SC6DI06692

Study on business models for Linked Open Government Data

(BM4LOGD)

Date: 12/11/2013

Document Metadata

Property	Value
Release date	2013-11-12
Status	Public
Version	1.00
Authors	Phil Archer – W3C Makx Dekkers – AMI Consult Stijn Goedertier – PwC EU Services Nikolaos Loutas – PwC EU Services
Project Officers	Vassilios Peristeras – European Commission Susanne Wigard – European Commission

Disclaimer

Table of Contents

List of Tables

List of Figures

EXECUTIVE SUMMARY

This report presents the outcome of a study on Linked Open Government Data (LOGD), a set of principles for publishing, linking and accessing open government data as a service on the Web. The study was commissioned by the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission to address the following research topics:

1. the value of LOGD for businesses, citizens, and public administrations;
2. the cost structures behind the provision of LOGD;
3. the revenue streams linked to the consumption of LOGD services; and
4. enablers and barriers with regard to the value creation of LODG.

The report consists of three main blocks. First a theoretical framework is presented in Chapter 2. Second, the theoretical framework is applied to fourteen case studies in Chapter 3. Finally, Chapters 4, 5 and 6 summarise the findings, list the identified enablers and barriers, and conclude the report.

Chapter 2 provides a theoretical framework to analyse the LOGD ecosystem. The framework is structured according to the nine areas in the Business Model Canvas [Osterwalder]. In the LOGD ecosystem, public administrations are data providers that provide Open Government Data as an online LOGD ***service*** to data consumers – citizens, businesses and other public administrations. Instead of downloading and processing a whole dataset, LOGD allows a data consumer to retrieve specific information about the entity of his interest, by resolving its Web identifier (URI). The data is provided in different machine-readable formats, ready to be linked and meshed-up with other data.
The theoretical framework hence puts the following value proposition of LOGD forward:

- LOGD offers flexible data integration;
- LOGD leads to an increase in data quality;
- the use of LOGD gives rise to new services; and
- LOGD reduces data integration costs.

To enable this value proposition, LOGD providers must have a **URI policy** that lays down the expected service levels of the Linked Data service; long-term persistence being one of the most important service gurantees. Governments should make their URI policy explicit, so that LOGD consumers can rely on LOGD services with confidence and other data providers can link to these URIs or reuse these URIs to denote identical concepts. The use of URIs as *common* identifiers to identify identical concepts in disparate datasets is a prerequisite to unlock the **positive network effects** of LOGD. The theoretical framework also provides a number of favourable conditions under which public administrations could consider providing LOGD:

- **Nature of the data**: There are no restrictions (e.g. no personal data protection and/or privacy concerns).
- **Positive network effects**: the publication of LOGD can reduce the costs of resolving interoperability conflicts in information exchange and ease data

integration. This may be particularly the case for data models and reference data that is used by many in different contexts.

- **No other economic agent wants to/can offer it.** In case no other economic agent wants to or can offer some reference data as LOGD, it may make sense for governments to assume this task, e.g. in the case of Base Registers. Otherwise, providing the LOGD service would compete with the services of other market players.
- **Economies of scale**. Governments may already have the infrastructure in place to provide the service and can provide LOGD with little additional costs.
- **Guarantees of stability and persistence**. Governments are in a good position to guarantee stability and persistence of the LOGD service.

In Chapter 3, thirty seven cases are identified in which public administrations have used LOGD to make open government data available as a service on the Web; fourteen cases have been selected for further analysis according to the aforementioned theoretical framework. The selected case studies are:

- Austria: Renewable Energy and Energy Efficiency Partnership (REEEP);
- Germany: German National Library (DNB);
- European Union: Europeana;
- European Union: European Commission Directorate-General Health and Consumers (DG SANCO);
- European Union: European Environment Agency (EEA);
- European Union: Publications Office of the European Union (OP);
- Italy: Agenzia per l'Italia Digitale (AgID);
- United Kingdom: BBC;
- United Kingdom: Companies House;
- United Kingdom: Department of Environment, Food and Rural Affairs (DEFRA);
- United Kingdom: National Archives;
- United Kingdom: OpenCorporates;
- United Kingdom: Ordnance Survey (OS); and
- International: Food and Agriculture Organization of the United Nations (FAO).

