THE Open GROUP

Business Scenario: Open Public Sector Data

Business Scenario: Open Public Sector Data

Document No.: K140

Published by The Open Group, December 2014.

Any comments relating to the material contained in this document may be submitted to:

The Open Group, 44 Montgomery St. #960, San Francisco, CA 94104, USA

or by email to:

ogspecs@opengroup.org

Contents

List of Figures

List of Tables

Management Summary

Governments worldwide are making much of their data publicly available. They have two main reasons for this: to improve the functioning of government and society as a whole, and to stimulate economic development. These goals have so far been achieved only partially.

This Business Scenario describes the problems that are limiting the achievement of these goals, the business and technology environments and processes for publication of public sector data, and the people and computing components (called "actors") who execute the scenario. These considerations lead to requirements for standards and best practices for publication of public sector information.

The requirements cover:

- Human readability and machine processing
- Ease of publication
- Interaction with subjects and subject owners
- Rights and concerns of stakeholders
- Optimization for search engines
- Publication with common metadata
- Catalogs and indexes

The Business Scenario is an input to the SHARE-PSI project, of which The Open Group is a member [SHARE-PSI]. The results of this project will in turn form input to the W3C Data on the Web Best Practices Working Group [W3C Web Data], which will develop standards and best practices in the light of the discussions at and conclusions reached by the SHARE-PSI workshops.

Common standards and best practices for publication of public sector information will contribute significantly to the goals of improving the functioning of government and society as a whole and stimulating economic development. It will do this by making it easier for political interest groups, social enterprises, commercial companies, and citizens generally to understand their social and political environments, and to interact with governments, and by making it easier for commercial companies to define, develop, and supply products and services that use public sector information.

Business Scenario

Open Public Sector Data

Business Scenario Problem Description

Governments worldwide are making much of their data publicly available. They have two main reasons for this: to improve the functioning of government and society as a whole, and to stimulate economic development.

There are directives on this topic from many governments, including the European Parliament and Council [EU PSI Directive] and the US Government [US 2009]. There are also economic analyses from such august bodies as OECD [OECD] and the World Bank [World Bank].

The goals have so far been achieved only partially. The problem is to find ways of improving the process of publishing public sector data so that it can be achieved more completely.

The Currency of Democracy

The idea that "information is the currency of democracy" is often attributed to Thomas Jefferson, although there is no evidence that he actually said or wrote those words. Whoever originated them, they capture perfectly the importance of information as an underpinning of democratic society.

People use economic and social statistics to measure the performance of national governments. At a local level, they want to know about proposals that affect them; for example, for new roads or for building developments. At all levels, they are interested to see how the actual performance of politicians compares with the promises.

Publication of information about how public money is used helps ensure that it is used properly. For example, *Supervizor* is an online application that provides information on business transactions of public sector bodies in Slovenia. In 2012, it brought to light 68 cases of violation of a restriction of doing business between public institutions and private entities. This led to effective enforcement, and there were no new cases in 2013 [Supervizor].

Transparency of government operation leads to improved services and service delivery. It can also cut costs; for example, through reduced personal interaction with citizens, including reduced freedom of information requests, and through data sharing within government.

When people lack information about matters that affect them, they feel disenfranchised and excluded. When they have that information, they feel involved, and are more likely to express their views.

The guiding principle of democracy is that the citizens take the decisions. To do this, they must understand the important factors and underlying considerations. Public availability of good information enables a well-functioning, transparent, and democratic society, with increased citizen participation in government, and more efficient administration.

The New Oil

The second major goal of making public data openly available is to stimulate economic activity.

This goal was well expressed by Neelie Kroes [Kroes] when, as the European Commissioner for the Digital Agenda, she said that:

"Data really is the new oil. Data is a raw material for information businesses, just as oil is a raw material for fuel and plastics businesses. Data is also everywhere, it is cheap, and it can deliver huge rewards both in terms of services and financial returns."

There are some notable examples of economic activity enabled by data. The Global Navigation Satellite System market is estimated to be worth €100 billion per year, growing at 13% [GNSS]. This is in part based on availability of geographic map data, provided in many cases by government or government-sponsored bodies. A second example is that online purchasing is routinely simplified by the use of postcode data. The economic benefit of this is hard to quantify, but is clearly very large.

Open public sector information should lead to new business opportunities and jobs in the private sector, through stimulation of existing businesses (e.g., increased tourism), and through companies making new products and services (e.g., information-based "apps").

Information-based products and services are increasingly important in today's digital societies. When public sector information is openly available, private sector companies can provide products and services that use it and add value, contributing to the growth of national economies.

Problems

There are, however, a number of problems that limit the realization of these goals.

Cost of Acquisition and Processing

It costs money to acquire data, just as it costs money to extract oil from the earth.

