided to use the NHS Choices API, an interface which allows users to provide a hospital name and returns an XML feed of the coordinates of the hospital, its postcode, what services are offered, whether there is an accident and emergency department and so on.

  However, despite the fact that the information was available in a useful machine-readable format, he found that there was no way of getting a list of this data for every hospital. In order to get around this he had to write a piece of code which searched for a space character in either the address or postcode, which ultimately enabled him to laboriously reconstruct a list of every hospital. He thinks that the NHS API is a very good start but is flawed because it only aims to solve one problem (finding the nearest accident and emergency department) and would much prefer if the full database were made available for download to permit other uses of it to be developed.[39]

- **Information on Planning Applications**. After the unexpected demolition of a local pub, Richard Pope was interested in creating a service to automatically notify UK citizens of planning applications near them. Unfortunately he found that this information was scattered across many different government website in many different formats. There was no single database of planning applications, so he had to write a script to scrape every single different council website in order to republish the information in a common format. According to Pope, this situation repeats itself with lots of other types of local data.[40]

The recommended solution is to ensure that the entire database is available for downloading. This does not mean that interactive interfaces which allow users to search on the public body's website for a specific piece of information should be removed or abolished. Rather it means that both forms of access should be provided.

➢ Recommendations

It is recommended that governments and IGOs should:

- Ensure that datasets which are published on interactive web interfaces are also available for downloading in bulk;

[38] Interview by authors with Jack Thurston of farmsubsidy.org, April 2010
[39] Interview by authors with Sym Roe of farmsubsidy.org, April 2010
[40] Interview by authors with Richard Pope, April 2010

- Make data available in the same disaggregated format as is accessible to public officials, subject to legitimate exceptions;
- Ensure that, in cases where certain data fields need to be removed because the data falls under exceptions in the access to information law, access is provided to the remainder of the data.

2.3.2 Up-to-date Data

For members of the public to be able to take advantage of the full value of direct access to information held by public authorities, it is essential that the information be published in a timely fashion.

Whether the data is traffic reports, crime statistics, public works contracts, or educational services, it is essential that builders of applications have access to the very latest data.

To do this, information should be compiled in ways that facilitate rapid disclosure. The Washington DC Data Catalogue has a number of live feeds for regularly updated material.[41]

A potential obstacle to rapid disclosure is the need to apply exceptions to documents or datasets before releasing them. When information is stored electronically, it is possible to anticipate disclosure when building databases. For example, if a register of information containing the names of private individuals is to be made public, one solution is to build the database in a way that permits officials to see the names while members of the public can see the remainder of the information but not identify the individuals, thereby respecting personal data protection rules.

The need to ensure that released data is up-to-date is now becoming part of government disclosure policies. An Australian government task force on "Government 2.0" recommended that public sector information "should be released as early as practicable and regularly updated to ensure its currency is maintained."[42]

➢ Recommendations

It is recommended that governments and IGOs should:

- Make provisions for informing potential users of updates to published datasets, for example by use of RSS feeds;
- Design datasets with disclosure in mind, anticipating which data may need to be exempted, in order that disclosure of the remainder can be as rapid as possible.

2.3.3 Linked Data and the Semantic Web

The World Wide Web is a vast array of resources, documents, data, and services which are communicated via the internet, linked by hyperlinks and URLs.

The web makes it possible to connect related documents and data by using different linking technologies. This "linked data" makes the web more meaningful and more useful.

[41] See http://data.octo.dc.gov/

[42] See On 3 May 2010, the Minister for Finance and Deregulation released the Government Response to the report of the Government 2.0 Taskforce, at http://www.finance.gov.au/publications/govresponse20report/index.html#recommendation-06

Examples of key technologies that support linked data are Uniform Resource Identifiers (URIs - a generic means to identify entities or concepts in the world); HTTP (a simple yet universal mechanism for retrieving resources, or descriptions of resources); and Resource Description Framework (RDF - a generic graph-based data model which is used to structure and link data that describes things in the world).[43]

These formats enable computers to analyse and "make sense of" data sets. So while an encyclopaedia might contain an ordinary sentence such as "Paris is the capital of France" which means nothing to a computer, an equivalent statement in RDF could express that "Paris" is the name of a capital city, "France" is the name of a country, and that the first "is a capital of" the second. Hence, RDF allows for the creation of structured relationships between entities that computers can parse and query – rather than unstructured text that the computer can do much less with.

As a concrete example, projects such as DBPedia[44] attempt to extract structured information from Wikipedia to allow users to make sophisticated queries such as:

> [...] soccer players with number 11 (on their jersey), who play in a club whose stadium has a capacity of more than 40000 people and were born in a country with more than 10 million inhabitants.[45]

The ability for machines (computers) to extract meaning from the information on the internet has led to the term "semantic web", used to describe the processing of the meaning - or "semantics" - of information.

Many proponents and users of semantic web technologies are interested in the longer term potential of a vast "web of data". Connecting together many different semantic web data sets would allow increasingly more sophisticated queries to be made, thus enabling a new generation of smarter web applications and services.[46] The move towards greater linking of open data holds significant potential for making information which is already on the web more meaningful and accessible. The possibility of linking to government information is also an important part of the future potential of the semantic web.

There are several **government initiatives** involving linked open data. For example, the US national data catalogue, data.gov, publishes a variety of datasets in RDF[47], as does the UK national catalogue, data.gov.uk. Both catalogues have an interface[48] which allows users to run

---

[43] See Linked Data, http://linkeddata.org/faq

[44] DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia, and to link other data sets on the Web to Wikipedia data. The goal is to make information in Wikipedia useable in new and interesting ways which might inspire new mechanisms for navigating, linking and improving the encyclopaedia itself. See www.dbpedia.org.

[45] Sören Auer, Jens Lehmann, What have Innsbruck and Leipzig in common? Extracting Semantics from Wiki Content (PDF), p.11. In Franconi et al. (eds), Proceedings of 4th European Semantic Web Conference, ESWC 2007, Innsbruck, Austria, June 3-7, 2007, LNCS 4519, pp. 503–517, ISBN 978-3-540-72666-1, Springer, 2007, at www.informatik.uni-leipzig.de/~auer/publication/ExtractingSemantics.pdf.

[46] The World Wide Web Consortium's Linking Open Data initiative, driven by a loose knit community of technologists, is currently undertaking to convert many large open and freely available datasets to formats such as RDF, linking these together and creating interfaces through which they can be queried. See http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

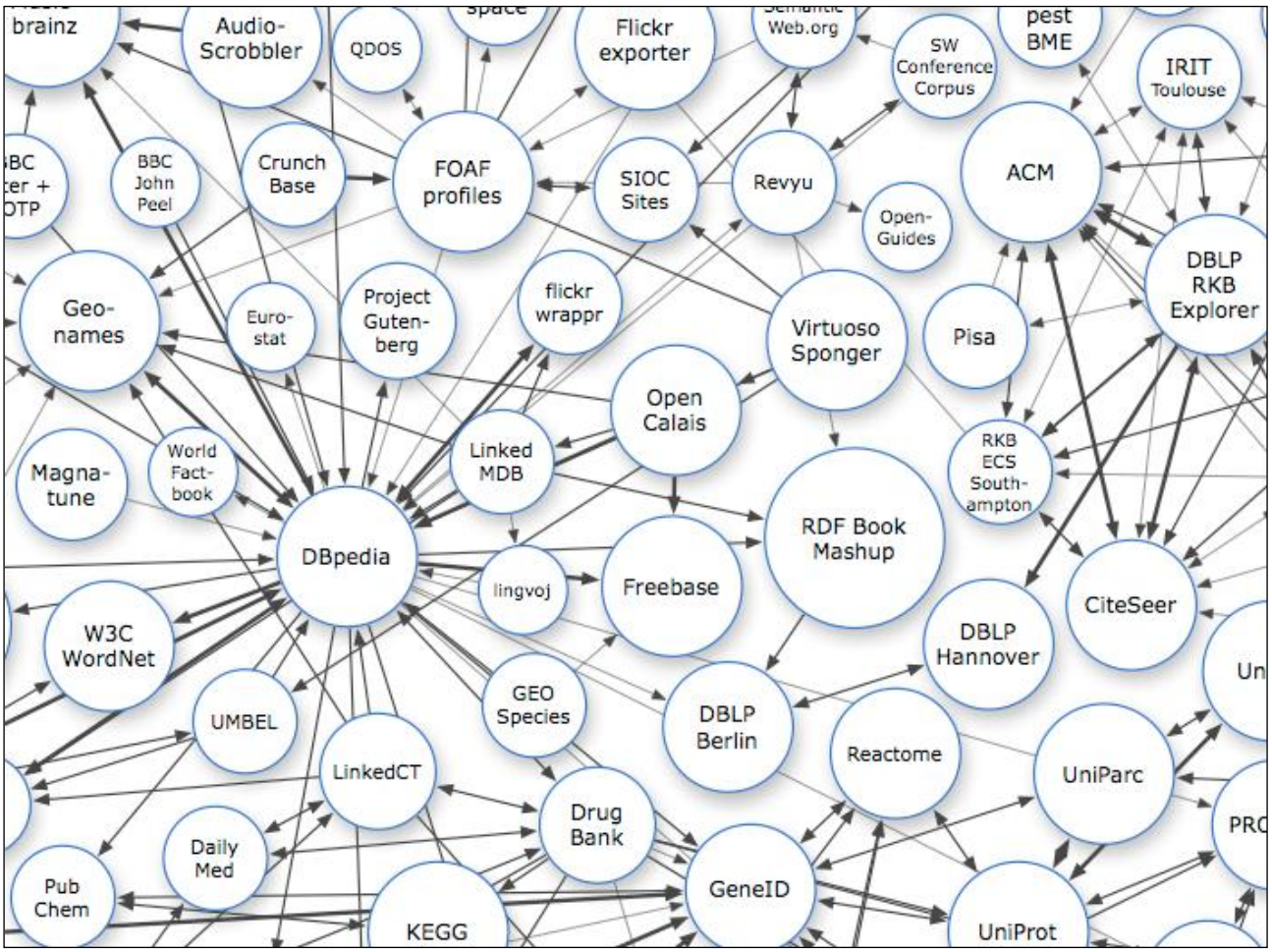
[47] http://www.data.gov/semantic/catalog

[48] Using SPARQL, a recursive acronym which stands for 'SPARQL Protocol and RDF Query Language

queries against the RDF data.[49] In 2009 the Australian government commissioned a report about how government departments can better utilise semantic web technologies when publishing official information.[50]

There are also a number of **unofficial projects** (driven by academics, technologists and others) using semantic web technologies to publish and link together government datasets.[51] The linked data diagram below gives a sense of some of the datasets which have been linked together.

The Linking Open Data (LOD) Project Cloud Diagram[52]



It remains to be seen whether and to what extent the Linked Open Data community's long-term vision of a 'web of data' will be realised. There are critics of the real potential of the semantic web and of the value of converting large quantities of data to linked data formats.[53] That said, there are also clearly interesting and valuable prototypes, as well as traction with key players involved in open government data initiatives in the UK and US.

---

[49] See http://semantic.data.gov/sparql and http://data.gov.uk/sparql

[50] See http://gov2.net.au/blog/2009/09/01/open-for-business/#semantic

[51] See http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

[52] Diagram downloaded from: http://linkeddata.org/

[53] Prominent critics such as Cory Doctorow and Clay Shirky argue that there are fundamental flaws in semantic web advocates' assumptions about human behaviour and classifying information See: http://www.well.com/~doctorow/metacrap.htm and http://www.shirky.com/writings/semantic_syllogism.html

Box D

Applications of governmental 'Linked Open Data'

To illustrate the practical value of the governmental adoption of linked open data technologies, Leigh Dodds, a programmer working with the UK government, gave the following examples of the kinds of queries that are now possible with education and transport interfaces at data.gov.uk:[54]

- Which schools in the Bath and North-East Somerset area have a nursery?
- In which parliamentary constituencies did schools close in 2008?
- What are the names and locations of the 100 newest schools in the UK?
- Give the uniform resource identifier (URI) with latitude and longitude for road traffic monitoring points on the M5

Examples of applications of linked data from the US data.gov site include:[55]

- Combining information on US foreign aid budgets with relevant articles from the New York Times and background information from the CIA Factbook.
- Linking a timeline of US government agency budgets to related New York Times articles.
- Linking information about the incidence of wildland fires in the US from various sources (including government data and structured information from Wikipedia) with information about US government spending on fighting wildfires.

For now it seems prudent to conclude that while publishing government information as Linked Open Data is desirable if resources are available, this is by no means necessary. In the short-term, there are other easier and quicker ways to put datasets online, and the desire to maximise the potential of the semantic web should not in any way slow down the immediate release of government data.

➢ Recommendations:

- Governments and IGOs should explore and monitor the potential of linked data, exploiting the possibilities of permitting users to make more sophisticated queries of the information. At the same time, these future possibilities should not impede the release of current government datasets.

2.4 Cost Considerations

Open government data and access to information advocates can present to government bodies the many medium- to long-term advantages of opening up government data. The social and democratic benefits of greater transparency are accompanied by the potential stimulation to the economy, which is one of the justifications spurring current open government data initiatives by public authorities.

[54] See http://blogs.talis.com/n2/archives/818 and http://blogs.talis.com/n2/archives/836
[55] For links and details see: http://www.data.gov/semantic/index

Nevertheless, governments might have concerns about the cost of opening up government data. There are undoubtedly some potential costs which need to be considered. For example, there are human-resource costs associated with organising and preparing information to put it online, particularly if the decision is taken to develop a special portal which may require an IT and design team. In addition, converting large volumes of data into re-useable formats as recommended in this report can have cost implications, particularly if there is a high level of use of proprietary software.

Some of the technical solutions considered here, particularly initiatives such as converting government data to semantic web and linked data formats, are time consuming and therefore costly. There are, however, many solutions to releasing information which come at a very low cost and should be seen as part of the day-to-day activity of public bodies, such as posting full datasets in open source formats on government websites, properly tagged with metadata so that the information can be found, but with no other special formatting or presentation.

Indeed, the evidence of recent initiatives to put raw data online demonstrates that it can be considerably cheaper than presenting the data to the user via a custom web interface. Hence openness can be achieved at relatively little cost or even with cost savings.

Furthermore, the use of open file formats and preferably of open source software can also significantly reduce the costs for government, as well as lowering the cost barriers to use for members of the public.

➢ Recommendations

It is recommended that governments and IGOs should:

- Look into potential cost savings associated with publishing government data in the same raw format that is used internally inside public bodies, as opposed to presenting it with a custom web interface;
- Consider making greater use of open source software to facilitate and reduce the transaction cost of converting information into open formats;
- Plan for and design in disclosure in order to avoid incurring additional human resources and IT costs when releasing datasets.
- Recognise the lack of knowledge and awareness within many public bodies and conduct necessary awareness-raising exercises such as training programmes run by central government bringing in relevant experts.

3. Legal Issues affecting access to and reuse of government data

There are a number of ways in which information held by public bodies may not be completely "open" to information seekers, from a legal perspective.

The first is that the information falls under legal exceptions on grounds such as national security or protection of privacy and is therefore not released to the public, even when a member of the public files an information request.

The second is when the information and documents produced by public bodies are subject to intellectual property rights such as copyright and database rights, which place possible limitations on the use and re-dissemination of the information. (These are limitations on the actual information, in addition to the copyright applying to the software support in which information is held, examined in the previous section).

The third is when the public body assesses that the information can be commercialised by being sold to for-profit companies which produce added-value products. The information will therefore be released to members of the public or to private companies only upon payment of a fee.

In all three cases the information is not completely "open", as will be examined further in this section.

There is a long tradition of governments claiming ownership of the information the state holds and collects. In some countries, as will be examined in Sections 4.2, this includes the assertion of intellectual property rights over the information created by government bodies. In other countries, even where intellectual property rights cannot by law pertain to certain government information, there is in effect an assertion of ownership in that it is the state that grants permission to reuse the information. The mechanisms for doing this, usually through licenses, are now enshrined in the various national laws adopted in line with the EU's Directive on the Reuse of Public Sector Information, which is be examined in Section 3.

Hence the licences issued for the reuse of government information are usually based on at least one, or even both of the legal considerations below:

- Licences granted under copyright law, which give users permission to reuse the information. These can vary significantly, ranging from automatic licences which are part of the copyright notice to complex licences which have to be negotiated between the copyright holder and the user on a case-by-case basis.
- Licences granted under laws governing the reuse of public sector information (information which may or may not be copyrighted), which set out the terms and conditions that apply to this reuse, including possibly the fees to be paid. These licences often have to be negotiated on a case-by-case basis but can be automatic.

In spite of the fact that in many countries governments assert rights over public sector information, including intellectual property rights, there is need for serious debate as to whether or not this is appropriate in modern democracies, given that the information has been created with tax-payers' money. This debate should consider the principle that public bodies are mere custodians of the information created for and on behalf of members of the public. These are issues which need to be addressed in the long term, as part of the ongoing democratic challenge to promote fully open and accountable government.

A range of solutions are possible to make government data more open, even is some licences apply. In the shorter term, the most important issue is whether it is possible to access and use government information in practice. Given the reality that much information falls under either intellectual property or other restrictions on reuse, the strategy adopted by the open government data movement is to call for the licences which pertain to the reuse of this information to be as flexible as possible, giving members of the public the automatic permission to take the information and use it in any way they want.

Whichever licence is used, it should meet the standard for openness promoted by organisations such as the Open Knowledge Foundation. An open licence is one which results in the information being "open" in the sense that "you are free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and share-alike."[56], according to the Open Knowledge Foundation's Definition.