The case studies were carried out in a structured, uniform way according to a predefined 'guide to conducting the case studies' (Annex I) to ensure comparability of the results. The information that was gathered for each case study, both by conducting an interview with the stakeholders and by performing desk research, is contained in Annex II of the report.

Chapters 4 summarises the main findings of the case studies, according to the nine areas of the Business Model Canvas:

- **value proposition**: the main driver for the use of LOGD in the cases investigated is that it allows for flexible data integration (see for example the cases of the BBC, REEEP, DG SANCO and OpenCorporates); this helps to increase data quality by allowing cross-references to authoritative data to be included and may drive future development of new services. The use of LOGD increases the efficiency of the internal operation of the provider and allows them to fulfil their public task more effectively and efficiently (e.g. the cases of DG SANCO and the National Archives). Whether this has given rise to new services or actually reduces costs varies from one case to another. The case studies do not provide evidence that LOGD has contributed to increased data quality, for example via self-service or crowd-sourcing mechanisms. The case studies also reveal that little effort is currently spent on quantitatively measuring the usage and benefits of LOGD. This may be put down to the fact that case study participants have implemented LOGD firstly for their internal data consumption, and are not yet in contact with external consumers.
- **key resources**: LOGD is applied most successfully in reference data; URI design policies are generally in place, while persistence is not often made explicit; many organisations cite a lack of tools that meet their specific need in their specific context; skill and competencies are mostly acquired in-house with some help from external consultants.
- **key partners**: most providers apply LOGD in the context of existing peer networks (such as between EEA and FAO); there is, as yet, little use of LOGD outside of those networks or by businesses.
- **key activities**: in general, providers consider development and maintenance of LOGD services as part of their normal system maintenance and operational activities; few invest in promotional activities, e.g. Europeana and DNB.
- **cost structure**: given that many providers see LOGD activities as part of their core business, the study did not bring out the cost structure of the Linked Data activities alone as most providers do not yet separately account for this; where figures in terms of finances or staff resources were mentioned, e.g. in the cases of FAO and DG SANCO, these spanned a wide range depending on the approach taken.
- **customer segments**: most cases showed either internal use or reuse in existing peer networks of government and non-government organisations; the study did not find much reuse of the LOGD by businesses.
- **revenue streams**: the predominant revenue model is public funding, as part of the normal budgets of the organisations surveyed; in all cases, the data is provided free of charge; licences are either open or not explicitly defined.
- **channels**: distribution channels include direct URI resolution and SPARQL endpoints. Bulk downloads are almost always offered; proprietary apps and Web applications are less common.

- **customer relationships**: there is little branding or advertisement of LOGD services, and little user support; feedback is typically through informal communications as part of institutional collaborations, e.g. in the case of DG SANCO, EEA, OP and the Companies House.

Chapter 5 lists a number of enablers and roadblocks to the value proposition LOGD that were identified in the case studies. The enablers are:

- **Efficiency gains in data integration – the network effect**: providers are more likely to engage in LOGD activities if they can see an immediate benefit for themselves. This was one of the drivers in the cases of REEEP, DNB, DG SANCO, EEA, BBC, TNA, OS and FAO. In the same way that each new telephone added value to the existing ones at the birth of telecommunications, the addition of each new Linked Open Government Data set adds value to those that are already published.
- **Forward-looking strategies**: as providers see the thrust of LOGD they may want to align themselves with modern techniques and technologies as a way to maintain their reputation as thought leaders in their domain, see for example the cases of Europeana, Companies House, TNA, OS.
- **Increased linking and integrated services**: providers who rely on connections with peer organisations, such as REEEP and the BBC, will value the possibilities for easier linking and increased interoperability that LOGD offers.
- **Ease of model updates**: LOGD makes future upgrades of data models much easier, for example to include new data or connect data from different sources together. BBC and DEFRA showed particular interest in this aspect of LOGD.
- **Ease of navigation**: URIs allow a 'follow-your-nose' navigation structure that provides better navigation through complex data as demonstrated for instance in the cases of the BBC and DNB.
- **Open licensing and free access**: LOGD considered in the study is mostly provided free of charge and under open licences which enables further use and reuse of data. Europeana and DG SANCO have worked a lot on releasing their LOGD under an appropriate licence.
- **Enthusiasm from 'champions'**: the knowledge and enthusiasm of individuals in organisations who create awareness of possibilities and potential benefits help organisations to consider engaging in LOGD activities. When their efforts show real benefits, their employers are usually quick to offer support. The role of champions was quite important in the cases of Europeana, AgID, DEFRA, OpenCorporates and OS.
- **Emerging best practice guidance**: availability of guidelines and dissemination of best practices create common approaches and reduce risk in implementation by enabling organisations to learn from each other. Most of the organisations interviewed, such as Europeana, DG SANCO, National Archives and the BBC, are convinced on the importance of sharing knowledge and experiences with others and contribute actively to the development of such best practices.