Much public sector data is obtained through people filling in forms, in the course of administrative transactions, or in response to censuses and surveys. The forms have to be designed, printed, and distributed. The input has to be checked. Often, manual input is transcribed to machine-readable format. Action may be needed to ensure that everyone who should do actually supplies the information.

Online input is increasingly possible. This is often cheaper, but there is still the cost of development of data collection applications, and of data processing and storage.

In some cases, specialized people or equipment are needed to collect the information. Geographic data for maps, for example, requires both.

Quality of Data

As with data of any kind, ensuring the quality of public sector data can be difficult. Problems described at the Samos workshop [Samos] included duplication, lack of updates, lack of completeness, and incorrect manual input.

Data is collected for administrative purposes; collection is not an end in itself. Complete records are only available in cases where some legal entity is the subject of inspections.

Measures can be taken to improve quality, but they are often expensive. There is a trade-off to be made between quality and cost.

Quantity of Data

While the quantity of published open public sector data is not large currently, there is a data explosion in prospect. For instance, considering the public administration of Trentino, Italy, it is estimated that the trend for growth will be up to hundred times per year, easily reaching terrabytes of data by 2018 [Trentino].

This expected growth, not only in the amount but also in the variety and rate of change of open public sector data, presents challenges in using traditional data processing systems and applications.

Ability to Find Data

Data is published by the responsible administrative departments, usually acting independently. Descriptions of what is published are not always produced. Where they are produced, they are not co-ordinated, so that similar data published by different administrations may be described in quite different terms. There generally are no catalogs or indexes showing the totality of what is available.

This makes it hard to establish whether the data required for a particular purpose is available, and to obtain it where it is available.

Lack of Confidence in Data Providers

Some potential data re-users fear that the data will not be truly open, and that the providers will apply discrimination in various subtle ways; for example, in the speed with which they notify different users of updates, or by endorsing some apps but not others. There are particular concerns where the data provider is also involved in downstream service provision.

Also, it is difficult to bind governments to behaving in certain ways. A government may set up commercial expectations by following a particular policy, but cannot guarantee not to change the policy, with unfortunate consequences for companies that rely on the expectations. The desire to serve the best interests of the voters often overrides the desire to treat business partners consistently and fairly.

So, for example, a policy of providing a particular kind of data free-of-charge, or at a low cost, could be changed, and this could destroy the commercial viability of products and services that use the data. This is a risk that a commercial re-user must assess before investing in product or service development.

Ability of Computers to Process Data

Much public sector data is published in formats that are designed to be interpreted by people (for example, as text, graphics, or video) rather than formats that are designed for machine processing (for example, as database tables). Even where it is provided in machine-processable form, there is a lack of metadata that would enable computer applications to interpret the data.

Different administrations may make machine-processable data available in different ways. For example, one might supply it in CSV files, while another makes the same data available in JSON through a web service API.

A particular issue is that location information is represented differently in different datasets.

As noted above, similar data published by different administrations may be described in quite different terms, and this also applies to any supplied metadata. This makes it hard for computer applications to integrate and use machine-processable data published by different bodies.

The variety of publication formats and metadata makes it hard to develop large-scale applications that can be used across administrative areas. For example, it is easy to develop an application to plan bus journeys for a particular city, but hard to develop one that can be used in any city in Europe. It also makes it hard to integrate data produced by different bodies (as, for example, in an application that planned bus journeys between any two points in Europe).

Existing Administration Practices and Systems

Open data implies a radical change to the approach to data and to working with it inside public administrations. Public sector bodies today often do not have a culture of making data available, publish data independently of each other, and have "stovepipe" systems that do not easily share data.

When publication reveals inaccuracies in the data, or inconsistencies in administrative procedures, this may be seen as reflecting badly on the people concerned, rather than as an opportunity for improvement.

When administrations publish independently, they will be likely to use different formats and to describe their data in different ways, with the consequences noted above.

Use of "stovepipe" systems makes it costly to share or integrate data, despite the complexity of modern public services which rely on interoperability. This is a general problem, not restricted to the public sector. The problem caused by the lack of the right information to the right person at the right time, preventing organizations from achieving their business objectives, is described in The Open Group Business Scenario: Interoperable Enterprise [Interoperable Enterprise].

Citizens' Attitudes

Obtaining the desired benefits from open public sector data also implies some changes to the attitudes of citizens towards their governments and administrations. They are often unaware of public decision-making processes, of the possibilities for their involvement in these processes, and of the desirability of being involved.

Defining Commercial Products and Services

While data is a raw material for information businesses, just as oil is a raw material for fuel and plastics businesses, it is not easy for those businesses to define commercial information-based products and services.

This is reflected in a point made at the Samos SHARE-PSI meeting [Samos] that, even though valuable data is released, additional effort is required to energize external stakeholders to create something with it. For commercial companies, there is little point in making attempts to energize by explanations or requests; they will be energized only if they are paid directly to do something, or see that a profit is to be made.