Box E

Clarification of Terms: "Public Domain"

For legal experts working on issues of freedom of expression and information, to say that "information is in the public domain" means that the information is circulating freely in society. This term is used, for example, by the European Court of Human Rights to refer to information whose circulation has not been restricted by government censorship.[57] The general principle is that "Once information has been made generally available, by whatever means, whether or not lawful, any justification for trying to stop further publication will be overridden by the public's right to know."[58]

For those working on copyright issues, the term "the works are in the public domain" means that the works are not covered by intellectual property rights (copyright and similar rights) either because the property rights have expired or because they have been formally waived.[59] If a work, which could include a government-produced document, is in the public domain, this means that it can be circulated freely, quoted, and reused, without any need to consider copyright restrictions.

In this report we will refer to both of these versions of the "public domain": the public domain in the sense that the information is circulating freely in society, and the public domain in the sense of being free of property rights. It is important to note that information can be in the public domain from a freedom of expression perspective but can then still be subject to copyright: a book which is free of censorship and has been widely circulated is a good example. To avoid confusion, we will clarify which term is being used at each instance.

---

[56] See the Open Knowledge Definition at www.opendefinition.org

[57] See for example, Case of the Sunday Times vs. The UK (No.1), European Court of Human Rights judgement of 26 November 1991, available at http://cmiskp.echr.coe.int/tkp197/view.asp?action=html&documentId=695461&portal=hbkm&source=externalbydocnumber&table=F69A27FD8FB86142BF01C1166DEA398649

[58] See also The Johannesburg Principles, at Principle 17 on Information in the Public Domain, available from Article 19, at: http://www.article19.org/pdfs/standards/joburgprinciples.pdf

[59] See http://en.wikipedia.org/wiki/Public_domain and for a useful explanation from a US perspective see Stanford University page http://fairuse.stanford.edu/Copyright_and_Fair_Use_Overview/chapter8/

3.1 Exceptions on the Right of Access to Information

Many access to information laws assert a "principle of publicity" which states that information should be presumed to be in the public domain (accessible to members of the public) unless one the recognised exceptions permitted by international law applies.

Typical exceptions found in most access to information laws include protection of national security, protection of international relations, protection of criminal investigations, protection of commercial interests, and protection of privacy. Of these, only the latter, protection of privacy is an internationally protected human right; the remainder are interests which may sometimes justify withholding information.

If a public body thinks that an exception applies to information that has been requested or is scheduled for release, then it may be refused. The public body has the obligation to demonstrate or convincingly reason that release of the information would harm one of the legitimate interests which the legal exceptions protect. Sometimes the exception is time-sensitive. For example, to release all the information about an ongoing police criminal inquiry might harm the possibility that the police will catch the criminal suspect. After the inquiry is finished and the criminal arrested, the information could be released.

In addition, consideration should be given to the overriding public interest in releasing the information. For example, disclosing a contract between a public body and a private contractor might potentially harm the commercial interests of the contractor in that their pricing structures would be made public, but there is also a public interest in knowing this information and it should therefore be released.[60]

A case-by-case assessment of these exceptions should always be made. This is generally the case in the world's over 80 access to information laws, although in some cases there are blanket exceptions which permit refusal of information without the need to apply a harm or a public interest test.[61]

In addition, public authorities should consider granting at least partial access to documents when they contain information which might harm a protected interest. Hence release of a document which contains the names of private individuals or other personal data can be achieved quite easily, simply by blacking out the sensitive information and releasing the remainder of the document. International standards such as the Council of Europe Convention on Access to Official Documents require such partial access.

In the early years of access to information laws, there was a tendency simply to refuse access to sensitive documents in full, rather than granting partial access. This use of blanket exceptions is now changing following legal challenges and advocacy from civil society, and it is increasingly common to receive documents with part of the information blacked out rather than receiving an outright refusal.

Partial access should in theory apply to large volumes of data as well. For example, a field in a database which contains personal data such as the e-mail addresses of private individuals can

[60] This example is based on a real case from Ireland: Irish Freedom of Information Commission, Case 99183 - McKeever Rowan Solicitors and the Department of Finance.

[61] For example, in the UK, information from or about the security services is excluded in its entirety from the scope of the Freedom of Information Act and similarly in Turkey information about the duties and activities of civil and military intelligence units is excluded. In Germany, information may be excepted if it causes harm to listed interests but there is no public interest test in the law, although this test may be applied by the courts when considering appeals.

be removed before the remainder of the information is released in order to protect personal privacy but at the same time respect the right of access to information.

There is, however, some reticence on the part of government bodies to do this, which can result in refusal of even partial access to a requested database. One reason is that a huge database may contain many bits of data, which takes time to identify and remove. However, to comply with the right of access to information, public bodies often have no option but to take the time to remove the sensitive data and to grant access.

In the longer term, a solution to this is to design databases with the right of public access in mind. This is now increasingly easy, at least from a technical perspective. It is possible to build a database which performs "one way encryption". This permits e-mail addresses to be included in a database, but in another table that is linked via a "hash value" so that when the data is shared, the e-mail addresses can be separated, thus ensuring that any personal data is kept secure. One way to think about this is to take the analogy of a bank: it is possible to enter a bank and look around but part of the bank – the vault where the money is kept – is locked and secure. Hence it is possible to achieve both security and transparency/access.[62]

It is estimated that 90% of programmers know how to do such a hash, and at least 20% of them understand that it is easy. Other technical solutions such as tagging certain potentially sensitive information contained as text in a database can help rapid review of this information upon receipt of an access to information request or when the database is being prepared for proactive release.

Government officials who are not IT specialists will often not understand how these technical solutions work in practice, and nor will many civil society activists or information commissioners. This is an area where open government data and access to information activists could usefully combine forces to promote solutions for ensuring the appropriate application of exceptions and for achieving maximum openness.

From a human rights perspective once the issue of removing information which falls under legitimate exceptions has been resolved, it should be possible to reuse the information without further restrictions. This is because, as discussed in Section 4, the right of access to information is directly linked to the right to freedom of expression protected under international law. This should mean that once members of the public have been granted access to government data, this data is in the "public domain" not only in the sense that it is circulating freely, but also in the sense that members of the public are free to use it in any way they want.

➢ Recommendations

It is recommended that governments and IGOs should:

- Take a clear position in law and policy that all government-held information is presumed to be accessible to the public unless legitimate exceptions apply;
- Apply exceptions for access to information in a way that ensures maximum disclosure, including by providing partial access;

[62] Example given in interview with FOI and open data expert Greg Elin, conducted via Skype by Helen Darbishire, Access Info Europe, 25 June 2009.

- Make clear that exceptions apply to databases in the same way as they do to any other information held by public authorities, in other words, that partial access should be provided wherever possible;
- Adopt legal or policy requirements to design all future databases with access in mind, ensuring that the technical solutions are in place to permit rapid separation of fields containing sensitive data.
- Ensure that law and practice permits that once information has been released pursuant to access to information requests (after the application of any necessary exceptions), the public has a right to make use of it for any purpose.

3.2 Intellectual Property Issues

International copyright standards, in particular the Berne Convention, first adopted in 1886, establish that the creator of material holds the intellectual property rights in it.[63] The Berne Convention gives national governments the leeway to decide whether copyright applies to "official texts of a legislative, administrative and legal nature, and to official translations of such texts."[64] Other government-generated information may be default be considered as the intellectual property of the body which creates it, unless there is an explicit exemption in national copyright legislation.

The fact that information is subject to intellectual property rights does not necessarily mean that it cannot be accessed and reused at all. It does mean, however, that depending on the type of licence under which it is released, there may be limits on its use, or an obligation to obtain permission before reproducing or disseminating the information.

This section examines the relationship between intellectual property rights (copyright and related rights) and the right to use information obtained from public bodies. It looks at the different types of copyright licence which can apply to information created and released by public bodies and assesses which of these would meet the test of the information being sufficiently "open" for it to be used freely, in line with open government data and access to information standards.

3.2.1 Copyright and Government Information

The existence of copyright in government data has traditionally been an issue for only a relatively small number of users with an interest in republishing government documents and data in bulk, usually for commercial purposes. In practice, around the world, responses to access to information requests almost never make reference to copyright issues, and simply provide the information or documents to the requestor; hence the reuse or publication of these was rarely a problem for the person who received the information. Furthermore, in many countries, the concept of "fair use" allows for citing or reusing of small copies of information even when it is protected by copyright, particularly when the reuse is for journalistic, academic, or non-commercial purposes. There are, however, some notable exceptions to this rule, as the case studies below illustrate.

[63] Berne Convention for the Protection of Literary and Artistic Works, first adopted in 1886, see: http://www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html

[64] Ibid at Article 2.4 which establishes that "It shall be a matter for legislation in the countries of the Union to determine the protection to be granted to official texts of a legislative, administrative and legal nature, and to official translations of such texts."

The relatively lax attitude towards enforcing copyright with respect to information provided in response to information requests is now changing as the open government data movement requests access to entire data sets, and as the advent of digital technologies gradually enables more members of the public to download large quantities of information from government websites and reproduce it in other websites or formats.

There is now an imperative to establish a clearer picture of the legal framework for copyright and database rights in each country and to understand the different options available which will allow public bodies both to uphold their intellectual property rights, where appropriate, and to ensure maximum use of that information which has been created using taxpayers money.

The following three case studies demonstrate how the assertion by public authorities of copyright or database rights can limit public use of information. In the cases from the UK and Belgium, users of information obtained from public bodies were threatened with legal action for violating intellectual property rights. In the UK case the information was taken off–line, but in the Belgian case the online community mobilised, legal support was provided, and the application created using this public information is still online. In the third case, from the Netherlands, a private company went to court and secured a ruling in favour of their right to use a database compiled by a city authority, thus overturning the assertion of a database right. The case study from Belgium also references European Court of Justice jurisprudence that limits the extent of database rights and could therefore be of relevance for open government data campaigners.

Case Study1: "Murder in Samarkand" and UK Crown Copyright

Perhaps one of the most egregious examples of the enforcement of government copyright to prevent dissemination of public information came in 2006 when Craig Murray, the former UK ambassador to Uzbekistan, posted documents obtained under the UK Freedom of Information Act on his website, linked to his book "Murder in Samarkand" which alleged UK government complicity in the torture of war on terror detainees in Iraq.

The UK government's lawyers wrote to Murray stating that although the documents had been received under the FOI Act, the Crown still held copyright in them and that "even if a document is released under the Data Protection Act or Freedom of Information Act that does not entitle you to make further reproductions of that document by, for example, putting them on your website or making further copies to be provided to third parties. The copyright remains enforceable."

The letter from lawyers acting on behalf of the Foreign Secretary (a copy of the letter, dated 7 July 2006, is on file with the authors of this report), calls for the documents to be removed immediately from the website within three days, under threat of legal proceedings. The letter noted that if legal proceedings were initiated, Murray would be charged for legal costs and that "[s]uch costs are likely to be substantial." Murray received legal advice that the costs could run to millions of pounds and so he complied with the take-down order.

The UK government's legal position is confirmed in Office of Public Sector Information guidance which stresses that "provision of information under this access legislation does not mean that the recipient has an automatic right to re-use it, for example to publish it,

or adapt it in some way. Most information supplied in response to an access request will be protected by copyright and permission to re-use it will be required." [65]

Case Study2: Yeri Tiete and Belgian Railway Timetables

In late 2008 Belgian IT student Yeri Tiete, who also goes by the nickname "Tuinslak" ("Garden Snail), developed a mobile website (iRail.be) for accessing the timetables of Belgian railways using the iPhone. In doing so he filled a gap, as no such application existed at the time. He took the data from the websites of the Belgian railway company SNCB/NMBS. According to statements on his blog, Tiete affirms that he "never got paid (or earned any kind of money) for this website. ... iRail.be was only created due to a lack of alternatives ... as a hobby and for personal experience."[66]

In October 2008 Tiete had written to SNCB/NMBS informing them of his plans. The first reply he received was on 8 June 2010, almost 2 years later. The letter states in no uncertain terms that "*Your website makes reuse of SNCB data. This violates its intellectual property rights, including copyright and database rights. It also makes you guilty of the criminal offense of counterfeiting. ... Any violation of the above [copyright notice] leads to prosecution, civil or criminal, without prejudice to the right to compensation for the SNCB. You are hereby formally in default for such violations and we urge you to cease forthwith. Failing this, SNCB will initiate without further notice appropriate legal action*."[67]

A mid-June plea on Tiete's blog urges "Please don't force me to kill this website." The online community mobilised and legal support was provided from a law firm specialising in technology issues. On 6 July 2010 Tiete announced that he had put iRail.be back on line and that his lawyers had sent a letter to SNCB/NMBS asserting his right to use the data.[68]

The legal basis for defending the use of the SNCB/NMBS data includes arguments derived from the jurisprudence of the European Court of Justice in a series of linked cases decided in 2004.[69] The Court ruled that when information in databases is generated as part of the regular activities of a company, then that data is not protected by database rights because the creation of the information has not required "substantial investment" in the terms meant by EU Directive 96/9/EC on the Legal Protection of Databases, and hence may be used by third parties without them needing to seek permission.

Case Study3: Landmark Decision from the Netherlands

In April 2009 the Judicial Division of the Dutch Council of State (*Raad van State*), the highest Administrative Court in the Netherlands, placed limits on the possibility for public bodies to charge for access to databases they have created when it ruled that a public authority could not assert database rights over, nor charge for, data collected with public funds as part of its regular activities.[70]

---

[65] UK Office of Public Sector Information, Guidance Note 1, July 2008, at http://www.nationalarchives.gov.uk/documents/links-between-access-and-reuse.pdf

[66] See the Tunislak blog at http://blog.tuinslak.org/2010/05/irail-2/

[67] Text of the letter from Belgian railways, to be found in Dutch at http://blog.tuinslak.org/2010/06/stopping-irail-be/, translated with help of Google translate.

[68] See Tunislak blog, http://blog.tuinslak.org/, posting of 6 July 2010

[69] Judgments of the Court in Cases C-46/02, C-203/02, C-338/02 and C-444/02 (9 November 2004). A good overview of the cases can be found at http://portal.unesco.org/culture/en/files/25220/11514870521ecj_databases.pdf/ecj_databases.pdf

[70] This case study is based on material published on the EPSIPlatform website http://www.epsiplatform.eu/examples/cases/landmark_nederland_bv_v_amsterdam_city_council and the

> The case was taken to the Court by Landmark Nederland, a large supplier of land and property search information, which in 2006 brought together a national dataset of environmental risks such as contaminated land from a range of sources including Dutch council records. These reports were part of a portfolio of products to be sold to home buyers via estate agency brokers. The City of Amsterdam sought compensation for supplying the data and also wanted to limit its reuse, arguing a substantial investment had been made in compiling the original dataset.
>
> The Court rejected the appeal lodged by the City of Amsterdam for compensation costs for supplying information which would be sold on for profit. The Court ruled that, while the data could be considered to form a database because there had been a substantial investment in its collection, the City of Amsterdam had not borne the risk of this substantial investment, and was therefore not a producer of the database for the purposes of asserting database rights. Consequently the City was not entitled to attach financial conditions or other limitations on the use of this data by Landmark.[71]

These case studies underline the complexities of the various national legal frameworks for copyright and related rights as they apply to government information. Research conducted by Access Info Europe and the Open Knowledge Foundation reveals a complex landscape where, depending on the country, not all government information is subject to copyright and where different solutions are being employed to facilitate access to and use of public sector information.

Three broad spectrum of national legal solutions can be grouped into three categories:

- **Public Domain/Copyright Free Model:** It is possible for the government as a whole (via the legislature) to take a decision to exempt public sector information from copyright. This has been done for example in the United States, where since the passage of the 1976 Copyright Act, federal government information is exempted from copyright. As a result, third parties – including ordinary citizens, researchers, businesses, and NGOs – are able to access and use a wide variety of information without charge and without restriction. The information now available includes satellite images, census data, geospatial data, scientific information, and much else. It should be noted however that at the state and local level public bodies do hold, and often enforce, copyright on the information they produce.

- **Mixed Model**: In a number of European countries, copyright law makes an explicit distinction between which public sector information is free of copyright and which may qualify for copyright protection. Often, core government documents such as law, decrees, decisions, press releases, and other official texts of a political, legislative, administrative or judicial nature may not be subject to copyright whereas other information such as geographic and scientific data as well as photographs and drawings may be. Countries with this mixed model include Austria, Belgium, Finland, France, Germany, Italy, Luxembourg, Netherlands, New Zealand, Poland, Romania, Slovenia, Spain, and Sweden. More detail is given in Annex 4.

- **Broad Copyright**: In some countries, most government information is copyrighted. The broad copyright for government information applies in Commonwealth countries such as

---

press release from Landmark Nederland to be found at http://www.epsiplatform.eu/media/files/landmark_nederland_english_final

[71] Commentators note that this case has been criticized because the court claimed that the database is protected by database rights, but that there is no rights holder. It is likely that other cases will follow.

Australia, Canada, Cyprus, Malta and the UK. In the UK, Crown Copyright (held by the Queen) covers a wide range of material including legislation, government codes of practice, Ordnance Survey mapping, government reports, official press releases, government forms and many public records. In addition, there are other copyright owners, such as local authorities and non-departmental public bodies. Users will generally need some kind of permission to use and reproduce the content, whether through a generic licence or on a case-by-case basis.

For information that is subject to some kind of copyright, held by a public body, the next question is what kind of copyright is used and what the procedures are for getting permission to reuse the information.

Case Study 4: Copyright in the Netherlands

A typical example of these issues can be found in the Netherlands, where much of the information held by the public sector (such as databases, maps, reports, papers, opinions) qualifies for copyright protection. Even factual and functional information from government meets the Dutch standard of original intellectual creation in the literary, artistic or scientific domain.[72] This means that information posted on the websites of Dutch public authorities can carry a copyright protection.