In addition, the following roadblocks were identified in the case studies:

- **Necessary investments**: as with all new technologies, LOGD requires investments in infrastructure, software and people. Not all organisations may be able to make such investments in a time of shrinking budgets and increased scrutiny.

- **Lack of necessary competencies**: not all organisations have the necessary skills, and specific training materials for a particular domain or application may not be readily available.

- **Perceived lack of tools**: some organisations currently develop their own tools, as there is a perceived lack of production-grade tooling. This was a common concern in many cases, such as the ones of EEA, BBC and DG SANCO. This is perhaps surprising given that Oracle, IBM and YarcData (part of Cray) are already among the companies offering high specification Linked Data systems. Additionally, the European Commission has funded in the context of the LOD2 project a number of open-source tools for Linked Data. Although the performance of RDF stores is a long way short of relational databases which are now highly optimised, resilient, production-grade systems, this should not necessarily be perceived as a barrier, as Linked Data services can run on top relational database environments.

- **Lack of service level guarantees:** the reuse of LOGD services by external third parties is hindered as providers do not yet give explicit service level guarantees. The case studies, in particular those of REEEP, BBC and DG SANCO, show that this is largely because the use of Linked Data is first and foremost for the publisher's own benefit and the availability of the data for third parties is a side effect. Service Level Agreements do exist however in cases where the provision of the infrastructure is outsourced, as in the cases of REEEP and the National Archives.

- **Missing, restrictive, or incompatible licences**: interviewees, such as REEEP and the BBC, report that missing, restrictive, or incompatible data licences continue to be a barrier to providing and consuming LOGD. It is not trivial to keep track of licence information for LOGD, especially when the ownership is not well defined or if data originates from different sources.

- **Surfeit of standard vocabularies**: many Linked Data applications are developed within a specific community with specific agreements using specific standards; although there are common standards like Dublin Core and FOAF, not all implementations use those in the same way giving rise to fragmentation that hinders wide interoperability. This was a common concern among many of the interviewees, including DNB and DG SANCO.

- **The inertia of the status quo**: even more than other types of organisations, public sector bodies tend to favour incremental change such that new systems are seen as ways to replicate the same tasks as old ones. It was observed that in many cases, such as REEEP, AgID, BBC, DEFRA, OpenCorporates and FAO, Linked Data is seen as a more substantial change and therefore meets resistance. Additionally, as LOGD allows connections to be made and relations to be seen that were not visible in non-linked approaches, organisations see the technology as carrying a higher risk than more traditional approaches; uncertainties may lead to delays in adopting new approaches.

Chapter 6 concludes the report. The fact that 37 cases have been identified in which public administrations have made LOGD available demonstrates that LOGD is becoming increasingly adopted. It is particularly important in the provision and management of reference data (e.g. information about organisations, places, and controlled subject vocabularies). The prevalent business model that emerged from the study is the one where the investment and maintenance costs of a LOGD service are covered through on-going public funding with some help from occasional grants. In all cases, the LOGD service is provisioned free of charge. Remarkably, the provision of LOGD to external reusers is in almost all cases not the first objective of the organisations that create Linked Data. More often, it is used to increase efficiency of internal data integration, or to support data exchange in existing collaborations. Many providers do not yet have a clear view of the consumers of their data: in general, the vast majority does not monitor usage, does not offer feedback mechanisms, and does not give guarantees to external parties about the availability and/or the quality of their services.