While the released public sector data may have value, it is the value added by the information business that is the basis for a viable commercial product or service. Just re-publishing the data cannot be the basis for a product that costs money, because the customers can obtain the data for

free from its original source. Also, because the data is equally available to other companies that can create competitive products, a company is unlikely to invest in developing a product that uses open data unless there are factors that make it hard for other companies to develop similar products.

In fact, the perceived value of the original data is often low. Few people, for example, are interested in reading the minutes of the meetings of their local authority. They might, however, be very interested in a newspaper story about a local politician breaking election promises. The story is based on the politician's actions recorded in the minutes, but it is because of the value added by the journalist in writing the story that people read the article.

In the case of satellite navigation systems, the products are based not only on the geographical information, but also on GPS reception hardware and routing algorithms. Also, the geographical information is often obtained from a variety of sources, including commercial ones as well as the public sector. The product is not based on a simple addition of value to public sector information, but on a combination of inputs and processes. Its value is defined in simple terms related to the user ("find your way from place to place with ease") rather than in terms of the original information.

These characteristics are shared by many of the successful information-based products and services. Defining such products and services is not an easy matter. Looking at what public sector information is available and trying to work out what products and services it could be used for may not be the best approach. It is more likely that the successful companies start by looking at market needs, and happen to notice that public sector information could help to satisfy them.

Implications

Because of these problems, the aims of making public sector data openly available have not been fully achieved. It is difficult to quantify how far functioning of government and society as a whole could be improved by more effective publication, but there is some indication of how far the expected economic benefits have failed to materialize.

The economic benefits of open government data could be huge. The McKinsey Global Institute estimates a potential of between three and five trillion dollars annually [McKinsey]. Yet the direct impact of open data on the EU economy in 2010, seven years after the directive was issued, is estimated at only about 1% of that [Capgemini], although the EU accounts for nearly a quarter of world GDP. This suggests that only a small fraction of the economic potential has been realized.

There are a number of reasons why the potential has not yet been realized, including:

- Incomplete availability of data, due to costs of acquisition and publication, or provision of data in unusable formats, and inability to find data even when it is available in a usable format
- Lack of confidence in the data, due to actual or perceived lack of quality, or lack of confidence in government as a provider
- Fragmented markets due to lack of data, metadata, and interface standards
- Difficulty of identifying unique value that a commercial company can add

These reasons explain the lack of commercial products and services currently using public sector data, in comparison with the large expected potential.

Detailed Objectives

The aim of this Business Scenario is to help the standards community to identify and develop standards and best practices for the publication of public sector data that will improve the functioning of government and society and stimulate economic development.

The Business Scenario is developed as part of The Open Group Open Platform 3.0™ initiative [OP3.0]. This initiative focuses on new and emerging technology trends converging with each other and leading to new business models and system designs. These trends currently include: mobility; social networks and social enterprise; big data analytics; cloud computing, and the Internet of Things (networked sensors and controls). The initiative is identifying a set of new platform capabilities, and architecting and standardizing an IT platform by which enterprises can gain business benefit from use of the technologies. Support for creation, management, and publication of open public sector data is an important requirement for this platform.

The Business Scenario is developed for the EU SHARE-PSI project, in which The Open Group participates. This project is holding workshops over 2014-15. The output of the workshops will be offered as input to the W3C Data on the Web Best Practices Working Group [W3C Web Data], which is compiling a W3C standard that will help guide people and organizations around the world as they build the web of data. Towards the end of the SHARE-PSI 2.0 network's life (the first half of 2016), the partners will incorporate the W3C Best Practice in their own guidelines as relevant to them in their country or sector of interest.

The solution envisaged by this Business Scenario is the standard and associated guidelines that the W3C Working Group will develop. Its success can be measured in the first instance by their take-up by SHARE-PSI partners, and ultimately by the amount of public sector data published worldwide in accordance with them.

The increased availability of public sector data can be expected to result in increased economic activity, more efficient government, and a more democratic society. These effects will, however, be hard to measure, and must be taken on trust.

Views of Environments and Processes

Business Environment and Process

The overall business environment and process for re-use of public sector information is shown in the figure (Figure 1).