An analysis of public body websites by Mireille van Eechoud and Brenda van der Wal of the Institute for Information Law in Amsterdam found various problems, including that the copyright reservation is often difficult for users to find, since it is either a simple © sign followed by the name of the organization and a year, or a notice hidden somewhere on the website.

Problems with this approach, which is common in many countries around the world, include that the vague definitions of the scope and terms of the copyright might lead members of the public to assume that they cannot use the information. It was found that often public bodies simply failed to inform the public about legitimate uses of the information such as private copying or the right to quote from legitimately published works. The effect could therefore have been to discourage use of the information, even where this was not intended.

One concern identified in this case study is that the very concept of licensing public sector information seems at odds with the notion that citizens have a right to access to such information under the Dutch access to information law. A subject of debate in the Netherlands, it is generally recognised that copyright places a barrier between accessing information and using it. Hence there is a conflict, with intellectual property laws running counter to the principles of the access to information law, designed to stimulate openness of government, to promote public debate, and to permit citizens to influence and control the administration and participate in the democratic process.

The study found that these problems could be addressed by using standard public licences, such as Creative Commons licences. The authors enumerate the advantages of Creative Commons licences, including that they are ready to use, automated and standardised so that public sector bodies do not need to draw up their own licences but can instead benefit from the expertise brought together in CC. The also note that the use of the licences is expanding quickly around the world, thereby aiding recognition and acceptance.

[72] See "Creative commons licensing for public sector information: Opportunities and pitfalls", by Mireille van Eechoud and Brenda van der Wal, Institute for Information Law, Amsterdam, to be found at http://www.ivir.nl/publications/eechoud/CC_PublicSectorInformation_report_v3.pdf

Another advantage of CC licences is that the icons and the human readable Commons Deed are user-friendly and give citizens (including businesses and interest groups) a much clearer indication of which rights are reserved and to what extent, as well as explaining exactly what kind of use is allowed.

3.2.2 Copyright and Licence options for government information

There are three main ways by which public bodies can release copyrighted information to the public:

i. Case-by-Case: Copyright and other rights are enforced, and permission to reuse the information is granted on a case-by-case basis. For example**:** The Romanian National Institute of Statistics has a copyright notice on its website which says that **"***Reproducing the content of this website, completely or partly, in original or modified, as well as its storage in a retrieval system, or transmitted, in any form and by any means are forbidden without the written permission of the National Institute of Statistics*."[73] It is not clear what volume of information from the website would trigger this requirement.

ii. Re-use permitted / Automatic licences: Copyright and other rights are enforced through publication of the licence terms and conditions, or some other legal statement which permits reuse by the public. For example the EU's general legal notice on its website stipulates that anyone is permitted to reuse the material on it by default, so long as the EU is correctly attributed. Licences such as the Creative Commons or Open Data Commons licences also fall into this category: copyright is asserted but use is permitted. In some cases copyright and other rights are enforced but reuse is granted via an automatic licence which can be applied for online (For example, the UK's Click Licences, which are free and automatic but still require some form of action on the part of the user to seek out permission to use the information. The government is currently re-evaluating these licences).

iii. Public domain. Documents and data sets may be exempt from copyright either by a law that applies to all or some parts of government or by dedicating specific material to the public domain using an appropriate legal tool which places no restrictions whatsoever on public reuse. Examples of this are the Creative Commons Zero Licence or the Public Domain Dedication and Licence, which place the entire work in the public domain, waiving all rights. For example, the Dutch government has started to use CC0 for material on its main website.

The level of effort which needs to be made by members of the public to access the licence and to have permission to use the information varies for each of these alternatives. Securing a case-by-case licence inevitably requires some work, contacting the relevant government department, providing information about the user and about what will be done with the information, and so forth.

Copyright is a complex area of law and for those who are not legal experts with specific training in intellectual property issues it can be a confusing and off-putting area. For many requestors of information – and indeed for many public officials – it is often not at all clear precisely what copyright restrictions actually mean in practice.

[73] See National Institute of Statistics – Romania, http://www.insse.ro/cms/rw/pages/index.en.do

Including with the copyrighted material a notice that it may be used is probably the best solution for members of the public. An example is the notice on the European Union's main website which includes the following text:

> **Copyright notice**[74]
>
> © European Union, 1995-2010
> Reproduction is authorised, provided the source is acknowledged, save where otherwise stated. Where prior permission must be obtained for the reproduction or use of textual and multimedia information (sound, images, software, etc.), such permission shall cancel the above-mentioned general permission and shall clearly indicate any restrictions on use.

At first reading, this seems relatively straightforward: the public is allowed to use the information immediately with no further permission needed. At the same time, this copyright notice opens some doubts about how it may be used and precisely how the source should be cited. This may not be clear for future re-users, particularly if the information is combined with much information from other sources. There is no further guidance provided on the website which gives details or explains how this works in practice.

Many users and re-users of public data are private individuals – people such as programmers and citizen journalists as well as representatives of small civic groups – who have little knowledge of how copyright law works, who have received no training of any kind on how to properly reference the sources of the information they are using, and who could not afford to consult a lawyer to explain how this works.

In the research carried out for this report, we found that government publications and websites contain woefully little information to explain to the public how copyright works, the relationship between copyright and access to information, and what users should do if they wish to use that information. There is a pressing need for clear and simple information that explains copyright rules to users of government information and makes clear what they can and cannot do.

One of the great advantages of the Creative Commons licences and Open Data Commons licences from the perspective of the user is that they come with very clear and simple terms and conditions. For example, the standard Creative Commons attribution ('CC-BY') allows all types of use of the work as long as the author is credited and information about the copyright status of the work is kept intact.

It is a common misperception that Creative Commons is a single licence, whereas there is in fact an entire range of Creative Commons licences, with a corresponding variety of levels of reusability. For example, some Creative Commons licences are not open, in so far as they limit use to non-commercial actors or contain the 'non-derivative' clause, which only allows the user to make and distribute verbatim copies. The Open Knowledge Definition distinguishes between licences that are open and those that are not open. Government bodies are strongly

[74] See http://ec.europa.eu/geninfo/legal_notices_en.htm#copyright

encouraged to make use of the most permissive and least restrictive of these licences, i.e. the attribution-only licences such as the Creative Commons Attribution licence.

Another problem is that where individuals are granted permission to use official material, they may not be able to pass on this permission. This may prevent the public from improving or adding value to the information by sharing it and encouraging others to use it as well. This applies in particular to licences and terms of use which prohibit redistribution, derivative works, reuse or particular kinds of use (e.g. commercial) as they will not deliver many of the benefits of open government data discussed in this paper. For instance, they will reduce the likelihood of development of web applications and services that make official material easier to understand by combining information with other external datasets and so on.

The third method of publishing information is to release it into the public domain. This will be done either because the information is by law not subject to copyright or because the type of licence selected is one that dedicates the information to the public domain (most commonly under a Public Domain Dedication and Licence or a Creative Commons Zero licence), under which the rights holder waives all copyrights and related or neighbouring rights over the work, such as moral rights (to the extent that they can be waived), publicity or privacy rights, rights protecting against unfair competition, and database rights as well as rights protecting the extraction, dissemination and use of data.[75]

The great benefit of the third option, that of releasing information fully in the public domain, is that it ensures that the public can continue to use and add value to the material downstream. This automatically makes public data fully "open".

The creative commons movement and the open government data movement have been successful in promoting the use of Creative Commons licences by public authorities. Whilst not all of these are CC0 licences, there is a growing trend to facilitate reuse.[76]

- An example is the Australian Government's data catalogue which publishes all information under a Creative Commons Attribution 3.0 Australia licence.[77]

- In March 2010, the Netherlands government launched Rijksoverheid.nl, a new website which will hold information about all Dutch ministries. The website's copyright policy is that by default it will remove all copyright restrictions with the CC0 public domain waiver.[78]

A recent policy move by the New Zealand Government which in July 2010 introduced the New Zealand Government Open Access and Licensing framework (NZGOAL)[79] is examined in Box F

For public policy makers who wish to maximise the societal benefits of public sector information by making it as "open" as possible, restrictive licences should be discouraged. Rather government bodies should be encouraged to adopt open licences which place few or no restrictions upon users, nor on the use made of the material. For open government data and

[75] Creative Commons http://creativecommons.org/choose/zero/
[76] A good overview of the many countries using Creative Commons licences for at least some information can be found here: http://wiki.creativecommons.org/Government_use_of_Creative_Commons
[77] See Australian government website http://australia.gov.au/about/copyright#data
[78] See http://creativecommons.org/weblog/entry/21473 and http://www.rijksoverheid.nl/copyright
[79] See the NZGOAL framework at http://www.e.govt.nz/policy/nzgoal/html-version, see also summary reports of the main features at http://blog.okfn.org/2010/08/06/new-zealand-government-open-access-and-licensing-framework/ and http://epsiplatform.eu/news/news/new_zealand_goal

access to information advocates, now is a good moment to push for publication with open licences and to achieve a potential paradigm shift in the usability of government information.

➢ Recommendations

It is recommended that governments and individual public bodies should:

- Ensure that each and every time that public sector information is released, whether proactively or in response to an access to information request, clear and simple information should be provided about the precise conditions pertaining to the use of this information;
- Take steps to inform the public about how copyright functions. Specifically, government publications and websites should contain clear explanations;
- Before asserting copyright government bodies should verify whether it does actually apply to the data at issue.
- Ensure that even where government-created information is subject to copyright, members of the public should be granted permission to use it with no need to ask for a specific permission or licence;
- Accompany any requirements to obtain a licence to use the information with clear explanations about the procedures to be followed and ensure that the process is as simple and fast as possible;
- Re-evaluate their licensing and copyright frameworks with the goal of maximising the volume of information that is released either with no copyright or under an open licence;
- Support international efforts to harmonise and simplify copyright for government information;
- Wherever possible release information under an open licence;
- Design databases in ways that ensure that the intellectual property rights in the software platform do not prohibit access to the information they contain;
- In the long term, exempt government information from copyright or, failing that, release it under a Creative Commons zero licence for both content and data. In cases where this is not practicable, an attribution only licence such as the Creative Commons BY licence is the recommended option.

The access to information and open government data communities should:

- Work together for reforms in law and practice to promote greater use of open licences for the publication of government information;
- Engage in advocacy for legal reform to ensure that government information is exempt from copyright.

BOX F

NZGOAL: the New Zealand Government Open Access and Licensing framework

Launched on 6 August 2010 by the New Zealand Government, this Cabinet-approved framework meets many of the demands of the open government data movement. In particular, it stipulates that a Creative Commons Attribution (BY) licence is the default for all government information unless a specific restriction applies. The NZGOAL explicitly discourages non-commercial licenses to make sure that open government data can be used (and contributed to) by commercial users, thereby encouraging innovation and stimulating the economy. Likewise, non-copyright material that is or may be of interest to people should be posted online using a 'no-known rights' statement.

The NZGOAL framework also addresses open licensing and open access principles, and guides state agencies through the questions they need to consider before releasing material for re-use. It contains a principle of non-discrimination which prohibits agencies from discriminating when selecting an NZGOAL licence between individual, not-for-profit and commercial uses of the relevant copyright works.

The NZGOAL framework does not apply to information or works containing personal or other sensitive information, but does give guidance on anonymising datasets which, once stripped of personal information, might be licensed or released.

Importantly, NZGOAL urges departments to publish in formats which make the data easy to use, taking into account the wishes of likely users and making use of formats which are best suited for interoperability and re-use, and which are searchable and indexable by search engines. Released datasets should be posted on the data.govt.nz website.

An Open Format Principle requires that, when releasing works or material in proprietary formats, agencies should also release the works or material in open, non-proprietary formats.

3.3 Rules governing reuse of public sector information

As already noted, public sector information has great value for the society as a whole, which includes the democratic value of having more, better, and timely information from government bodies. This allows members of the public to participate actively in the life of their society, both by participating directly in governance and by using the information in other ways to engage with their communities.

The information generated by public bodies can also be considered to have an economic value which can be assessed in a number of ways. One is the immediate economic benefit which can

be obtained by selling the data, for example by selling it to private bodies which will reuse it. Selling data produced by public bodies is a common part of the economic model of government in many countries, particularly in some areas which are seen as peripheral to the core functions of government, such as collecting geo-spatial or meteorological data.

It is now widely considered that the income that can be generated by commercialising government data is only one of the ways of benefitting from the potential value of the information, and is perhaps the least interesting way in terms of the benefit to the economy as a whole. Recent economic analyses show that when information is provided to the public free of charge or at very low cost, then individuals and private enterprises can take that information and create added-value products which they can then market. This economic activity stimulates the economy and also provides revenue to the government in the form of taxes. In addition, the information now available free of charge from the public bodies can be used for other civic society projects, such those mentioned in Section 5.

Examples of such information that has this broader potential value include mapping, meteorological, legal, traffic, financial, and economic data. Much of this raw data could be used for or integrated into new products and services which we use on a daily basis, such as car navigation systems, weather forecasts, or financial and insurance services. According to a survey conducted by the European Commission in 2006, the overall market size for PSI in the EU is estimated at EURO 27 billion.[80]

In the United States where federal government information is in the public domain, there are no restrictions on its use, and the value of this has been resoundingly confirmed by the Obama administration. In a January 2009 memo, the White House noted that release of public sector information serves "[t]o increase accountability, promote informed participation by the public, and create economic opportunity." [81]

In Europe, however, there has been a long tradition of charging for the use of certain information generated by public sector bodies. Historically, a limited number of providers, maybe even only one, had access to a set of government information which they and they alone had a right to publish.

This monopolistic position was not consistent with the principles of the common and open European market. With the advent of the digital age, the possibilities opened for information users from across the EU to access the same information relatively easily. To this end the EU adopted the Directive 2003/98/EC on the re-use of public sector information (PSI Directive) in 2003.[82]

The objectives were to eliminate the barriers to access the multitude of information gathered by public bodies in Europe, including in areas such as social, economic, geographical, weather, tourist, business, patent and educational information. The goal was that non-governmental bodies, such as private businesses, would then have an equal opportunity to reap the potential economic benefits inherent in the use of such information.

---

[80] See the study on Measuring European Public Sector Information Resources, available from http://ec.europa.eu/information_society/policy/psi/docs/pdfs/mepsir/final_report.pdf

[81] See the Open Government Directive http://www.whitehouse.gov/sites/default/files/microsites/ogi-directive.pdf

[82] See the Directive 2003/98/EC at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF

In doing so the Directive aimed to create a level playing field for all users of publicly-gathered information, from both the public and private sectors. It also aimed to address the mindset of public sector bodies, which were failing to realise the economic potential of the information they held, making the most of the digitalisation of information to bring about this change in law and practice.

Hence the PSI Directive was originally conceived with the intention of encouraging EU member states to open up more of their public sector information for reuse by the public. This was motivated by, on the one hand, the potential societal benefits of being able to access the data and, on the other hand, the economic benefits to be obtained from the innovative reuse of government material by businesses in web, mobile and other information services.

However, many critics suggest that the PSI Directive is too weak and insufficiently clear regarding the obligations of EU member states in making their data available for reuse, particularly with regard to charging for PSI, and making it available for commercial use. The Directive is widely regarded to have failed in achieving its objective of creating a level playing field, even for commercial reuse of government information.

3.3.1 The Complications Caused by Dual Systems for ATI and PSI laws

A further set of issues surround the relationship between the PSI Directive and access to information laws. These are issues which are relevant in Europe but which carry cautionary lessons for other parts of the world where there is still a culture of charging for the right to use large datasets from the administrative, legislative and judicial branches of government.

A first problem is the need to harmonise definitions between rules on the use of public sector information and access to information laws in order to avoid creating an artificial division as to which rules apply to which kind of information, or worse, separating the right of access from the right to reuse.

For example, in France a chapter on the reuse of PSI was introduced into the 1978 Law on Access to Administrative Documents. In doing so, limits were created on the use of information obtained under access to information requests, with the need to seek permission to reuse the information and the possible threat of a fine for not doing so. Hence, although the law defines PSI as information contained in administrative documents, this leaves some doubt about what information is being referred to, particularly with respect to databases.[83]

A second, and potentially more serious problem, is that public sector information reuse rules can set up a dual charging regime. On the one hand, there is the principle that access to information under access to information requests should always be free of charge because the information generated by public bodies is created using taxpayers' money. The only permissible charges are copies of information (photocopies, copies on discs) and postage or other costs which are actually incurred in delivering the data.[84]

On the other hand, the licence fees for data released under the PSI Directive can run to millions of Euros. A 2009 study commissioned by the EU found that around Europe a single

[83] See France's Law of 17 July 1978 on Access to Administrative Documents, available at: http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006068643&dateTexte=20100720

[84] This is confirmed by the Council of Europe Convention on Access to Official Documents which clearly states that: "Inspection of official documents on the premises of a public authority shall be free of charge. ... A fee may be charged to the applicant for a copy of the official document, which should be reasonable and not exceed the actual costs of reproduction and delivery of the document."

piece of geographic data – an aerial photo of 10 km^2 of land – can range in price from free of charge to €292 with the average price being €62; of course for any serious reuser, many such units would be required.[85] The same study found that 27 holders of geographic information from 24 EU countries had an income in 2007 for their products of a total of €356 million. Whilst geographic data seems to be a particularly lucrative market (so prices may be different in other sectors), these figures do indicate the potentially very high costs of obtaining public sector information. From the perspective of the right of access to information as an inherent part of freedom of expression it is hard to see how information accessed under access to information laws or information which has been proactively disclosed, can be subject to any further fees when any use is made of it. It is this kind of charging regime that has given rise to campaigns such as the UK's "Free Our Data" Campaign reviewed in Section 5.4.