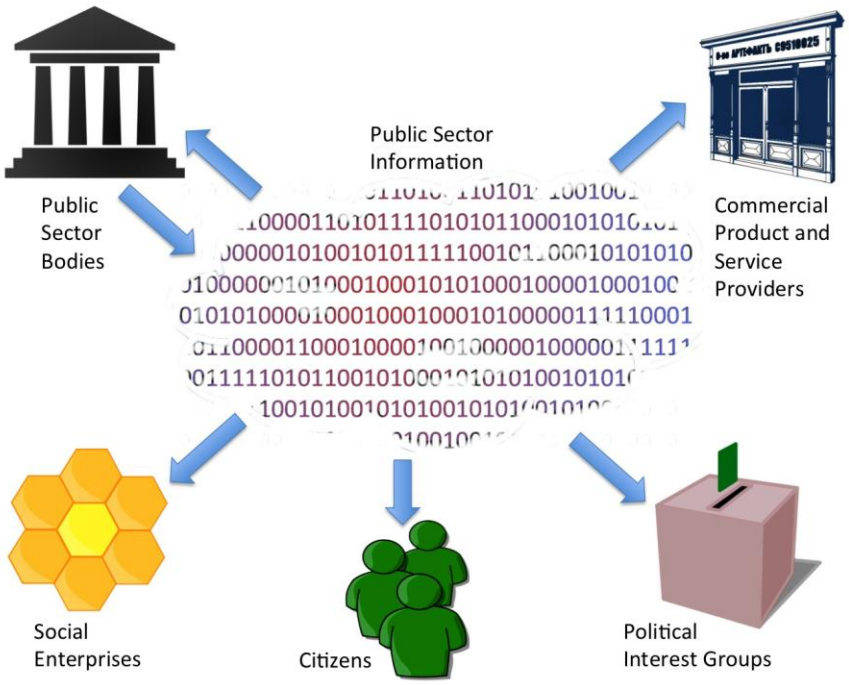

Figure 1: Business Environment and Process

Public sector bodies collect the information in the course of their activities, primarily for their own use, and make it available for use by consumers, including other public sector bodies, commercial product and service providers, social enterprises, political interest groups, and citizens. The commercial product and service providers may use the information in their products and services, which they make available to other consumers.

Public Sector Bodies

Public sector bodies perform a range of services for their communities and the people in them. These services involve the collection of information of various kinds.

The information topics in question can include, for example: agriculture; businesses and companies; crime and justice; demographics; education; energy, resources, and utilities; environment; geospatial and mapping; government operations, spending, and business; health and social care; housing; personal finance; social and community; and transport [Deloitte].

For the purposes of this Business Scenario, public sector bodies are as defined in EU Directive 2003/98/EC on the re-use of public sector information [EU PSI Directive]. They are 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 boundary between public and private sector activity is not well defined in functional terms. It can vary from time to time and from place to place. For example, 50 years ago, most post and telecommunications services worldwide were performed by public sector bodies, but they are now mostly performed by companies in the private sector. Today, healthcare is largely carried out by the public sector in some countries, and by the private sector in others. Also, it can sometimes be hard to determine whether a particular body is in the public or the private sector. For example, the UK Ordnance Survey is a government agency responsible for producing maps, but it operates in many respects as a commercial organization.

The fact that information is collected in the course of public sector activities has implications for how access to it is governed. The information can be regarded as public property. Access to it can be charged for, but should be offered on fair and equal terms. A private sector company can arbitrarily make information that it collects available to some parties but not others; a public sector body should not do this.

Commercial Product and Service Providers

These are enterprises that provide products or services to make a profit.

Provision of a product or service typically requires an initial investment in its development. The enterprise concerned assesses the investment required and the expected return, and makes a risk-based decision. The expected return depends on a number of factors, including the value of the product or service to consumers, and the possible competition from similar products or services.

There are many kinds of product and service that can use public sector information. They include journalism, insurance, credit-rating, market analysis, sales lead generation, freight logistics, and journey planning.

A study of companies developing such products and services in Spain in 2012 [Infomediary] showed that the most frequently used kinds of public sector information were geographic/cartographic information (used by 51% of companies), business/financial information (47%), social demographic/statistical information (30%), and legal information (28%). Note that over 70% of the companies combined two or three types of information in their products. The companies were of a mixture of sizes. Only about a quarter of them classed re-use of public sector information as their main activity. Use of data no doubt varies between countries, but this is probably a reasonably representative pattern. For comparison, a UK study found that the most popular, and potentially most valuable, datasets include geo-spatial, environmental, transport, health, and economic data [Deloitte].

Many information-based products take the form of computer applications, and many information-based services take the form of web services.

The income that a company obtains from its products or services traditionally comes from straightforward sales, but there are a number of other commercial models that can apply to computer applications and web services. (See, for example, Free [Anderson].) The trend for these models to replace the traditional sales model is particularly marked for applications that run on mobile devices (*apps*). An app typically makes little money through direct sales [Louis]. The *freemium* model, involving free download with in-app purchases, is the most successful monetization strategy for app developers today, followed by in-app advertising [App Annie].

Social Enterprises

A social enterprise is an organization that applies commercial strategies to maximize improvements in human and environmental well-being, rather than maximizing profits for external shareholders. Its activities thus fall somewhere between public services and commercial product and service providers.

There are many such enterprises today. It is estimated, for example, that there are 70,000 of them in the UK, employing nearly a million people, and with a combined turnover of over £50 billion [UK Social Enterprises].

Such enterprises are important users of public sector information, and can be considered as contributing to the economic benefits of making that information openly available.

Political Interest Groups

Political interest groups also fall between public services and commercial product and service providers. A political interest group is an organization that applies political strategies to bring about (or in some cases to prevent) social or environmental change. Such organizations include political parties, pressure groups, and parliamentary lobbyists. They also include groups formed to address local issues; for example, to object to particular road or housing developments.

Such a group has a particular interest in public sector information that relates to its specialist topic. A group that objects to a road development, for example, may wish to research traffic levels in the area, noise levels associated with similar roads elsewhere, and the interests and voting records of the politicians who will make the decision.

Such a group may also re-publish the information (generally with added opinion), in the form of newsletters, leaflets, etc.

Citizens

Citizens are in theory responsible for taking the decisions in a democracy. In practice, many of them appear to take little interest in direct political activity. For example, the average turnout in the 2014 EU parliamentary elections was less than 45% [EU 2014]. The typical citizen does not engage in significant study of public sector data.

Most citizens do, however, follow public affairs through newspapers, broadcast media, and, increasingly, online. They rely on these channels to filter and organize their input information, and to bring to their attention anything that they might feel to be of major importance.

When they do engage in significant political activity, it is likely to be as members of political interest groups.

Technical Environment and Process

Public sector information can be published, and historically usually was published, using printed paper, but digital technology is expected to be mainly used in the foreseeable future. This Business Scenario only considers use of digital technology.

The overall technical environment and process is shown in the next figure (Figure 2).

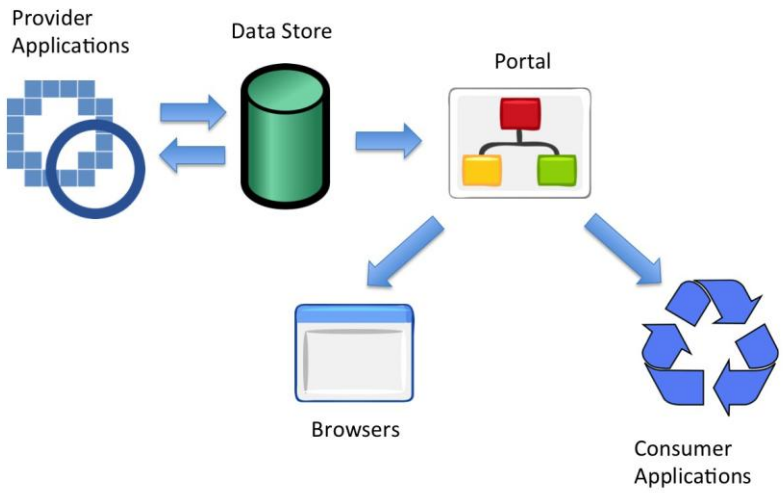

Figure 2: Technical Environment and Process

The public sector body responsible for the information holds it as data in a data store, using provider applications to create, manage, and process it, and makes it available for re-use through a portal. The consumers, including other public sector bodies, commercial product and service providers, social enterprises, political interest groups, and citizens, access it using browsers and consumer applications. Commercial product and service providers may use such applications to implement products and services, or may supply them as products to other consumers.

Provider Applications

The provider applications can be implemented using any of the wide variety of programming languages and environments available today. They can run on systems operated by the public sector body itself, or using cloud computing. They can gather information in a number of ways, including by direct user input, by transcription of written input (possibly using optical character recognition), and by using sensors in the Internet of Things.

Data Stores

Data that is primarily intended for machine processing is most commonly held in relational databases that can be queried using SQL. Such data is often referred to as *structured data*.

Another kind of data store that is sometimes used is the triple store, which can support semantic processing based on W3C's Resource Description Framework (RDF) [RDF] and Web Ontology Language (OWL) [OWL], and can be queried using the RDF Query Language [SPARQL].

Data that is primarily intended to be used by people can be held in the fields of relational databases or triple stores, but is most commonly held in flat files, using a variety of text, graphic, audio, and video formats. Such data is classed as *unstructured data*.

Portals

A portal is a computer system through which the public sector body makes data available to consumers. The web is the normal way of making such data available today, and its use is assumed in this Business Scenario.

A portal will generally act as a server of HTTP [HTTP]. This is the generic data access protocol of the web. Other data access protocols such as FTP [FTP] are sometimes used instead.

A portal may support a web service API that provides for specific data items to be retrieved in response to specific requests, in a defined way.

Documents containing unstructured data that are published on the web in HTML can be linked together and can have added metadata that facilitates its categorization and discovery. Search engines can use this metadata, but are also able to discover and index unstructured data in other formats, such as Portable Document Format (PDF), by analyzing its content.