In some countries it has been established that the right of access to information takes precedence over any charges, apart from those permitted for making copies of and delivering the information. This is the case in the Netherlands where the PSI Directive has been incorporated into the Dutch access to information law, the "WOB", and the government has decided that information should be provided for re-use against payment, but only of the costs of the carrier of the data (such as a CD-ROM or DVD).[86]

In many countries, however, the PSI reuse rules have forced an artificial distinction between smaller chunks of information, which are provided in responses to access to information requests, which may be used without charge, and larger collections of information such as databases which fall under a potential charging regime, particularly when that reuse is for commercial purposes.

A halfway-house solution has been found in Slovenia, where the Information Commissioner has ruled that access to public databases shall always be free of charge when the use to be made of the information is in the public interest, a ruling which should encourage proactive disclosure of databases.[87]

There is a widespread lack of clarity about what the rules permit country by country. Access Info Europe and Open Knowledge Foundation conducted a survey about the relationship between access and reuse and received responses from 15 countries globally.[88] The survey

[85] MICUS Management Consulting report on "Assessment of the Re-use of Public Sector Information (PSI) in the Geographical Information, Meteorological Information and Legal Information Sectors", published March 2009, available at: http://www.micus.de/pdf/MICUS-Studie_PSI_EU_March_2009.pdf.

[86] See report on the European Public Sector Information Platform website: http://www.epsiplatform.eu/examples/cases/landmark_nederland_bv_v_amsterdam_city_council

[87] The case which raised this issue in Slovenia followed a request by the Slovenian Speleological Society to topographic maps produced by the Slovenian Mapping Agency. The commissioner ruled that the information should be provided free of charge in order that the Speleological Society could produce the world's first online caves registry which includes 9262 caves and links to another 4293 documents. Most other countries in the European Union allow public bodies to charge even for non-commercial reuse of public sector information, but in Slovenia this is now clearly prohibited. The decision, in Slovenian only, can be found at http://www.ip-rs.si/informacije-javnega-znacaja/iskalnik-po-odlocbah/odlocbe-informacije-javnega-znacaja/?tx_jzdecisions_pi1[showUid]=229&tx_jzdecisions_pi1[highlightWord]=raziskovanje%20jam&cHash=5b26f278d5, last accessed 28 March 2010. Links to the maps can be found at the Ljubljana Caving Society (Društvo za raziskovanje jam Ljubljana) at http://dzrjl.speleo.net/, last accessed 28 March 2010.

[88] Answers were received from at least one respondent in Austria, Canada, Croatia, the Czech Republic, Estonia, France, Germany, Hungary, Macedonia, Montenegro, the Netherlands, Moldova, Romania, Spain and the USA.

aimed to map out problems for users of public sector information with regard to access to databases and the legal issues related to copyright and licenses.

The survey revealed that in many countries there is a lack of legal clarity, either because there are conflicting rules or because the way the rules are applied by different public bodies is not consistent. Civil society representatives, lawyers, and other respondents to the survey highlighted unresolved conflicts between access to information and reuse laws, which sometimes cross-reference each other without resolving differences. This can make it difficult for public officials who are not sure which rules should be followed and/or can apply them arbitrarily.

For members of the public, there is often no list of data that need licenses to be used nor is there a uniform mechanism for applying for those licenses. Fees can be charged for collection of data, production of data, reproduction of data, dissemination of data and for the time invested by the public authorities, even though this sometimes conflicts with the provisions of access to information laws. Finally, a number of respondents noted that it is not clear in their countries whether public documents can have copyright or not.

Sometimes there is a creeping infringement on the principles established in the access to information law. In Estonia for example, the drafters of the access to information law followed the principle that information had already been paid for once by the taxpayer and that there should not be any additional charges. This started to change over time with increased pressure on public funds and a relaxing of the rules to give some entities the right to recuperate their development costs through charges. Advocates in Estonia were successful in getting rid of the user fees for the population registry, but they report that the land and property registries continue to charge fees.[89]

By addressing these challenges, the open government data movement is having an impact on policy makers' understanding of public sector information and this is leading to a debate about rewriting the PSI Directive, something which is scheduled for 2012. It is not clear yet what the nature of the reforms will be, but there are indications that the European Commission intends to broaden and strengthen the directive so that EU member states will be encouraged to make Public Sector Information freely available for all to use without restriction. An alternative approach would be to abolish the PSI Directive entirely and revert to the right of access and the right of reuse that are already established under national access to information laws and by the right to freedom of expression, but it is reported that this seems unlikely to happen. There is, therefore, an important work to be carried out by open government data and access to information advocates in Europe during the next 2 years to push for making as much data as possible available in the public domain, in both senses of the term.

➢ Recommendations

It is recommended that governments and IGOs should:

- Re-evaluate the relationship between access to information and reuse of public sector information laws to ensure that all information held by public bodies, including databases, falls within the scope of the right to information;

[89] Information provided via e-mail by Ivar Tallo, founder and president of the e-Governance Academy and former member of the Estonian Parliament, April 2010.

- Re-evaluate pricing and licensing frameworks to ensure that, by default, government information, including raw data sets, should be available for anyone to use for any purpose (including added-value commercial purposes) without charge;

- Ensure that the public is able to use information published proactively or released under access to information laws free of charge.

Open government data and access to information activists should:

- Engage in the discussion and work together around the reform of the PSI Directive in Europe, a process which offers an opportunity to set global standards on reuse of public sector information.

PART II. THE RIGHT TO INFORMATION AND OPEN DATA MOVEMENTS

4. The Right of Access to Information Movement

The first "freedom of information activists" were probably the Enlightenment thinkers in Sweden and Finland who successfully promoted the adoption of Sweden's 1766 Freedom of the Press Act, which establishes the principle of the openness of official documents and is widely considered to be the world's first access to information law.[90]

The right to access and use information were intrinsic to freedom of the press according to this constitutional law, which established a freedom to print in whole or in part extracts from

> *correspondence, documents, protocols, judgments and awards* [produced by] *courts and government departments, our senior administrators and consistories or other public bodies ... which, when requested, shall immediately be issued to anyone who applies for them on penalty of the provisions following paragraph*.

Documents should be provided "immediately" and the penalty foreseen is loss of office for the public official who fails to provide the documents or in any way obstructs their release.[91]

Similarly, the French Declaration of the Rights of Man and the Citizen of 1779 provided that:

> *All citizens have the right to ascertain, by themselves, or through their representatives, the need for a public tax, to consent to it freely, to watch over its use, and to determine its proportion, basis, collection and duration.*

The historical rule of law and accountability grounds for government transparency were joined in the second half of the 20th Century by arguments in favour of efficient and effective administration, with provision of information about administrative procedures being one of the good administrative practices promoted by access to information laws adopted in countries such as France and the Netherlands in the 1970s. More recently, it has been recognised that information is needed for public participation in government decision-making.

Hence four primary drivers of government transparency throughout history can be identified:

1. **Rule of law**: The right to be informed about the laws and decisions with which the public has to comply as well as the right to know about ones rights.
2. **Accountability**: The right to hold governments accountable both at and between elections is contingent on the right of access to information.
3. **Access to services**: The right of equitable access to government services and funding places an obligation on public bodies to publish information; the volume of information published has expanded significantly in the past decade with growth of electronic access to services or "e-government".
4. **Participation**: The right to participate directly in governance has developed in recent years through initiatives such as participatory budgeting and public consultations on a

[90] One of these men was Anders Chydenius, the promoter of the Freedom of the Press Act in the Swedish parliament. The Chydenius Foundation, founded in 2001, works in favour of the right to information, see http://www.chydenius.net/eng/index.asp. Another was Peter Forsskål whose "Thoughts on Civil Liberty" was first published in 1759 and was published in English translation in 2009 by Atlantis Stockholm.

[91] Text from "The The World's First Freedom of Information Act", published by the Chydenius Foundation (2006) available at: http://www.chydenius.net/pdf/worlds_first_foia.pdf

range of issues, this in turn is contributing to expanding access to information at increasingly earlier stages in the decision-making cycle.

These drivers have contributed both to the development of extensive proactive disclosure of information by governments and inter-governmental bodies, as well as to the more recent development of the right to request and receive information.

4.1 The global access to information movement

In spite of the drivers listed above, even other Nordic countries were slow to copy Sweden in enshrining the "principle of openness" in law: Finland adopted an access to information law in 1951, Norway and Denmark in 1970, and the Netherlands in 1978.

The post World War II growth of public administrations and the welfare state combined with the development of democratic thinking in general and greater empowerment of the public vis-à-vis government resulted in a second wave of access to information laws. In the United States the US Freedom of Information Act was passed in 1966 (the law was strengthened in the wake of the 1974 Watergate scandal) and then in countries such as Australia (1982), New Zealand (1982) and Canada (1983). France's 1978 law is part of this wave, focusing as it does on improving relations between the administration and citizens.

The starting line for the campaign for a universal right of access to information was the fall of the Berlin wall. The post-Communist civil society movement in central and eastern Europe recognised the importance of the shift in power which such a right represented and mounted vigorous campaigns for broad right of access to all information held by all public bodies irrespective who was asking or why they wanted the information.[92] This was not simply a call for recognition of a right that already existed in established democracies; rather, it was an expansion of that right, moving it from being a privilege granted by the administration to the status of a fundamental human right.

Hungary led the way in central and eastern Europe with the adoption of its 1992 Act on Protection of Personal Data and Public Access to Data of Public Interest[93], which became just the 12th access to information law in the world. Inspired by both Sweden's 1766 law[94] and the United States 1966 Freedom of Information Act[95], the Hungarian act grants a broad scope to the right in terms of both the bodies and the information to which it applies, being: "*any information or knowledge, not falling under the definition of personal data, processed by an organ or person performing a state or local government function or other public function*

92 Much has been written about the role of civil society in the drafting, adoption, and implementation of access to information laws. See, inter alia, Andrew Puddephatt "Exploring the Role of Civil Society in the Formulation and Adoption of Access to Information Laws: The Cases of Bulgaria, India, Mexico, South Africa, and the United Kingdom" (2009) published by the World Bank Institute, available at http://siteresources.worldbank.org/EXTGOVACC/Resources/atICivSocietyFinalWeb.pdf. See also "Transparency and Silence: A survey of access to information laws and practices in 14 countries" (2006) by the Open Society Justice Initiative, available at: http://www.soros.org/initiatives/justice/focus/foi/articles_publications/publications/transparency_20060928

93 Hungary's Act LXIII of 1992 on the Protection of Personal Data and Public Access to Data of Public Interest http://abiweb.obh.hu/dpc/index.php?menu=gyoker/relevant/national/1992_LXIII

94 See David Goldberg writing in introduction to English translation of Peter Forsskal's "*Thoughts on liberty*" (1759), http://www.peterforsskal.com/thoughts-on-civil-liberty

95 United States, The Freedom of Information Act, 5 U.S.C. § 552, As Amended By, Public Law No. 104-231, 110 Stat. 3048 http://www.justice.gov/oip/foia_updates/Vol_XVII_4/page2.htm

determined by a rule of law, or any information or knowledge pertaining to the activities thereof, recorded in any way or any form, irrespective of the manner it is processed and its independent or collected character."[96]

Subsequent laws adopted in the new eastern European democracies echoed and confirmed this breadth of scope of the right, securing a "follow the money" principle by placing an obligation on bodies which "*manage public funds or operate with state property*" (Slovakia's 2000 Act on Free Access to Information[97]) or and even to individuals which "*as far as only their activities financed with funds from the consolidated state budget, subsidies from the European Union funds or allocated through EU projects and programs*" (Bulgaria's 2000 Access to Public Information Act[98]).

The exercise of public power in any form also creates obligations to uphold the right of access to information in many countries, such as in Estonia's Public Information Act (2000)[99] which establishes that: "*The obligations of holders of information extend to legal persons in private law and natural persons if the persons perform public duties pursuant to law, administrative legislation or contracts, including the provision of educational, health care, social or other public services – with regard to information concerning the performance of their duties*."

The impact of this new wave of constitutions and laws was therefore to establish access to information both as a fundamental human right and as a *sine qua non* of a democratic society. Democratic reformers in and out of government combined constitutional and rights-based arguments with comparative examples to press for strong access to information laws. The result was a rapid adoption of access to information laws which conformed to a common high standard, granting a broad right to information that placed obligations on all branches of power.

A boost to this process was the successful campaign to secure adoption in 1998 of the sector-specific Aarhus Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters (Aarhus Convention).[100] The Aarhus Convention was in many countries a foot in the door for the transparency movement, being actively promoted by activists. For example, in Bulgaria, the Access to Information Programme, the regions first dedicated access to information organisation founded in 1996 had its roots in campaigns for environmental information as well as the fight against corruption and broader human rights goals.[101]

The European Union Accession process was made full use of by campaigners to lever this change and so, even though there was no formal *aquis Communitaire* (requirement to have a law) for states wishing to join the European Union, having an access to information law became part of the de facto signposts on the road to joining the club of the European Community. All eight former communist countries which joined on 1 May 2004 had an access to information law, as did Romania and Bulgaria when they joined on 1 January 2007.

---

[96] Supra note 2, Article

[97] Slovakia Act on Free Access to Information and Amendments of Certain Acts available in English at http://www.aip-bg.org/pdf/slov_faia.pdf

[98] Bulgaria Access to Public Information Act (2000) available at http://www.aip-bg.org/library/laws/apia.htm

[99] Estonia Public Information Act (2000), available at http://www.legaltext.ee/text/en/X40095K2.htm

[100] Aarhus Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters, Adopted at Aarhus, Denmark, 25 June 1998, http://www.unece.org/env/pp/documents/cep43e.pdf, last accessed 28 March 2010.

[101] Access to Information Programme report "10 Years Access to Information Programme" (2007) available at: http://www.aip-bg.org/pdf/aip_en_10y.pdf

It is hard to identify precisely the impact of the right to information developments in central and eastern Europe on the situation in western Europe, but there undoubtedly was an effect, a kind of looping of democratic standard-setting from west to east and back to the west again. For example, campaigners in Germany used the fact that the country was increasingly isolated by not having a national access to information law,[102] and successfully promoted the adoption of what became the 2005 Federal Act Governing Access to Information held by the Federal Government.[103]

These developments were reflected at the international level by the Council of Europe which in 2002 updated its 1981 Recommendation with a new text that included a broad definition of "official documents" as any information recorded in any form and defined public bodies as all levels of government and administration, including bodies which are "*natural or legal persons insofar as they perform public functions or exercise administrative authority*".[104]

The access to information movement in Europe has achieved stunning results in a short period of time: From 8 laws in 1989, today 40 of the of the 47 Council of Europe member countries have rules setting out mechanisms for access to information and although they vary in quality, most meet the minimum standard for this right as set out in the new Council of Europe Convention on Access to Official Documents adopted in 2009, which has already been signed by 25% of member states and ratified by two, Norway and Hungary.

A similar civil society movement is accompanied by the democratic transitions in Latin America, with the vanguard being the *Grupo Oaxaca* in Mexico, a gathering of intellectuals, academics, journalists and activists which promoted and helped to draft Mexico's Federal Law on Transparency and Access to Public Information.[105]

The Mexican access to information law is a remarkably strong and broad instrument whose functioning is assured by a well-resourced Information Commission, the Federal Institute for Access to Information (IFAI by its Spanish acronym) which is headed by five information commissioners empowered, inter alia, to search the files of public bodies and to order the disclosure of information.

Other laws adopted in the Latin American region, such as Peru's 2002 Law[106] were inspired by the movement to promote the right to information in Eastern Europe, by the adoption of South

---

[102] See Manfred Redfels of Netzwerk Recherche, "*Informationsfreiheit: Deutschland als verspätete Nation Warum die Bundesrepublik sich schwer tut mit dem Abschied vom „Amtsgeheimnis*" available at: http://www.netzwerkrecherche.de/files/ifg-deutschland-als-verspaetete-nation.pdf

[103] Federal Act Governing Access to Information held by the Federal Government - (Freedom of Information Act) available at http://www.bfdi.bund.de/cae/servlet/contentblob/412040/publicationFile/24681/TextIFG_EN.pdf

[104] Recommendation (2002)2 on Access to Official Documents http://www.coe.int/T/E/Human_rights/rec%282002%292_eng.pdf

[105] See Issa Luna Pla, *Movimiento social del derecho de acceso a la información en México* published by the Universidad Nacional Autónoma de México (2009), available at http://www.bibliojuridica.org/libros/libro.htm?l=2629. The Mexican law was adopted in 2002 by the government of President Vicente Fox who in 2000 had become the first president of Mexico to be elected from an opposition party, ending the 71 years of one-party rule by Mexico's Institutional Revolutionary Party, the PRI.

[106] Peru's 2002 Law on Transparency and Access to Information in Spanish can be found at: http://www.pcm.gob.pe/transparencia/Ley_de_Transparencia_y_Acceso_a_la_InformacionPublica.pdf, last accessed 27 March 2010.

Africa's strong Promotion of Access to Information Law[107] as well as by the proximity to the United States and knowledge of its FOIA. Numerous exchanges and study visits took place ensuring a strong cross-fertilization of experiences, strategies and standards. The Council of Europe's Recommendation was translated into Spanish and distributed in the region. It was also from Latin American that an international human rights tribunal ruled for the first time that access to information is a fundamental human right (see Section 4.4).

Table 3: Adoption of access to information laws around the world

Year	Countries*	Number of Countries Adopting Access to Info law this period	Cumulative Total
1766-1950	Sweden (1766)	**1**	**1**
1951-1960	Finland (1951)	**1**	**2**
1961-1970	United States (1966)	**1**	**3**
1971-1980	Denmark (1970), Norway (1970), France (1978), Netherlands (1978)	**4**	**7**
1981-1990	Australia (1982), Canada (1982), New Zealand (1982), Colombia (1985), Greece (1986), Austria (1987), Italy (1990)	**7**	**14**
1991-2000	Hungary (1992), Ukraine (1992), Portugal (1993), Belgium (1994), Belize (1994), Iceland (1996), Lithuania (1996), South Korea (1996), Ireland (1997), Thailand (1997), Israel (1997), Latvia (1998), Albania (1999), Czech Republic (1999), Georgia (1999), Japan (1999), Liechtenstein (1999), Trinidad & Tobago (1999), Bulgaria (2000), Estonia (2000), Moldova (2000), Slovakia (2000), South Africa (2000), United Kingdom (2000)	**24**	**48**
2001-2010	Bosnia & Herzegovina (2001), Poland (2001), Romania (2001), Jamaica (2002), Angola (2002), Mexico (2002), Pakistan (2002), Panama (2002), Peru (2002), Uzbekistan (, Tajikistan, Zimbabwe (2002), Armenia (2003), Croatia (2003), Kosovo (2003), Slovenia (2003), Turkey (2003), St. Vincent & Grenadines (2003), Dominican Republic (2004), Ecuador (2004), Serbia (2004), Switzerland (2004), Antigua & Barbuda (2004), Azerbaijan (2005), Germany (2005), India (2005), Montenegro (2005), Taiwan (2005), Uganda (2005), Honduras (2006), Macedonia (2006), Jordan (2007), Kyrgyzstan (2007), Nepal (2007), Nicaragua (2007), China (2008), Chile (2008), Cook Islands (2008), Uruguay (2008), Indonesia (2010), Bangladesh (2009), Russia (2010)	**42**	**80**

[107] See South African History Archives, http://www.saha.org.za/interests_and_links/access_to_information.htm

4.2 The Right to Information Movement Today

The right of access to information has developed significantly in recent years, with at least 80 countries worldwide currently having a dedicated legal framework for requesting and receiving information.

The right is also enshrined in at least 50 national constitutions.[108] A typical example is the Constitution of Poland:

> *A citizen shall have the right to obtain information on the activities of organs of public authority as well as persons discharging public functions Such right shall also include receipt of information on the activities of self-governing economic or professional organs and other persons or organizational units relating to the field in which they perform the duties of public authorities and manage communal assets or property of the State Treasury.*
>
> *The right to obtain information shall ensure access to documents and entry to sittings of collective organs of public authority formed by universal elections, with the opportunity to make sound and visual recordings.*[109]

Today the global movement for the right of access to information consists of civil society organisations, academics, lawyers, journalists, information commissioners and other activists in as many as 90 countries around the world. There is a global network, the Freedom of Information Advocates Network, which has around 190 civil society organisations and an active discussion list of 500 specialists working on access to information. There are also a number of regional networks as well as national networks, which link to hundreds more activists. Section 2.1 reviews the main actors in and strategies in the right to information movement.

In addition to the recent growth in the number of access to information laws (See Table 3), the movement has had an impact on the legal recognition of the right of access to information as a fundamental human right globally. Still emerging as an essential part of the right of access to information is obligation on governments to publish information proactively, which is reviewed in Section 4.3.

In the past four years, the right of access to information has been linked by international human rights tribunals to the right to freedom of expression. The impact of this linkage for the open government data movement and the right to reuse government data is examined in Section 4.4

Whilst many elements of the right of access to information have now been firmly established (Section 4.5) there are still a number of future challenges. Three main questions yet to be settled are: who owns public sector information, whether there is a right to information in machine-readable or open file formats, and whether there is a right to reuse information released under access to information laws. All of these questions are pertinent to the challenges facing the open government data movement.

---

[108] Right2INFO.org, Good Law and Practice, edited by Sandra Coliver, Open Society Justice Initiative, www.Right2INFO.org, page on Constitutional Provisions. This website contains sections on the access to information provisions of national constitutions, details of national laws, and relevant comparative analysis about the right to information.

[109] Constitution of Poland, 1997, at Article 61, http://www.sejm.gov.pl/prawo/konst/angielski/kon1.htm

The activities carried out by information advocates will vary according to the current status of access to information norms in their countries. Typically, such activities include advocacy to promote the adoption of access to information laws, monitoring of the application of the laws, litigation to promote compliance with the right to information, and technical assistance to government bodies to encourage greater transparency in practice.

Many organisations engaged in promoting access to information work closely with other user communities such as civil society organisations, journalists, and lawyers, and provide them with guidance on how to file requests as well as offering legal support to challenge refusals to provide information. In addition, a significant number of access to information organisations are engaged directly in requesting information, particularly when they are conducting monitoring work in areas such as human rights defence, democratisation and good governance, the fight against corruption, or environmental protection.

The access to information organisations surveyed a relatively low level of engagement with the open government data movement and many expressed unfamiliarity with recent developments and with the technical terms. Nevertheless, some access to information groups have undertaken activities which directly relate to the goals of the open government data movement. Examples include:

- In the **United States**, the National Security Archive has litigated successfully for access to records in electronic format, which is now enshrined in the US FOIA.

- In **Israel**, the NGO Movement for Freedom of Information has successfully litigated for the right to get access to budget information in an Excel or similar machine readable format. The ruling from the Tel-Aviv district court in early July 2010 ordered the Tel Aviv municipality to publish its budget in machine readable format, starting with the 2011 budget, rather than in the current PDF format. Following this campaign, the main state budget was released in XLS format and the civil society groups have held meetings with a leading financial newspaper to launch a joint visualization project focused on the budget.

- In **Argentina**, the NGO "*Poder Ciudadano*" (Citizen Power) which has long campaigned for the right of access to information, launched in 2009 a project called "Dinero *y Política*" (Money and Politics) which allows people to access the reports submitted by Argentinean political parties on their expenses during the 2009 legislative election campaign. This information was made available in a way that could be freely accessed, used, and distributed by anyone.[110]

In addition, a large number of ATI CSOs expressed an interest in learning more about the open government data movement and in having training on the specific issues in order to have a better grasp of this newly emerging field. Such requests came in particular from a number of countries in Europe (including Russia and South-East Europe) and Latin America.

[110] Poder Ciudadano has also joined the University of San Andres and The Garage Lab group to organize an event to preset some new independent initiatives in Argentina. Check: http://garagelab.tumblr.com/ http://www.udesa.edu.ar/Unidades-Academicas/Centros/Centro-de-Tecnologia-y-Sociedad/eventos?eid=3377

Box B

The Access to Information Programme, Bulgaria

An example of a typical access to information NGO is the Access to Information Program from Bulgaria. Founded in 1996 it was the first civil society organisation in the new democracies of central and eastern Europe dedicated to promoting the right of access to information.

Formed by journalists, lawyers, sociologists, and economists working in the area of human rights, AIP campaigned for the adoption of the law (in 2000) and monitors its implementation. They have successfully engaged in campaigns to block amendments which would weaken the law, including when the transposition of the EU's Directive on Re-use of Public Sector Information threatened to introduce a requirement that requestors give reasons for asking for information.

AIP runs a network of journalists in 26 cities around Bulgaria providing support to information requestors. They have held over 600 trainings on access to information issues, attended by representatives of NGOs, journalists and students, as well as a total of 7000 civil servants. In addition to over 50 publications, AIP has appeared in over 2120 articles in the printed media and has participated in 1430 radio and TV broadcasts.

As a result of AIP's activities, public awareness about the right of access to government held information has increased from 3% in 2000 to 39% in 2010, and Bulgaria ranks second in the world in terms of number of requests per capita among 86 countries with effective access to information laws.

AIP has provided legal advice in more than 4000 access to information cases and has taken over 150 cases to court, helping citizens, journalists, NGOs and business receive thousands of documents, including access to previously classified documents which revealed who murdered Bulgarian dissident journalist Gyorgy Markov with a poisoned umbrella in London in 1978.[111]

AIP has successfully advocated free access to a number of important registers and databases, including the securing publication on the internet of the Trade Companies Register, the State Gazette, the Register of Property Owned by High Government Officials, the minutes of the government sessions, and government contracts with private companies.

Members of the team have presented AIP's experience at meetings and forums in: Azerbaijan, Albania, Argentina, Armenia, Belgium, Bosnia and Herzegovina, Chile, Croatia, the Czech Republic, Denmark, France, Georgia, Hungary, Italy, Kazakhstan, Macedonia, Mexico, Moldova, Montenegro, New Zealand, Nigeria, Poland, Rumania, Spain, Slovakia, Slovenia, Serbia, the Netherlands, Turkey, Ukraine, United Kingdom, USA, and Vietnam.

[111] As a result of information obtained following litigation on access to information requests, journalist Hristo Hristov was able to write two books about the subject: *Kill the Tramp (2005)* and *The Double Life of Agent Piccadilly (2009)* which resulted in the police investigations being reopened in the UK and Bulgaria. See http://hristo-hristov.com/

4.2.1 Sectoral Transparency Movements

This core right-to-information movement has over time forged links to sectoral transparency movements in a such a way that thematic actors have developed strong right to information skills and regularly make use of the relevant legal provisions to access information. At the same time, access to information groups have acquired expert knowledge on these thematic areas. The main thematic movements linked to the access to information community are:

- **Environmental**: The right of access to information is most extensively recognised when it comes to environmental information. In many countries citizens have a legal right to request environmental information even where laws do not exist to cover other types of information, as it is the case in Spain. Furthermore, in the environmental sector this right is generally wider in scope, not only recognising a right to request information but also a right to make use of that information to participate in the decision-making process. A special obligation is also put on governments to proactively collect certain classes of environmental information. These advances have largely been due to the successful campaign for the Aarhus Convention (Aarhus Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters)[112] which continues to be a well used tool within the environmental community. Spearheading this movement is the NGO the Access Initiative[113] which has chapters in many countries worldwide and which maintains a dedicated focus on access to environmental information.

- **Budget Transparency:** The budget transparency movement calls for governments to proactively publish core information that is in the public interest, namely proposed budgets and accounts of actual spending. The International Budget Partnership[114] is an important NGO dedicated to promoting budget transparency which works with organisations all over the world to monitor national budgets and produce a budget transparency index. Every two years, the IBP publishes the Open Budget Index report on comparative levels of budget transparency; the next report due in late 2010 surveys over 90 countries worldwide following research conducted by national budget transparency and access to information organisations.

- **Anti-Corruption:** Access to Information is proving to be a key tool in the fight against corruption. As a consequence, many anti-corruption organisations have significant programmes dedicated to advocating for robust access to information regimes. Many chapters of Transparency International[115], for example, actively work on promoting access to information laws and subsequently use these laws to request information to expose and prevent corruption.

- **IFI Transparency:** The movement to promote the transparency of international organisations has focused on the international financial institutions (such as the World Bank and the regional development banks). This movement has been led by a coalition of civil society organisations from around the world, who formed the Global

[112] Aarhus Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters, Adopted at Aarhus, Denmark, 25 June 1998, http://www.unece.org/env/pp/documents/cep43e.pdf, last accessed 28 March 2010.
[113] See http://www.accessinitiative.org/
[114] See International Budget Partnership at http://www.internationalbudget.org/
[115] Transparency International http://www.transparency.org/

Transparency Initiative[116] to press for IFIs to adopt transparency rules akin to national access to information laws. This movement has achieved a number of notable successes, including persuading IFIs to switch from a paradigm of presumption of non-disclosure to presumption of openness (as was the case, for example, with the Asian Development Bank in 2005) or, perhaps most significantly, convincing the World Bank to adopt a disclosure policy which came into force on 1 July 2010 and which for the first time gives a full right of access to information to World Bank documents by means of information requests, a significant change from the previous policy which relied on proactive disclosure by the Bank.[117]

- **Aid Transparency:** There is widespread recognition of the need for reform of international development aid and in particular of the need to make international aid more transparent. For this reason the aid community has turned in recent years to access to information, emphasising that access is a right, not just a preference or an option. The movement for aid transparency is currently focused on the development of the International Aid Transparency Initiative,[118] a multi-stakeholder initiative to develop a voluntary standard which, once adhered to, would oblige all participating donor organisations to publish information in a way that makes it easy to access and free for all to use; participating NGOs include aidinfo and Publish What You Fund. As with the budget transparency part of the access to information movement, aid transparency activists place emphasis on the proactive dimension of the right of access to information as well as the legal right to access information across country borders. Aid activists insist on the importance of stakeholders in recipient countries accessing information from donor countries in order to permit their full participation in decision making on development.

- **Journalists and the right to information:** Journalists are regularly engaged in campaigns to promote access to information laws and in due course tend to become a key user community of these laws. In parallel with defending the right to protect sources, particularly when these are whistleblowers inside government, journalists regularly call for more information to be made available via legal means such as in responses to access to information requests in order to avoid the need to use information which is leaked from secret sources. In the current "war on terror" climate, both freedom of expression and freedom of information are under pressure, and there is increasing cooperation between freedom of expression, media freedom, and journalists' rights organisations, on the one hand, and access to information advocates, on the other hand. One project which helps journalists make use of access to information laws is the Legal Leaks initiative.[119]

---

[116] See Global Transparency Initiative at http://www.ifitransparency.org/

[117] The information note from the World Bank about this new policy recognises that this is a significant shift to a presumption of openness by the Bank, and will result in significantly greater transparency: http://web.worldbank.org/WBSITE/EXTERNAL/NEWS/0,,contentMDK:22635372~pagePK:64257043~piPK:437376~theSitePK:4607,00.html. The Bank states the policy was based on the US and Indian access to information laws, although in fact the campaign for this new policy promoted principles based on a much wider set of comparative standards, best practices, and lessons learned. The Bank itself held consultations in 33 countries before adopting the policy. An independent oversight panel is another key feature of the new transparency regime.

[118] See International Aid Transparency Initiative at http://www.aidtransparency.net/

[119] See the Legal Leaks initiative run by Access Info Europe and n-ost, at http://www.legalleaks.info/

The FOIAnet, the international network of freedom of information specialists, also has activists from each of these issue-specific areas as member organisations in the network and actively participating in its discussion list.

Activists engaged in these thematic social movements are often interested in bulk access to large government datasets and are therefore linked to the demands of the open government data movement. For example, aid transparency organisations have worked with open government data experts to identify and develop policy recommendations on how to achieve greater transparency of information about aid flows.[120]

4.2.2 National and International Actors Promoting Access to Information

As the right of access to information has become established over the course of the past two decades, a number of other key actors at the national and international level have engaged in the promotion and defence of this right. These include information commissioners, special rapporteurs appointed by international human rights bodies, and inter-governmental organisations working on democratisation and development.

- **Information Commissioners**: These independent bodies which oversee implementation of national access to information laws are increasingly engaged in the global movement to promote and defend the right of access to information. They promote awareness and use of the right, receive appeals from citizens when their information requests have been turned down, and monitor compliance with the law, compiling statistics and making recommendations to improve transparency.

  In many cases Information Commissioners are empowered to conduct searches of public bodies in order to verify whether or not they hold the requested information, and to sanction public bodies who do not comply with the law, either by refusing to release information following an order by the Information Commission or by failing to publish proactively the information that by law they are required to make automatically available.

  In an increasing number of countries, such as Germany, Hungary, Mexico, Slovenia, and the UK, information commissioners have the dual responsibility of overseeing both the access to information and the data protection laws. In these cases they often develop specific expertise on one of the most debated areas of information law: the balancing of the right to information and the right to privacy. This model of oversight body has proved particularly effective in dealing with overlaps between the two laws.

  Another oversight model is that of appeals to national ombudsmen. Generally the opinions of ombudsmen are not binding and no sanctions can be issued, but their decisions are nevertheless influential in countries such as Sweden and Norway and in many countries around the world the ombudsman's office is an important ally in promoting the right of access to information.

- **Special Rapporteurs**: A number of inter-governmental bodies with a human rights mandate have appointed special rapporteurs whose role is to monitor respect for the right of access to information in their region of competence. The bodies which have such special rapporteurs are the United Nations (Special Rapporteur for Freedom of Expression), the Organisation for Security and Cooperation in Europe (Representative

[120] See the report Unlocking the Potential of Aid Information, by the Open Knowledge Foundation and aidinfo, at: http://www.unlockingaid.info/

on Freedom of the Media), the Organisation of American States (Special Rapporteur on Freedom of Expression) and the African Commission on Human and People's Rights (Special Rapporteur on Freedom of Expression and Access to Information).

As can be seen from the names of these special mandates, they link defence of the right of access to information with the right to freedom of expression. These individuals are empowered to investigate cases of abuse of the rights they oversee, to report to their respective IGOs, to raise concerns with governments, to participate in efforts to promote the rights and to engage in standard-setting at the international level.[121] The special mandates have been influential in securing greater recognition for the right of access to information as a fundamental human right.

- **Inter-governmental organisations**: Key international organisations such as the Council of Europe, World Bank, OSCE, and United Nations Development Programme have been engaged in promoting democracy and human rights and fighting against corruption have been actively engaged in the movement to promote the right of access to information over the past 15 years (since the mid 1990s). They have conducted standard-setting exercises, supported government and civil society programmes to defend the right of access to information, and worked with activists operating at the national level to encourage governments to adopt and implement access to information laws. The Council of Europe, for example, developed for governments (such as the 2002 Recommendation on Access to Official Documents) and then the world's first binding treaty, the Convention on Access to Official Documents, which opened for signature in 2009. The OSCE has been active in research and standard-setting, and in some countries has directly supported governments and experts in drafting access to information legislation, thereby helping to ensure a democratic debate around the main features of future access to information laws. The World Bank Institute has played a key role in advancing the empirical and conceptual development of the right of access to information, supporting forums for debate and think pieces by leading specialists in the field. The World Bank and the UNDP support the provision of technical assistance to governments in order to make the right to information work as a reality in practice.