The term *linked data* refers to a set of best practices for publishing and interlinking structured data on the web [HEATH]. These practices facilitate the processing of such data by computers, and provide for the addition of metadata so that the data can more easily be discovered and integrated.

A portal may use syndication protocols to advise consumers of updates. ATOM [ATOM] and RSS [RSS] are the principal standard web syndication protocols.

CKAN is an open source data portal platform used by many public-service data publishers [CKAN]. It supports publishing, sharing, finding, and using data.

Browsers

A web browser is a program that enables its users to read documents on the web. It is designed particularly to read documents in HTML, but can typically cope with most formats in which unstructured data is stored on the web by invoking other programs that are designed specifically for those formats. It can also cope with some structured data, particularly spreadsheets in CSV format, in this way.

Web browsers are universally available, and are used by consumers of all kinds, including citizens who have no other data reading or processing capabilities.

Consumer Applications

These are applications that consumers use to retrieve and process data that is available from portals.

They are most frequently used to process structured data. Through sophisticated analysis techniques, they are sometimes able to interpret and process unstructured data.

Like provider applications, they can be implemented using any of the wide variety of programming languages and environments available today. They can run on systems operated by the consumers, or using cloud computing.

They can incorporate internal data stores. For structured data, these typically use relational database technology.

They can use “big data” technology to process datasets that are large or that change frequently.

In some cases they, or components of them, run on mobile devices.

Actors and their Roles and Responsibilities

Human Actors and Roles

"Human" is here taken to mean "legal person". The human actors include organizations as well as individual people. They are listed with their roles and concerns in Table 1.

Table 1: Human Actors and their Roles

Human Actor	Role(s)
Subject	Person or organization that the data is about: • May be concerned about the quality of the data • Can input to improve the quality of the data • May be concerned about visibility of the data • Can request privacy or publicity, and may have a right to privacy
Subject Owner	Person or organization that owns the entity that the data is about, when this is not a person or organization. For example, the owner of a building that appears in a street plan. Has a similar role to a subject, with similar concerns.
Creator	Person or organization that creates the data: • Makes data available to the publisher • May be the subject (e.g., citizen completing census form) • May create data with inanimate subject (e.g., a map) • Responsible for quality of the data • Often has IPR in the data
Publisher	Public sector organization that selects, formats, and publishes the data: • May be the creator, and possibly also the subject (e.g., when a council publishes its meeting minutes) • Responsible for availability, discoverability, and quality • Responsible for protecting rights of creator, subject, and subject owner
Consumer	Person or organization that uses the data; e.g., citizen, political interest group, social enterprise, commercial company: • Concerned about availability, discoverability, and quality • May wish to process the data using computers, in which case wants machine-readable format • May wish to read it manually, in which case wants human-readable format

Human Actor	Role(s)
Commercial Product or Service Supplier	Person or organization that produces and sells product or service using the data; e.g., media company, insurance company: • Responsible for product or service quality • Responsible for protecting rights of creator, etc. • Concerned about availability, discoverability, and quality • May wish to process data using computers or manually

Computer Actors and Roles

The computer actors and their roles are listed in Table 2.

Table 2: Computer Actors and their Roles

Computer Actor	Role(s)
Data Store	Holds the data.
Provider Application	Used by creator or publisher to create, manage, or process the data.
Portal	Used by publisher to make the data available.
Browser	Used by consumer to read the data, particularly when it is unstructured.
Consumer Application	Used by consumer to process the data and display the results, particularly when it is structured.

Requirements

A common set of standards and best practices for publication of public sector information will contribute significantly to the goals of improving the functioning of government and society as a whole, and stimulating economic development. It will do this by making it easier for political interest groups, social enterprises, commercial companies, and citizens generally to understand their social and political environments, and to interact with governments, and by making it easier for commercial companies to define, develop, and supply products and services that use public sector information.

The following requirements are for the form that those standards and best practices should take in order to optimize the achievement of these goals.

Human Readability and Machine Processing

The standards and best practices should cover publication of information for human readers and for machine processing.

Human readership and machine processing are both important but require different publication formats, typically including unstructured and structured data formats.

Ease of Publication

The standards and best practices should facilitate the publication of information by public sector bodies in the forms in which it is available to them, rather than requiring significant transformation.

It is difficult for public sector bodies to estimate the commercial or other value of their information. The risk of deciding what publication form will best deliver that value, and the work of converting it to that form, should be left to commercial product and service providers, and other consumers.

Cost of acquisition and processing is one of the identified problems that limits achievement of the goals.

Interaction with Subjects and Subject Owners

The standards and best practices should not only cover publication, they should also address input of information, input of requests for the visibility of information (particularly for privacy), and input of feedback on the quality of published information.