In addition to these activities, as noted above, inter-governmental bodies are increasingly a target for the access to information activists who are calling for them to practice what they preach so as to increase their internal transparency.

The intergovernmental, governmental, and independent bodies listed above, as well as access to information advocates and civil society organisations from other sectors, are all potentially strong allies for open government data activists when trying to reform policy and practice in a particular country. To tap this potential support, there is a need to inform and educate these actors, for example by inviting them to participate in relevant events, in order for them to become acquainted with and to engage in the open government data debate.

4.3 The Right to Know and Proactive Transparency

As noted in the introduction, the right of access to information places two key obligations on governments. The first is the obligation to respond to requests for information from the public,

[121] For an example of the work of the Special Rapporteurs, see their annual declarations, the 2009 declaration can be found at: http://www.article19.org/pdfs/press/international-special-rapporteurs-for-free-expression-highlight-critical-ten.pdf and focuses on the Ten Key Challenges to Freedom of Expression in the Next Decade, which includes "The failure by a majority of states to adopt laws guaranteeing the right of access to information, and the weak implementation of such laws in many states which have."

also known as *reactive disclosure* of information. The second obligation is to make information public at the initiative of the public body, without a request being filed, which is known as *proactive disclosure.*[122]

Proactive transparency can be achieved using a multiplicity of means, ranging from institutional publications and official gazettes, to publicly accessible notice boards, to radio and television announcements, to posting on the internet via a public institution's website.[123] Of particular relevance to the open government data movement is proactive disclosure of entire datasets via the websites of public bodies.

The standards for proactive disclosure are still in development. If the only channel for members of the public to enjoy the right to information were via requests filed by individuals, then huge information inequalities would rapidly arise with different people knowing different things about the functioning of government. Furthermore, in this scenario, large sections of the population would remain ill-informed, to the detriment of society as a whole. Such a regime would also place an intolerable burden on public officials who would have to strive to answer huge volumes of requests from information-hungry citizens. Hence proactive disclosure ensures that all members of the society are able to enjoy the right to information on a more equal footing.

In countries with access to information regimes, proactive disclosure has the benefit of reducing the burden on public administration of having to process requests for information that may be filed under an access to information law. There is a body of evidence which suggests that proactive disclosure encourages better information management and hence improves a public authority's internal information flows, thereby contributing to increased efficiency.

For members of the public, the automatic availability of information means timely access to information and hence reduces the need to file information requests. Additionally, in countries still emerging from authoritarian regimes or where corruption is widespread, proactive disclosure permits anonymous access[124] and so gives some protection to applicants from weaker segments of society who might not feel comfortable writing to government bodies to ask for information for fear of repercussions.[125]

In the information era and provided that a number of technical and legal conditions are met, proactive disclosure can also place information at the public's disposal in a digital format. This paves the way the public to take the information and to do interesting and innovative things with it. The proactive disclosure of data has already been proved to have a positive social and

---

[122] Proactive disclosure is also known as *active disclosure* (this term is used in the United States for example) and as *suo moto* disclosure from the Latin for "upon its own initiative" (this term is used, in particular, in India).

[123] Proactive disclosure by public bodies is distinct from the obligations that government places on private bodies to disclose information, such as nutritional labelling on food, warning notices with medicines, financial reporting by companies to shareholders, or car safety ratings. These are referred to as "targeted transparency" requirements by researchers Archon Fung, Mary Graham, and David Weil at the Harvard Kennedy School of Government's Transparency Policy Project. See http://www.whitehouse.gov/files/documents/ostp/opengov_inbox/opengovernmentcommentsfinal.pdf, and more generally www.transparencypolicy.net.

[124] In some countries, such as Mexico, information requesters have the option of filing requests anonymously, but this is an unusual option. In India it was argued that anonymous requests would detract from the incentive to mobilise people against corruption.

[125] Examples would be public officials who should provide information on tenders to companies that ask for them or provide subsidies to farmers, but who withhold information in order to control the allocation of resources.

economic impact, as is discussed further in Section 5. Hence proactive disclosure is an essential component of the right of access to information.

In spite of the strong rights-based and practical benefits arguments in favour of proactive disclosure, the standards as to what exactly should be disclosed are still emerging. Historically proactive disclosure has been limited to certain fields of government activity such as laws, regulations, decisions, summaries of budgets and spending, and information about government services and funding opportunities; yet much other detailed data remained out of reach. Even with the advent of the internet, public bodies tend to regard their websites as shop windows for promoting their services and achievements, through short reports and press releases, rather than letting look directly at the raw data about the day-to-day functioning of government.

That is now changing because of the greater demand from a number of sectors of civil society, including those working in particular thematic areas (environment, health, anti-corruption, and participatory budgeting, for example), for access to complete data sets. Access to information advocates supporting these communities are increasingly getting involved in debates about large-scale proactive disclosure from a rights-based perspective, at precisely the same time as the open government data movement is making a series of social and economic arguments to support its calls for large scale disclosure of raw data in the hands of government.

Some of the newer access to information laws do contain extensive chapters on proactive disclosure and the obligation to release information proactively is recommended in a number of international standards.[126] From these norms it is possible to argue that there exists an obligation to proactively publish information about at least the core functioning of any particular public body. Typical classes of information for proactive disclosure include Organisational and legal information about the public body as well as operational information such as plans, policies, activities, procedures, reports, and evaluations; decisions and acts; descriptions of services offered to the public; and information on open meetings and on mechanisms for consultations and public participation in decision-making. Financial information should include: budget information including the projected budget and actual income and expenditure; information on subsidies and grants issues; and detailed information on public procurement processes and on contracts awarded.[127] More standard-setting work is needed, however, and there is an opportunity for open government data advocates to work with access to information experts to define further the scope and nature of the proactive disclosure obligations.

4.4 Freedom of Expression and Access to Information

International law protects the right of access to information as an integral and intrinsic component of freedom of expression. This should mean that when information is obtained through the exercise of the right of access to information, people are free to use it without further constraints. This section examines the link between the right of access to information and freedom of expression, and considers the relevance of this connection for the open government data movement.

---

[126] For a detailed analysis of the international standards see "Proactive Transparency: The future of the right to information? A review of standards, challenges, and opportunities by Helen Darbishire, published by the World Bank Institute, July 2009, available at: http://siteresources.worldbank.org/WBI/Resources/213798-1259011531325/6598384-1268250334206/Darbishire_Proactive_Transparency.pdf

[127] The paper referenced in footnote 24 gives detailed analysis of the classes of information for proactive disclosure.

The right to freedom of expression has long been regarded as one of the most fundamental rights in any democratic society. The French Declaration of the Rights of Man and the Citizen[128] (1779) asserted that the right to "free communication of ideas and of opinions" is "one of the most precious rights". As Article 19 of the UDHR makes clear, the right includes seeking, receiving and sharing information:

> *Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.*

No international treaty or constitution explicitly states that people have the right to access government data in order to use it to create social and economic value, for example by mashing it up with data obtained from other sources and by applying information design techniques. Nevertheless, international human rights tribunals have ruled that access to information is implicit in this universal guarantee of the freedom to disseminate information as part of the right to freedom of expression.

The two leading cases which resulted in formal recognition of access to information as a human right by international human rights tribunals came from eastern Europe and Latin America. It was in these two regions that the efforts of the democratisation movements of the 1990s and early 2000s had secured a strong legal and constitutional basis for the right to information, thereby creating a context in which the international human rights system was able formally to confirm a fundamental human right of access to government-held information.

The key cases are *Claude Reyes vs. Chile* and *TASZ vs. Hungary*. Both were cases taken to the international human rights tribunals by representatives of civil society organisations who needed to access information in order to participate in public debate on matters of public importance: an environmental protection campaign in the Chilean case, and debate about a new drugs law in Hungary.

In the Chilean case, a request for information about foreign companies investing in logging native forests had met with silence. The Inter-American Court of Human Rights issued a clear and strong decision on 19 September 2006 in which it found that the Chilean government had violated the right of access to information as protected by the freedom of expression and information provision of the American Convention on Human Rights, Article 13.[129] The decision states in its key paragraph 77:

> *... the Court finds that, by expressly stipulating the right to "seek" and "receive" "information," Article 13 of the Convention protects the right of all individuals to request access to State-held information, with the exceptions permitted by the restrictions established in the Convention. Consequently, this article protects the right of the individual to receive such information and the positive obligation of the State to provide it, so that the individual may have access to such information or receive an answer that includes a justification when, for any reason permitted by the Convention, the State is allowed to restrict access to the information in a specific case.*

---

[128] The Declaration of the Rights of Man and the Citizen of 1979, Article 11, English translation at: http://www.diplomatie.gouv.fr/en/france_159/institutions-and-politics_6814/the-symbols-of-the-republic-and-bastille-day_2002/the-declaration-of-the-rights-of-man-and-the-citizen_1505.html

[129] Case of Claude Reyes and others v. Chile, see http://www.corteidh.or.cr/casos.cfm?idCaso=245 (Spanish original) and http://www.corteidh.or.cr/casos.cfm?idCaso=245&CFID=525202&CFTOKEN=97319768 (English).

The court also made clear that no particular interest needed to be proved by the requestor of the information, the public interest in access and dissemination being sufficient:

> *The information should be provided without the need to prove direct interest or personal involvement in order to obtain it, except in cases in which a legitimate restriction is applied. The delivery of information to an individual can, in turn, permit it to circulate in society, so that the latter can become acquainted with it, have access to it, and assess it. In this way, the right to freedom of thought and expression includes the protection of the right of access to State-held information, which also clearly includes the two dimensions, individual and social, of the right to freedom of thought and expression that must be guaranteed simultaneously by the State.*

This decision, which also ordered the Chilean government to take the necessary measures to ensure respect for the right, including through training of public officials, had a significant impact in Chile, which in 2008 incorporated the right into its Constitution at Article 8, and adopted an access to information law that entered into force in April 2009.

In Europe, the case before the European Court of Human Rights revolved around a request submitted by the Hungarian Civil Liberties Union to Hungary's Constitutional Court for access to a parliamentarian's complaint questioning the constitutionality of criminal legislation concerning drug-related offences. The Constitutional Court refused to release the information arguing, somewhat bizarrely, that it was the personal data of the parliamentarian, and with no higher instance to turn to, the applicants appealed to Strasbourg.

In its ruling of 14 April 2009[130], the European Court of Human Rights dismissed the privacy argument, noting that it seemed "*quite implausible that any reference to the private life of the MP, hence to a protected private sphere, could be discerned from his constitutional complaint.*"

The Court underlined that it would be "*fatal for freedom of expression in the sphere of politics if public figures could censor the press and public debate in the name of their personality rights*" and that such arguments could not be called upon to justify the restriction on access to information and consequent interference with freedom of expression as protected by Article 10 of the European Convention on Human Rights.

The Court argued that when a public body holds information which is essential either for the media to play their role as "public watchdogs" or for civil society to play a "social watchdog" function, then to withhold that information is an interference with freedom of expression. The judges achieved this paradigm shift by arguing that when a public body holds information and refuses to release it, it is exercising the "censorial power of an information monopoly" and hence should have supplied the information to those requesting it.

The existence of the right to information protected by Article 10 of the Convention was confirmed on 26 May 2009 by second ruling of the European Court of Human Rights, once again in a case against Hungary. This second case was brought by an historian, János Kenedi, who had applied for access to historical documents about the functioning of the Hungarian State Security Service, had been granted access by the Hungarian Courts, but had not been provided with the documents by the Ministry of Interior.

---

[130] *Társaság a Szabadságjogokért v. Hungary* (App no 37374/05), ECHR, 14 April 2009, paragraph 36, http://cmiskp.echr.coe.int/tkp197/view.asp?action=html&documentId=849278&portal=hbkm&source=externalbydocnumber&table=F69A27FD8FB86142BF01C1166DEA398649.

In its ruling, the European Court of Human Rights noted that the Hungarian Government accepted that there had been an interference with the applicant's Article 10 rights and confirmed that "access to original documentary sources for legitimate historical research was an essential element of the exercise of the applicant's right to freedom of expression."[131]

The linkage of the right of access to information to the right to freedom of expression is relatively recent and its full implications for the national legal structure still need to be worked out, particularly in those countries where the constitution does not already make that connection. Nevertheless, it is clear that the right of access to information entails both the right to access information and to use it without limitations, save for those few exceptions which are actually permitted by international treaties as acceptable limitations on the right to freedom of expression.[132]

4.5 Key elements of the right to information

The global movement to promote the right of access to information has achieved commonly-agreed standards for this right. These standards have been agreed upon by the main international human rights standard-setting bodies and are reflected in the strong commonality of content of national access to information laws around the world. It is therefore possible to map out the key elements of the right to information. These include:

- **The right applies to all public bodies**: The right of access to information should apply to all branches of state power, including the executive and administrative bodies, the judicial and legislative branches, as well as private performing public functions or operating with public funds.

- **Everyone has a right of access:** Being a fundamental human right, everyone should have a right to access information held by public bodies in all countries without discrimination. This means that not only citizens and residents but anyone can file an access to information requests in any country. Most access to information laws permit this although language is an obstacle as it is usually necessary to file requests in the language of the country.

- **The right applies to all information**: All information held by all public bodies is subject to the right of access to information, although limited exceptions may apply to release of some of that information. Whilst it is the case that some countries have "access to information" regimes and others have "access to documents" regimes, there is relatively little difference in practice. One area where these different definitions can cause problems is in relation to a right of access to entire databases. In a handful of countries, particularly those with older access to information laws, a database is not considered to be a document and is therefore cannot be requested under access to information laws. (See also Section 2.2.1 on Electronic Formats and Access to Databases for further consideration of this issue).

- **Short timeframes for access.** Most access to information laws require "prompt" or "rapid" response to information requests and set a maximum timeframe for providing

[131] *Kenedi v. Hungary* (Appl. no. 31475/05)

[132] The International Covenant on Civil and Political Rights at Article 19, establishes that the rights to freedom of expression and information may be limited only where the restrictions "*are provided by law and are necessary: (a) For respect of the rights or reputations of others; (b) For the protection of national security or of public order (ordre public), or of public health or morals*."

information. The global average for responding to access to information requests is about 15 working days, although access to information laws range from immediate access to 30 calendar days. In exceptional and complex cases, extensions are allowed, but these usually do not exceed 1-2 months and must be clearly justified in writing.

- **Access should be free of charges or low cost**: Filing access to information requests should always be free of charge. The prevailing standard is that submitting information requests is free of charge as is the inspection of original documents and the receipt of information by electronic means. The norm is that the only legitimate charges that may be made are for providing copies of information (photocopies, copies on discs) and for delivering these copies (postal charges). This is confirmed by the Council of Europe Convention on Access to Official Documents which clearly states that: "*Inspection of official documents on the premises of a public authority shall be free of charge. ... A fee may be charged to the applicant for a copy of the official document, which should be reasonable and not exceed the actual costs of reproduction and delivery of the document."* This raises the question of whether it is ever legitimate for public authorities to charge for access to and use of information created with public funds, as discussed further in Section 3.3 on Reuse of Public Sector Information.

- **Exceptions:** There are typical exceptions for access to information which are common in many countries and which are established in the freedom of expression and information provisions of binding international treaties. These commonly accepted exceptions include:
  - *Exceptions to protect state interests*: such as national security, international relations and protection of public order as well as protection of monetary policies.
  - *Protections aimed at ensuring effective government*: a) protection of internal deliberations within public authorities prior to decision-making –this is known as the "space to think" exception; b) protection of criminal investigations.
  - *Exceptions to protect private interests, human rights and other rights*: privacy protections, protection of legitimate commercial and economic interests, protection of the environment, and protection of the equality of parties in court proceedings.

  If properly applied, this broad set of exceptions will only exclude from disclosure a reasonable small percentage of all the information held by public bodies. Furthermore, even when a document contains some sensitive data which might fall under one of the exceptions, the public body must provide partial access to the remainder, which in some cases can amount to the majority of the information in the document.

  In addition, public bodies have to apply a public interest test so that even if some information potentially falls under one of the exceptions, it should be released if there is an overriding public interest in it. Determining the public interest has to be done on a case-by-case basis and should take into consideration the public scrutiny of government that characterizes functioning democracies. Hence, for example, information relating to the spending of public funds should almost always be provided to the public. This is explained in greater detail in Section 3.1 on Exceptions to the Right of Access to Information.

4.6 Future Challenges for the Right of Access to Information

The right of access to information as it has developed to date still has a number of shortcomings from the perspective of both open government data activists and access to information advocates. These challenges include a number of potential elements of the right of access to information which are not yet clearly established as forming part of the right.

- **Who owns public sector information?** There is a long tradition of public bodies asserting ownership and copyright rights over the information they produce. This seems to be at odds with modern democratic principles and with the "principle of publicity" statement in many access to information laws, which establishes that information is presumed to be public unless an exception applies. Given that information is produced by public officials for and on behalf of the people, using taxpayer's money, the assertion of copyright or other private property ownership rights over the information seems to be an anachronistic legacy which is hard to justify in twenty-first century democracies.

  This issue is not clearly resolved in the access to information regimes of many countries. As a result, right to information advocates around the world have to deal with and work to change the extensive perception of private ownership of the information by bureaucracies internal to all branches of government. This matter is a priority for standard-setting strategies which will advance the open government data and access to information agendas.[133]

- **Right to electronic access?** Many access to information laws give requestors the right to define the preferred format in which they would like to receive information. This includes an electronic format, if available. The Council of Europe Convention on Access to Official Documents also confirms this right: "*the applicant has the right to choose whether to inspect the original or a copy, or to receive a copy of it in any available form or format of his or her choice unless the preference expressed is unreasonable*." Some laws, however, particularly those written before the internet age, don't include this option, for example the Swedish law (Section 2). This is an issue which the right to information community still has to resolve, and is of great relevance for the open government data community given its call for access to data in digital formats.