Visibility of information is a concern of subjects and subject owners. Quality of information is a concern of all stakeholders, and is one of the identified problems that limits achievement of the goals. Improving quality and addressing concerns over visibility will also mitigate another of these problems: lack of confidence in the information. Finally, increased interaction will contribute directly to the quality of government and the health of democratic society.

Rights and Concerns of Stakeholders

The standards and best practices should cover the storage with the information of the rights of subjects, subject owners, creators, and other stakeholders, and of requested levels of visibility.

For systems to take account of stakeholders' rights and concerns, they must be aware of what the rights and concerns are.

Optimization for Search Engines

Publication formats should facilitate the operation of search engines on public sector data.

Ability to find data is one of the identified problems that limits achievement of the goals. Discovery by human readers can best be achieved through search engines.

Publication with Common Metadata

Information should be published with common metadata to facilitate discovery and integration.

Common metadata means that public sector bodies use the same terms, or terms that can be translated to each other directly, to describe the same information. It may not be possible to achieve a single standard terminology, because different bodies have historically used different terms, and because different countries use different languages, but it should at least be possible to map the used terminologies to each other so that translation is possible.

The Open Group UDEF standard [UDEF] provides a framework for the development of common metadata of this kind.

Search engines can use metadata, so its inclusion will assist optimization for them.

Most discovery of data for machine processing is currently performed by people, but it is desirable that computer applications should be able to discover the data that they process. Machine-processable metadata increases the possibility of this. Improved discoverability will help to address the identified problem of it being hard to define commercial products and services.

As well as enabling computer applications to discover data, machine-processable metadata improves their ability to process data and to integrate data from different sources. Inability of computers to process data is another of the identified problems that limits achievement of the goals.

Catalogs and Indexes

Publication on the web of catalogs and indexes of public sector information can make a significant contribution to discoverability, and should be considered as a recommended best practice.

The catalogs and indexes should include metadata for the documents that they reference. This will facilitate discovery of those documents even when they themselves do not include metadata.

Next Steps

The SHARE-PSI project will hold further workshops on the following topics:

- **Encouraging data usage by commercial developers**: Lisbon, hosted by the Portuguese Government, December 2014.
- **Identifying data sets for publication**: Timisoara, Romania, hosted by the West University, March 2015.
- **A self-sustaining business model for open data**: Krems, Austria, hosted by the Danube University, May 2015.
- **Maximizing interoperability – core vocabularies, location-aware data, and more**: Berlin, hosted by Fraunhofer Fokus, November 2015.

This Business Scenario provides input to those workshops, and may be updated in the light of their conclusions.

The W3C Data on the Web Best Practices Working Group will develop standards and best practices in the light of the discussions at and conclusions reached by the SHARE-PSI workshops.

Those standards and best practices will be available for use by public sector and other bodies worldwide, and by SHARE-PSI participants in particular. Promotion of those standards and best practices will be important, to facilitate their adoption.

Glossary of Terms and Abbreviations

API	Application Program Interface; see also *Web Service API*.
App	A computer application that runs on mobile devices.
CSV	Comma-Separated Values – a file format that can be processed by many spreadsheet programs.
EU	European Union
FTP	File Transfer Protocol
GDP	Gross Domestic Product
GPS	Global Positioning System
HTTP	Hypertext Transfer Protocol
IPR	Intellectual Property Rights
JSON	Javascript Object Notation; refer to: www.json.org.
Linked Data	Structured data on the web that follows a particular set of best practices for publication and interlinking [HEATH].
OECD	Organization for Economic Co-operation and Development
OWL	Web Ontology Language
PC	Personal Computer
PDF	Portable Document Format
RDF	Resource Description Framework
RSS	Really Simple Syndication
Structured Data	Data that is organized in identified fields within records or files. Relational databases and spreadsheets are examples of structured data.
SQL	Structured Query Language
UDEF	Universal Data Element Framework – a framework for describing data to enable interoperability (The Open Group).
Unstructured Data	Data other than *structured data*.
W3C	The World-Wide Web Consortium; refer to www.w3.org.
Web Service	A software system designed to support interoperable machine-to-machine interaction over a network (W3C).

Web Service API	An interface by which another system interacts with a *web service* using the World-Wide Web.

References

(Please note that the links below are good at the time of writing but cannot be guaranteed for the future.)

[Anderson] C. Anderson: Free: How today's smartest businesses profit by giving something for nothing, Random House (2010).

[App Annie] Mobile App Advertising and Monetization Trends 2012-2017: The Economics of Free, App Annie & IDC; refer to: www.idc.com/getdoc.jsp?containerId=prUS24771914.

[ATOM] IETF RFC 4287: The Atom Syndication Format, The Internet Society (2005); refer to: www.ietf.org/rfc/rfc4287.txt.

[Capgemini] The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data, Capgemini Consulting; refer to: www.capgemini-consulting.com/resource-file-access/resource/pdf/opendata_pov_6feb.pdf.