- **Right to information in machine-readable format?** Most access to information laws and the Convention on Access to Official Documents are silent on this (See Section 2.2.2 on machine readable formats). Further mapping of the law and, specifically, of practice is needed, and campaign by access to information advocates could usefully be developed to secure a right to information in machine-readable format wherever practicable. The campaign goal should be to promote policies and eventually laws which oblige governments to provide information in a variety of formats, including both readable by humans and processable by machines.

- **Right to information in open file format?** The research for this report did not find a single access to information law or international standard document relating to the right to information which considers this issue. A number of government policy statements linked to the open government data movement do make this commitment (See Section 5). The challenge for the open government data and access to information communities

[133] A useful reference might be the law and jurisprudence in some countries which denies public bodies the right to use for defamation, something which private individuals and private legal bodies have a right to do. The arguments on points of principle about the nature of public bodies are very similar.

is to promote government policies which encourage public bodies to use open file formats.

- **Right to reuse information?** The fact that the right of access to information is linked to the right to freedom of expression indicates that once information has been received from a government body, members of the public should be free to make use of it in any way they desire. This seems to be confirmed by the Council of Europe Convention on Access to Official Documents. Its Explanatory Memorandum asserts that information received under access to information can generally be used for any lawful purpose, including disseminating the information and publishing it.

  However, the Explanatory Memorandum also contradicts itself by stating that reuse shall be determined by other laws, "*such as those regulating intellectual property or data protection or transposed by the Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information*."[134]

  In Access Info Europe's analysis, there is a serious clash of principle on this regard: if the right of access to information is intrinsically linked to the right to freedom of expression, then this right should include not only access to the information but the right to reuse it in any way that a member of the public chooses. Such a right should not be constrained by copyright restrictions or by other licenses or charges for reuse of the information received.[135]

  It is unclear whether the international human rights system permits limits on the reuse of information created by public bodies for purposes of protecting the intellectual property rights of public bodies. It is even more debateable whether bodies funded by taxpayer's money have a right to charge for reuse of information which, in principle, can be accessed using access to information mechanisms. Although there are no specific international standards or jurisprudence on these potential limits on reuse of information, the existing standards and rulings make clear that information is not obtained from government bodies merely for the benefit of the knowledge of the requestor, but rather in order that that person can make full use of the information in exercising his or her freedom of opinion and freedom of expression rights, which includes the right to "*impart information and ideas through any media".*

  The realities in practice are different, as will be explored further in Section 3. This unresolved policy and rights issue is being laid bare by the open government data movement, which is pushing for greater access to large volumes of government data and the right to reuse it in any way that members of the public so choose. Further standard-setting work in this area is urgently required.

4.7 Recommendations for the Access to Information Movement

The right of access to information is sufficiently well established that access to information laws can usefully serve the open government data movement, helping those who wish to reuse information secure access to it. There nevertheless remain a number of outstanding shortcomings which could usefully be addressed by experts in access to information.

---

[134] For the Explanatory Memorandum to the Convention, see http://conventions.coe.int/Treaty/EN/Reports/Html/205.htm

[135] One of the exceptions to the non-restrictive copyright principle is when the information held by a public body was received from a third party which holds copyright in it. This would be the case, for example, of a book or DVD purchased by or donated to the public body.

➢ Recommendations

It is recommended that access to information organisations:

- Develop standards on the proactive dimension of the right of access to information to secure stronger obligations on public authorities to release core datasets without the need for information requests;
- Conduct further research and policy development on the link between access to information and freedom of expression in order to construct arguments about the right to reuse information obtained from public bodies;
- Research the legal framework country by country to identify current restrictions on reuse of information based on copyright, database rights, or other assertion of ownership rights over public sector information;
- Identify legal and policy solutions which will ensure that information released to the public can be used without restriction;
- Conduct research and engage in standard-setting with a view to:
  - confirming that the right to information/documents includes a right of access to entire databases or datasets held by public authorities;
  - ensuring that there is a right of access to documents in electronic format whenever they already exist in that format;
  - developing a right of access to government information in machine-readable and open formats wherever practicable in order to maximise reusability of the information;
- Collaborate more closely with developers and open data activists so that CSOs working on the legal aspects of the ATI field are up to speed on the technical issues;
- Engage information commissioners, special rapporteurs and inter-governmental organisations in these policy debates and standard-setting activities.

5: The Open Government Data Movement

New digital technologies enable public bodies to release the data that they hold rapidly, inexpensively and *en masse*. Digital technologies also make it possible to represent this information in innovative ways, combining it with data from other sources to deepen public understanding of the activities of government.

The social and commercial potential of opening up public datasets has strengthened calls for information held by public bodies to be disclosed in full and in open formats which facilitate its use, for example, by copying and disseminating it, by combining it with other material, or by republishing it. The benefits of open government data which are being used to argue for the right to reuse public sector information are examined in Section 5.1.

The open government data movement is a new generation of technically proficient, publicly-minded activists who are using their skills to build applications, often web-based, which process and represent visually large quantities of government data in ways that make it more comprehensible, for example, by connecting it with other sources of information. Some of the leading organisations which represent this new sector of civil society and the type of activities they engage in are mapped out in Section 5.2.

The successes of the first initiatives by the open government data movement have triggered a surge of interest in accessing and using government data. As technologists, journalists, and civil society organisations started to make use of data which had been obtained either because it was published proactively by public bodies or as a result of access to information requests, they demonstrated the tremendous social potential contained in government-held information. Examples of projects which show the way by making use of government data are described in Section 5.3.

In spite of the successes of this young movement, there remain many obstacles to accessing and using government data. These obstacles to reuse vary from country to country, but commonly include the need to pay for access to data created with public funds, the issuing of restrictive licences for the reuse of government databases, and the assertion of intellectual property rights over data collected by public bodies. Examples of campaigns which have successfully overcome these obstacles are given in Section 5.4 and details on how to overcome the technical and legal obstacles can be found above in Sections 2 and 3 respectively.

In addition to calling on public bodies to release information, there is a need to stimulate use of the data by civic hackers, civil society organisations, digital entrepreneurs and others. To this end, open government data activists organise events and competitions to provide incentives to potential users. These strategies and the types of activities organised are set out in Section 5.5.

This review of the open government data movement concludes with an analysis of how open government data advocates can work more closely with access to information activists to maximise the impact of campaigns for open government data. A series of recommendations have been developed, for civil society, for government and for funders on how to support this movement and, in particular, encourage open government data initiatives in a broader range of countries around the world.

5.1 Advocacy for Open Government Data

The open government data (OGD) movement has rallied a number of arguments in favour of opening up government data. These arguments stress the social and economic benefits which derive from allowing members of the public to use and add value to the already-existing information.

Some of these arguments about the positive potential of opening up government information overlap with those employed by the right to information movement to argue in favour of strong access to information laws.

The benefits of open government data as argued by the OGD activists include:

- **Transparency**. In a well-functioning democratic society citizens need to know what their government is doing. To do that, they must be able *both* to access government data and information *and* to share that information with other citizens. True transparency goes beyond access: information should be "open" in the sense that it is in a format and free of limitations on use so that it can be shared, combined, analysed, visualized, and distributed. An example is having access to budget spending data, to statistics, to contracts, to parliamentary voting records, to laws, or to court decisions. Transparency can only maximise its democratic value if citizens can make use of, discuss and distribute public information.
- **Innovation to release the social and commercial value of information.** Government can help drive the creation of innovative businesses and services that deliver social and commercial value by opening up public data**.** For example, geographic data about the location of schools and hospitals could be incorporated into a commercial service sold to prospective house buyers about services in the area they are thinking to live, but could also be used by associations of parents to map out the facilities available to their children and to develop a campaign for improving the organisation of those services. The example shows how public datasets can be freely combined to generate both social and entrepreneurial initiatives.
- **Collaboration**. Releasing government data creates a platform for collaborative projects between public bodies, private actors, civil society and the general public. Opening up data helps governments harness external expertise in order to improve public services. For example, by releasing fuller information about spending of donor aid funds, external experts and civil society organisations will be in a position to work with governments on joint projects to design strategies and to carry out the effective spending of these funds.

- **Participatory Governance**. Much of the time citizens are only able to engage with their own governance sporadically — maybe just at an election every 4 or 5 years. By opening up data, citizens have real-time information which allows them to be involved in decision-making more frequently basis, either informally or through structured consultations. Moving from transparency to participatory governance is about developing a full "read/write" society. For example, a department of health can hold consultations on its strategies for care of the mentally ill in order to ensure that they take account of the experiences of affected individuals and families. To do this effectively, more information about how these services are run should be released to the public.

- **More efficient governance.** Both the public and government itself have a clear interest in increasing the efficiency and effectiveness in the spending of public funds and the

operation of public bodies. With access to full datasets, the public can provide feedback to public bodies and, where necessary, hold them accountable for budgetary waste or plain corruption. For example, the release of information on the spending of government subsidies permits civil society, the media and the general public to play their watchdog role and identify problems such as inappropriate or illegal allocation of subsidies.

5.2 Organisations Promoting Open Government Policies and Practices

Advocacy work is normally conducted either by individual organizations or through campaigns that bring together a range of actors including CSOs, media organisations, individual programmers and academics. Examples of some of the leading open government data organisations from around the world which are promoting government policies and practices in favour of greater openness include:

- **The Open Knowledge Foundation** a UK-based not-for-profit organisation founded in 2004. The OKF is dedicated to improving the way information is shared in a wide variety of different domains. It has been working to promote open government data since 2005, and worked with the UK government to build the data.gov.uk registry. It hosts a working group on open government data, which is an international network for advocates, government representatives, policy makers, developers, and others. It is also engaged in policy work at a European level. In addition to organising events and community building, it undertakes work on a number of projects re-using official information such as 'Where Does My Money Go?' which aims to visualise the spending of the UK budget.[136]

- **The Sunlight Foundation**[137] is a US organisation working to secure greater government transparency and accountability. Founded in 2005, the Sunlight Foundation acts through lobbying, advocacy, and journalist training. They encourage and assist in the development of new policies to improve transparency, and aim to raise the expectations of citizens regarding the level of openness they demand from their government. They also develop software and online tools to enable easy access to and reuse of data. The Foundation is rooted in the belief so poetically expressed by Louis Brandeis that "sunlight is the best disinfectant" and that the power of the internet can and should be harnessed to the strengthening of democratic institutions by opening them up to public scrutiny. Successes have included the online and easily useable publication of roll call votes and expenditure reports as well as the reform of old rules that had previously made it illegal for members of Congress to use social networking sites such as Twitter and YouTube.

- **Iniciativa Chile Datos**[138] campaigns for open government data in Chile. The initiative was established in response to the adoption in 2008 of Chile's Law on Access to Public Information and the governmental ordinance of 13 April 2009 which sets out how public bodies should comply with the proactive disclosure obligations under that law. The initiative was inspired by the 8 Principles of Open Government Data (See Box A). This organisation focuses on the development of software to assist in the processing of data, and aims to make published data technically easier to use and combine in novel ways. Of particular importance to their campaign is ensuring that all data is published under

[136] See the Open Knowledge Foundation website: http://www.okfn.org/
[137] See the Sunlight Foundation website: http://sunlightfoundation.com/
[138] See Iniciativa Chile Datos website: http://chile-datos.degu.cl/index.html

licences which permit re-use and redistribution as the Chilean government continues to publish its data without any specific licensing information, implying a reservation of rights by the government.

- **Pro Bono Público**[139] in Spain is typical of a new generation of open government data organisations. It was founded in January 2010 by a group of programmers that aim to extract the social value out of public sector information. *Pro Bono Público* aims to promote a culture of openness and accountability in government, and foster a greater appreciation of the importance of open government data among citizens as well as the private and public sectors. Their work includes the development of online libraries, platforms, and other open source projects, driven by the conviction that new technologies have the potential to substantially alter the relationship between citizens and their governments. *Pro Bono Público* was involved in the organisation of Spain's first open government data competition, the 48-hour *Abre Datos* contest in April 2010.

- **The Open Data Network**[140] in Germany, also founded in 2010, is an initiative to promote fully open access to data from government, the public administration, and the scientific sector. Through a mixture of advisory and educational events, campaigning, and software production, the Open Data Network encourages citizens to press for open government. One aspect of their work is the development of tools for citizens wishing to use public data, such as the recent Mapnificent project, inspired by Mapumental (See Section 5.4 below) which visualises environmental data about Berlin in an interactive map. Through these activities, Open Data Network aims to promote the social and political engagement of citizens.

5.3 Projects which use government data

There are a number of civil projects which are seen as part of the open government data movement because they make use of datasets obtained from government, often to maximise the social and democratic benefits of government data outlined in Section 5.1. In doing so, these projects push for the goals of the open government data movement, such as to make information available in formats which facilitate reuse, even if such advocacy is not the primary focus of these projects. Some of these projects make use of data from intergovernmental organisations as well as from national governments, often combining them with datasets from other sources.

Leadings examples of these projects include:

- **TheyWorkForYou**[141] is an online tool designed to give detailed and user-friendly information about the activities of Members of the UK and Scottish Parliaments as well as of the Northern Ireland Assembly, TheyWorkForYou makes use of the official parliamentary record, called Hansard.[142] The Hansard transcripts were initially offered

[139] See Pro Bono Público website: http://blog.probp.org/
[140] See the Open Data Network website at http://opendata-network.org/
[141] See TheyWorkForYou website at http://www.theyworkforyou.com/
[142] The Hansard Reports are named after their first official publisher, Luke Hansard (1752 - 1828) who was the printer of the House of Commons Journal from 1774. http://www.parliament.uk/site-information/glossary/hansard-official-report/. Prior to 1771 it had been an offence to publish the proceedings of the House of Commons but at that time the Mayor of London refused to punish a publisher for doing so, see http://en.wikipedia.org/wiki/Hansard. These developments took place at around the same time (1766) that the Swedish Law on Freedom of the Press legalised the reprinting of parliamentary

as a subscription-only service, but as a result of the project the transcripts have become fully open although they still carry Parliamentary Copyright[143]. The project's website enables users to search the parliamentary transcripts, and to subscribe to email alerts, for instance when their member of parliament speaks, when chosen topics or keywords are discussed in Parliament, or when their local area is mentioned in those discussions. Additional data sets are presented used to allow members of the public to search their MP's voting record, expenses claims, frequently raised issues, and registered interests. The value of the project in allowing people to keep track of their elected representatives' daily work has lead to similar projects in Australia, Ireland and New Zealand and has influenced the development of parliamentary monitoring sites in France, Italy, Germany, and at a European Union level.

- **Mapumental**[144] is a mapping tool which helps users to decide where in the UK they might like to live, work, or visit based on a range of variables such as commuting time, house prices, and how scenic a place is. Combining open access maps from the application OpenStreetMap[145] with travel data purchased from Traveline[146], house price information from the Land Registry[147], and survey results from the purpose-made survey ScenicOrNot[148], users are thereby enable to make informed decisions based on factors which they deem important to their quality of life. It should be pointed out that while some of the data used in Mapumental is fully open, other data had to be paid for, and the developers note that the project would not have been possible without the support of Channel 4 to buy the data. The map does not include house pricing information for Scotland due to the prohibitively high price of acquiring this information.

- **Where Does My Money Go?**[149] gives the UK public an interactive overview of how their taxes are spent through use analysis by using analysis and visualisation techniques. The project was established in response to the realisation that the very abundance of published spending data was restricting the ability of citizens to get a real understanding of public finances. The developers saw the great potential of new media technologies to amass, analyse, and represent the data in ways that made it more useful. The project is based on the premise that citizens have a right to know where their taxes go. Furthermore by accessing and engaging with this information citizens can engage more effectively with in their political and social institutions, thereby strengthening those institutions.

  Additionally, *Where Does My Money Go?* encourages transparency and accountability across the public sector by improving the possible level of public scrutiny. Visitors to the

---

proceedings and court records, with the difference that the Swedish law also gave an explicit right of access to this material.

[143] See http://www.parliament.uk/site-information/copyright/ where it states that "*The material listed may be reproduced without formal permission for the purposes of non-commercial research, private study and for criticism, review and news reporting provided that the material is appropriately attributed.*"

[144] More information on Mapumental can be found on the MySociety website at: http://www.mysociety.org/projects/mapumental/

[145] See http://www.openstreetmap.org/

[146] See Traveline website at http://www.traveline.org.uk/index.htm. The information is under UK government (Crown) copyright.

[147] See http://www.landregistry.gov.uk/, the information is also under Crown Copyright.

[148] See http://scenic.mysociety.org/ run by MySociety.

[149] See http://www.wheredoesmymoneygo.org/dashboard/#/uk-bubble-chart/focus=TOTAL&year=2010-2011

site can obtain information specifically tailored to their own locality, in order to understand where their money goes to or comes from, as well as accessing details or government spending on a national level. Recent developments have included visualisations of the impact of the new budget on social services and infrastructures. Other applications help users navigate the financial data in the UK Treasury's Combined Online Information System (COINS) database[150], which was released in June 2010 following advocacy from such as (see more on this campaign in Section 2.2.1). The *Where Does My Money Go?* Project is also working to represent spending data at a European level and in a number of other countries around the world. The software is also being used to track UK bailout funds as well as the spending of international development aid worldwide.

- **Afghanistan Election Data[151]** was designed to increase transparency of the August 2009 Afghan election results as part of the election-monitoring mission of the National Democratic Institute.[152] It was developed in partnership with the online communications consultancy Development Seed,[153] the application uses preliminary raw vote count data released by Afghanistan's Independent Election Commission in September 2009. By combining this raw data with demographic, ethnographic, topographical, and security information, the site generates a range of interactive visualisations with multi-layered maps from OpenStreetMap. Users can clearly identify areas which had significant electoral irregularities, for example by searching for polling stations which returned over 95 percent for one candidate, or stations which returned far more votes than the normal number of ballots. The goal of the site is to help those involved in the Afghan democracy-construction process, both within the country and in the international community. For example, future elections can be improved by considering ways to deter fraud or identifying areas in which factors such as security risks or topographical barriers seem to have prevented people from casting their vote. All the data used in creating the visualisations can be extracted from the website, processed and redistributed.

- **Gapminder**[154] is an international-level project which aims to "unveil the beauty of statistics for a fact based world," by providing visualisations and developing tools for statistical graphics which are more interactive, dynamic, and easier to handle. In this case, the focus is on the representation of international development data with the objective of improving how social and economic trends are understood. They hold data for a vast range of indicators such as fertility rates, employment figures, aid donations, and carbon dioxide emissions, with data going back as far as 1800 for some sets. The data comes from a range of sources, such as the United Nations, the International Labour Organisation, the World Health Organisation, and the World Bank. It is represented using the Trendalyzer software that Gapminder developed and was subsequently purchased by Google, which creates animated graphics that make the statistics clear, intuitive, and even playful. Visitors to the site can use the raw data, transform the visualisations generated by Gapminder, find thematically organised packs for teaching and presentations, and watch videos which explain development issues in

[150] For further information see: http://www.hm-treasury.gov.uk/psr_coins_data.htm
[151] See Afghanistan Election Data at http://afghanistanelectiondata.org
[152] See National Democratic Institute at http://www.ndi.org/
[153] See Development Seed strategic consultancy and an open source product development company specialising in implementing technology projects for large international development organizations and leading open source research and development initiatives. See developmentseed.org
[154] See GapMinder at http://www.gapminder.org/ See also *Mapping for results* http://blog.aiddata.org/2010/08/mapping-for-results.html

ways which are accessible and entertaining. The project's vision is that all people should have access to global data sets in an array of formats and styles that will help them to understand the complexities of our global society without necessarily being advanced statisticians.

It is important to note that some of these initiatives have versions in other countries. For example, in New Zealand there is a fixmystreet.org.nz and theyworkforyou.co.nz. In Spain a version of WhatDoTheyKnow will soon be launched as MiPregunta.es.

5.4 Campaigns that target areas of non-publication

When civil society organisations need to get access to information which governments are reluctant to release, it is sometimes necessary to mount a specific campaign to target the areas of non-publication. Examples of campaigns which have successfully done this include:

- **The Free Our Data Campaign**[155] started on 9 March 2006 the UK newspaper the Guardian ran an article entitled "Give us back our crown jewels".[156] The article argued that many public bodies collect data as part of activities which are funded by tax-payers' money and that this data should therefore be available to the public free of charge, without having to pay for it for a second time. As is noted in Section 2.3, the costs for accessing such data can run into millions. Particular "culprits" identified by the article included the Ordnance Survey's mapping data and data collected by the UK Hydrographic Office and the Highways Agency as well as all government-owned agencies which receive a significant proportion of their income from taxes – as much as 50 in the case of the Ordnance Survey.[157] The article also noted that many public authorities have to pay for access to such data, resulting in costs being incurred inside government as well as for members of the public wanting to use the data, even when they want it for non-commercial purposes. The combination of the economic model of these public bodies and onerous copyright restrictions preventing re-use of the data were restricting innovation, according to the campaigners.

  The successes of the Free Our Data campaign have included that on 1 April 2010 the Ordnance Survey started releasing mapping data free of charge.[158] The campaign also served to raise wider public awareness of the right of access to government data and helped advance the open government data movement in the UK, having an impact on the open data policies of the new UK government elected in May 2010.

- **Farmsubsidy.org**[159] is an initiative whose goal is to obtain detailed information about payments and recipients of farm subsidies in every European Union member state under the EU's Common Agricultural Policy (CAP). The activists and journalists who run Farmsubidy.org aim is to increase transparency of the spending of the around €55 billion of agricultural subsidies spent annually in the EU (which represents about 40% of the EU's annual budget, or about €100 for each EU taxpayer). Their goal is also to make this information available in a way that is useful and meaningful to European citizens and, without taking any position on the CAP, to improve the quality of the debate around it.

[155] Free Our Data website at http://www.freeourdata.org.uk/
[156] See the article at http://www.guardian.co.uk/technology/2006/mar/09/education.epublic
[157] See http://www.theyworkforyou.com/wrans/?id=2007-07-17b.150285.h
[158] See BBC news item at http://news.bbc.co.uk/2/hi/8597779.stm and see also the Guardian news item at http://www.guardian.co.uk/technology/2010/apr/01/ordnance-survey-maps-download-free
[159] See Farmsubsidy.org at http://www.farmsubsidy.org/about/

The campaigners first used national access to information laws to obtain the data, securing access to data from Denmark in 2004, followed by the UK in 2005. The stories (and in some cases scandals[160]) that resulted from publication of this information generated a lot of media coverage and put pressure on the less transparent countries of the EU to open up their files. Eventually an EU directive was adopted requiring member states to release their data by 30 April of each year, starting in 2009.[161] The released data has revealed who are the largest recipients of funds (large sugar companies) and how some farmers are receiving small payments of as little as one euro, an amount which is probably vastly exceeded by the bureaucratic of processing the subsidy.

From an open government data perspective, one of the biggest challenges that the Farmsubsidy campaign has faced is that the data is often posted on government websites in ways that make it impossible to download in bulk. In May 2009 an Open Data Summit was organised in Brussels to scrape and process the data that had been released in multiple formats.[162] Since then a transparency ranking[163] has been created to lobby some states to change their policies for releasing information. For example, Hungary released its 2008 data in non-machine-readable PDF format. In April 2010, Hungary released the 2009 bulk data in a user-friendly and downloadable electronic format. Nevertheless, while some countries now follow the best practice of providing the 'raw data',[164] others still refuse to provide the data for bulk download, meaning that activists have to rely on "web-scraping" robots to harvest the data from government websites.[165]

The Farmsubidy.org initiative has had a global impact, spurring similar projects in countries such as Mexico.[166]

5.5 Projects which Promote Reuse

One of the more effective ways for CSOs to demonstrate to governments the value of opening up their datasets is to show the multiple ways in which the information can be managed to achieve the social and economic advantages set out in Section 5.1. CSO that promote reuse have been instrumental in countries where there have been advances in policy and law to ensure that datasets are both technically and legally open (see Sections 2 and 3).

---

[160] In 2010 the release of the 2009 data revealed that in Sweden (which includes personal identification numbers in the data), the youngest recipient of CAP funds is just 14 years old, while two Swedish recipients are 100 years old, though both are dead. The oldest living Swedish CAP recipient is a 98-year-old woman. Other curious recipients of CAP funds included: an accordion club (Sweden – €59,585.10), a billiard club (Denmark – €31,515, a payment for beer and soft drinks), a Juri High School alumni society (Estonia – €44,884), Ons Genoegen ice skating club (Netherlands – €162,444), the Sint Maarten amateur football club (Netherlands – €354,566.62) and Schipol Airport in the Netherlands (€98,864.33). In Bulgaria, the national CAP paying agency appears to have paid itself a sum in excess of a million euros as well as payments to the 26-year-old daughter of the former Bulgarian deputy agriculture minister (who had responsibility for EU funds) in excess of €700,000 euros.

[161] Ibid. A good summary can also be found in The Guardian, *Press forces EU Subsidies issue*, 7 February 2007, http://www.guardian.co.uk/media/2007/feb/07/pressandpublishing.newmedia1,

[162] See http://www.followthemoney.eu/analysis-presented-at-todays-press-conference-in-brussels/

[163] See https://docs.google.com/View?id=dknjc26_4gssshfkz

[164] These include: Belgium, Denmark, Estonia, Czech Republic, Lithuania, Poland, Romania, Slovenia and Sweden. See article at http://capreform.eu/2009-data-harvest/

[165] Ibid. The 2010 "transparency villains" included Austria, Cyprus, Greece, Ireland, Italy, Luxembourg, Malta, Netherlands, Portugal, Portugal and Slovakia.

[166] The "*Subsidios al Campo*" project in Mexico also uses access to information requests to obtain the data needed to track spending on agricultural subsidies. See http://www.subsidiosalcampo.org.mx/.

The typical activities which are undertaken as part of these initiatives normally include competitions, open government data conferences, "unconferences", workshops and "hack days". These activities are often organised by the user community with data that has already been published proactively or obtained using access to information requests. In other cases civil society advocates have worked with progressive public officials to secure new release of datasets that can be used by programmers to create innovative applications.

Box C

Geek Gatherings Explained

Unconferences: An "unconference" is so-called because its format is different from that of formal conferences which have registration fees and structured presentations. The unconference is more open and participant-driven than a traditional conference, but still has the feature of an event where people come together in the same physical space to discuss particular themes

BarCamps are open workshops on technology issues whose content is provided by participants. The events are initiated and organised using open wikis to define the content and agenda.

Hack Days are events where programmers, developers, designers and others come together and work, over a short space of time, typically 48 hours, to develop new applications.

MashUp: a "mashup" is web application that combines or remixes data and/or functionality from more than one source It can also be used as a verb "to mash-up" and is applied as a noun to events where programmers take data and mash it up.

5.5.1 Competitions

As an incentive to encourage computer programmers and others to design applications which make use of government data, one strategy now being employed is for public bodies to organise open competitions. The structure of these competitions is that a number of datasets are released and programmers participate within a short time-frame, running from as little as 48 hours to a few weeks, to develop applications. A prize is then awarded to the best application. Competitions have been held in a number of countries including the UK, the US, Norway, Australia, Spain, Denmark and Finland.[167] Annex 2 gives a full listing of such competitions.

- **Show us a better way**[168] was the first such competition in the world. It was initiated by the UK Government's "The Power of Information Taskforce" headed by Cabinet Office Minister Tom Watson in March 2008. This competition asked "What would you create

[167] In addition to the examples given in the text, other prominent competitions include: "NYC Big Apps" in 2009 in New York City; the "Competition Public Data In Play" in 2009 in Denmark; and the "Apps for Finland Competition" in 2009 in Finland. A more comprehensive listing can be found in Annex XX.

[168] See http://www.showusabetterway.co.uk/call/2008/11/and-the-winners-are.html information about the data used for the competition can be found at: http://www.showusabetterway.co.uk/call/data.html

with public information?" and was open to programmers from around the world, with a tempting £80,000 prize for the five best applications. The winners were:

- *Can I Recycle It?* Postcode can be used to find out about local government recycling schemes;
- *UK Cycling*: A website for planning a cycling route;
- *Catchment Areas*: Information about the school districts, including where children living in a zone have a choice of more than one school;
- *Location of Post Boxes*: Directs people to the nearest post box;
- *Loofinder*: A mobile texting or website service to find the nearest public toilet.

The organizers noted that the programmers were *"telling us [government] what information they want and how they want to use it."* A number of additional programmes were selected for being explored further by the government:

- *Road Works API*: An interface linking to road works so that other organisations (such as satnav companies) or individuals could build alert systems;
- *Oldienet*, on local services for the elderly population;
- *Free Legal Web*, an authoritative "mashup" of expert legal commentary and public-sector information;
- *Allotment Manager*, for better allocation of garden allotments;
- *Where Does My Money Go*, an interactive web application showing government budget data via maps, timelines, graphs and charts.

➢ **Apps for Democracy**, one of the first competitions in the United States, was launched in October 2008 by Vivek Kundra, at the time Chief Technology Officer (CTO) of the District of Columbia (DC) Government. Kundra had developed the groundbreaking DC data catalogue, http://data.octo.dc.gov/, which included datasets such as real-time crime feeds, school test scores, and poverty indicators. It was at the time the most comprehensive local data catalogue in the world. The challenge was to make it useful for citizens, visitors, businesses and government agencies of Washington, DC.

The creative solution was to create the Apps for Democracy contest. The strategy was to ask people to build applications using the data from the freshly launched data catalogue. It included an online submission for applications, small but many prices instead of big but few prizes, and several different categories as well as a "People's Choice" prize. The competition was open for 30 days and cost the DC government $50,000. In return, a total of 47 iPhone, Facebook and web applications were developed with an estimated value in excess of $2,600,000 for the local economy.[169] After the competition, Kundra commented that "Apps for Democracy produced more savings for the D.C. government than any other initiative." The winning applications included:

- iLive.at: by entering an address in DC, users have access to information tailored to that exact location organized into categories.
- DC Historic Tours, a Google Maps mashup that combines custom walking tour creation with Flickr photo feeds and Wikipedia entries.

The people's choice awards were:

- Car Pool Mashup Matchmaker
- DC Bikes - Your Guide to Biking in DC

➢ **The Abre Datos (Open Data) Challenge 2010.**[170] Held in Spain in April 2010, this contest invited developers to create open source applications making use of public data in just

[169] For a list of all applications: http://www.appsfordemocracy.org/application-directory/
[170] Abre Datos, more information can be found at http://www.abredatos.es/

48 hours. The competition had 29 teams of participants which developed applications that included a mobile phone programme for accessing traffic information in the Basque Country, and for accessing data on buses and bus stops in Madrid, which won the first and second prizes of €3,000 and €2,000 respectively.

- **Nettskap 2.0**.[171] In April 2010 the Norwegian Ministry for Government Administration held "Nettskap 2.0". Norwegian developers - companies, public agencies or individuals - were challenged to come up with web-based project ideas in the areas of service development, efficient work processes, and increased democratic participation. The use of government data was explicitly encouraged. Though the application deadline was just a month later, on May 9, the Minister Rigmor Aasrud said the response was "overwhelming". In total 137 applications were received, no less than 90 of which build on the reuse of government data. A total amount of NOK 2.5 million was distributed among the winners; the estimated value of the applications developed was NOK 28.4 million.

- **Mashup Australia**.[172] The Australian Government 2.0 Taskforce invited citizens to show why open access to Australian government information would be positive for the country's economy and social development. The contest ran from October 7th to November 13th 2009. The Taskforce released some datasets under an open licence and in a range of reusable formats. The 82 applications that entered into the contest are further evidence of the new and innovative applications which can result from releasing government data on open terms.

One strategic consideration for those in government or civil society planning to organise projects to promote reuse of government data is to make use of datasets which already have been or can easily be released. Going after these "low hanging fruit" often makes it easier for public officials to persuade their colleagues that the data can be released and used by members of the public. For this reasons, organising such events around datasets that contain sensitive information (such as personal data) which needs first to be removed (something which can be a timely process if it was not anticipated in the design of a database) is not recommended.

5.5.2 Open government data conferences, "unconferences", workshops and hack days

Events centred around government data permit both user communities (human rights activists, journalists, etc) and public officials gain a perspective on the diversity of potential ways in which the government-held information can be used. In particular, these events offer an occasion to forge closer ties with public officials and enable governments to identify and prepare the data sets that different social groups may be interested in. Conferences, unconferences, workshops and hack days provide an analogical and digital platform for brainstorming and initiating new projects to use government information. Some leading examples are given below:

- **Gov 2.0 Expo**[173] Technology lovers and data geeks gathered at this conference in Washington D.C. from the 25th to the 27th of May 2010 in order to explore the links between data, technology, and public information. The goal of Gov 2.0 was to

[171] See website of the Ministry of Government Administration, Reform and Church Affairs (in Norwegian) at http://www.regjeringen.no/nb/dep/fad/kampanjer/nettskap.html?id=599571
[172] See the winners of Mashup Australia at http://mashupaustralia.org/winners/
[173] See Gov 2.0 website at http://www.gov2expo.com/gov2expo2010/public/content/about

strengthen contacts between public bodies and private citizens and to discuss how to harness the combined potential of transparency, participation and collaboration in government to spark innovations in the administration and the technology sector in order to serve the public good.

- **Mashup Camp**,[174] one of the first events to be termed an "unconference" took place in February 2006 in Silicon Valley, California, and it was an open gathering of technology lovers in the area, that was free to attend with the only limitation being that of space – 300 people. It was organised on a first-come-first-serve basis, and was an inexpensive and informal gathering that allowed for greater participation of all attendees in the belief that the audience should not be a passive receptor of information from experts; rather they should be encouraged to share their own personal expertise with the objective of jointly arriving at innovative solutions.

- **BarCamp**[175] is a series of informal conferences where people working on open source and technology projects can meet, network, learn, and teach. The first took place in Amsterdam in 2005, and since then there have been many more all over the world, including in Dallas, Delhi, Cape Town and Nairobi. These are organised through wikis where people share information and debate about the issues to be discussed prior to the unconference. The spontaneity and informality of these events allows for more creative juices to get flowing, and for a community of data geeks to come up with some truly innovative for connecting citizens and software.

- **Hacks4Democracy**[176] was a two day barcamp-style meetup held in April 2010 in Berlin. The goal of the event was to demonstrate that it is possible, in a short time and on a low budget, to programme creative and innovative prototypes that make data from the public authority accessible and usable. About 60 people participated and created several applications and prototypes with the community-driven catalogue for open data in Germany[177], which uses the same underlying technology as the UK open data government-run website described above.

- **Rewired State**[178] is an example of how a community-driven approach to work with open government data can help build a better society by enabling a productive dialogue between "Geeks and Government". Rewired State is a community of developers which runs hack days to show government officials what is possible and for government to show developers what is needed. Rewired State has now been hired by the British government to run hackdays for government agencies to help them improve their services.

[174] See report on the event at http://money.cnn.com/2006/06/05/technology/business2_unconference0606/index.htm
[175] See the BarCamp website at http://barcamp.org/
[176] See the Hacks4Democracy website at http://opendata.hackday.net/
[177] See www.offenedaten.de
[178] See the full list of projects developed on the Rewired State website at http://rewiredstate.org/projects