[CKAN] CKAN: The Open Source Data Portal Platform; refer to: http://ckan.org.

[Deloitte] Market Assessment of Public Sector Information, written by Deloitte for the UK Department for Business, Innovation and Skills (May 2013); refer to: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198905/bis-13-743-market-assessment-of-public-sector-information.pdf.

[EU 2014] European Parliament/Results of the 2014 European Elections; refer to: www.europarl.europa.eu/elections2014-results/en/election-results-2014.html.

[EU PSI Directive] Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the Re-use of Public Sector Information; refer to: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF.

[FTP] ITEF RFC 959: File Transfer Protocol (FTP), The Internet Society (1985); refer to: www.ietf.org/rfc/rfc959.txt.

[GNSS] GNSS Market Report Issue 2, The European GNSS Agency; refer to: http://egnos-portal.gsa.europa.eu/sites/default/files/content/Market_Report_GSA_2012.pdf.

[HEATH] T. Heath, C. Bizer: Linked Data: Evolving the Web into a Global Data Space (1st Edition), Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool (2011); refer to: http://linkeddatabook.com/editions/1.0.

[HTTP] IETF RFC 2616: Hypertext Transfer Protocol – HTTP 1.1, The Internet Society (1999); refer to: www.ietf.org/rfc/rfc2616.txt.

[Infomediary] Characterization Study of the Infomediary Sector: 2012 Edition, Spanish Government; refer to: http://datos.gob.es/sites/default/files/files/Estudio_infomediario/121001%20RED%20007%20Final%20Report_2012%20Edition_vF_en.pdf.

[Interoperable Enterprise] Business Scenario: The Interoperable Enterprise (K022), October 2002, published by The Open Group; refer to: www.opengroup.org/bookstore/catalog/k022.htm.

[Kroes] N. Kroes: Information is the New Oil, Interview of the European Commissioner for the Digital Agenda, January 4 2012, Laurent Nicolas; refer to: www.thenewfederalist.eu/Neelie-Kroes-Information-is-the-new-oil,04732.

[Louis] How much do Average Apps Make? Tristan Louis in Forbes; refer to: www.forbes.com/sites/tristanlouis/2013/08/10/how-much-do-average-apps-make.

[McKinsey] Open Data: Unlocking Innovation and Performance with Liquid Information, The McKinsey Global Institute; refer to: www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information.

[OECD] Public Sector Innovation and e-Government: OECD Recommendation on Digital Government Strategies, Organization for Economic Co-operation and Development (OECD); refer to: www.oecd.org/gov/public-innovation/recommendation-on-digital-government-strategies.htm.

[OP3.0] Open Platform 3.0™, The Open Group; refer to: http://www.opengroup.org/subjectareas/platform3.0.

[OWL] Web Ontology Language (OWL), The World-Wide Web Consortium; refer to: www.w3.org/TR/owl-overview.

[RDF] Resource Description Framework (RDF), The World-Wide Web Consortium; refer to: www.w3.org/TR/2004/REC-rdf-primer-20040210.

[RSS] Really Simple Syndication (RSS) Specification; refer to: http://validator.w3.org/feed/docs/rss2.html.

[Samos] Uses of Open Data Within Government for Innovation and Efficiency, Workshop of the SHARE-PSI Project in Samos, Greece, June/July 2014; refer to: www.w3.org/2013/share-psi/workshop/samos/report.

[SHARE-PSI] SHARE-PSI 2.0: The Network for Innovation in European Public Sector Information; refer to: www.w3.org/2013/share-psi.

[SPARQL] SPARQL 1.1: RDF Query Language, The World-Wide Web Consortium; refer to: www.w3.org/TR/2013/REC-sparql11-overview-20130321.

[Supervizor] Supervizor: Providing Citizens and Institutions with Information about Public Spending, Republic of Slovenia, Ministry of the Interior; refer to: www.w3.org/2013/share-psi/workshop/samos/supervizor.

[Trentino] I. Bedini et al: Open Government Data: Fostering Innovation; refer to: www.w3.org/2013/share-psi/wiki/images/e/eb/Feroz.pdf.

[UDEF] Universal Data Element Framework (UDEF), The Open Group; refer to: www.opengroup.org/udef.

[UK Social Enterprises] Social Enterprise: Market Trends, The UK Cabinet Office; refer to: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/205291/Social_Enterprises_Market_Trends_-_report_v1.pdf.

[US 2009] The US Open Government Directive of 2009; refer to: www.whitehouse.gov/open/documents/open-government-directive.

[W3C Web Data] W3C Data on the Web Best Practices Working Group; refer to: www.w3.org/2013/dwbp/wiki/Main_Page.

[World Bank] The World Bank: Open Data Essentials; refer to: http://data.worldbank.org/about/open-government-data-toolkit/knowledge-repository.