In the study, we have not seen a wide reuse by third parties that take data from various providers and create new services from such mash-ups. This may be because providers do not yet provide operational guarantees or because Linked Data requires acquisition of new skills on the side of the reusers. As such, reuse by third parties is still very much in the Innovator phase with few examples of new services, such as the Forest Reproduction Material client application of DG SANCO or the upcoming insolvency register based on the official notices in the Gazettes of the United Kingdom. Nonetheless, as data providers are in the process of producing massive amounts of LOGD we can expect that more reusers will find their way to linked data technologies and LOGD, contributing to a thriving ecosystem.

1. INTRODUCTION

This is the report for the "business models for Linked Open Government Data" (BM4LOGD) study that was commissioned by the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission. This chapter explains the study's objectives, scope and approach.

1.1. Context

In Europe, **Open Government Data** (OGD) is seen as an enabler for Open Government and a goldmine of unrealised economic potential[1]. Open Data usually refers to public records (for example on transport, infrastructure, education, and environment) that can be used and redistributed by anyone - either for free or at marginal cost.

But opening-up data, e.g. in Open Data portals, often happens in an ad-hoc manner, and in many cases thousands of datasets are published without adhering to commonly-agreed data and metadata standards and without reusing common identifiers. Hence, a fragmented data-sphere is created where finding, reusing, integrating and making sense of data from different sources is a real challenge.

Linked Open Government Data (LOGD) can respond to these challenges and can be an enabler of eGovernment transformation. LOGD is a way of identifying, linking and accessing OGD Data according to the **Linked Data design principles** put forward by Tim Berners-Lee[2] and reflected in the "5 stars of Linked Open Data"[3]. Linked Data is a convention based on open Web standards such as HTTP URIs and the Resource Description Framework. It enables the provision of "data services" and conceives the Web as an open ecosystem where data owners, data publishers, and data consumers can interconnect and integrate disparate datasets. It converts the Web from a "Web of documents" into a "Web of interconnected data". Applied to eGovernment, LOGD has the potential to lead to smarter and more efficient government services and applications, and to foster creativity and innovation in the digital economy.[4]

[1] Press release: Digital Agenda: Turning government data into gold http://europa.eu/rapid/press-release_IP-11-1524_en.htm

[2] http://www.w3.org/DesignIssues/LinkedData.html

[3] http://5stardata.info/

[4] How Linked Data is transforming eGovernment https://joinup.ec.europa.eu/sites/default/files/D4.3.2_Case_Study_Linked_Data_eGov.pdf

1.2. Objectives

This study tries to answer the following research questions:

1. **Value**: What value does Linked Open Government Data (LOGD) bring to businesses, citizens, and public administrations? For example, can LODG lead to cost reductions? Can LOGD foster semantic interoperability of information exchanges?
2. **Cost structures**: What does it cost to provide LOGD services?
3. **Revenue streams**: Who pays for the provisioning of LOGD?
4. **Barriers and enablers**: What are enablers and barriers with regard to the value creation of LODG?

The table below provides a brief overview of the study objectives.

Table 1 Study objectives

Objective	Description
Objective 1	To analyse the supply-side of the LOGD ecosystem with regards to key resources, partners, activities, and cost models.
Objective 2	To analyse the demand-side of the LOGD ecosystem, including value proposition, customer segments, and revenue systems.
Objective 3	To identify enablers and roadblocks towards realising the benefits of LOGD.

1.3. Study methodology

The study approach consisted of the following steps:

1. **Define LOGD terminology and theoretical framework to analyse the ecosystem**: First, we defined key terminology and set an analysis framework that is referred to throughout the study.
    a. **Glossary**: A glossary of terms is provided in Section 1.4.
    b. **LOGD ecosystem**: The analysis framework is based on relevant existing work. It is outlined in Chapter 2.
2. **Identify, select, and analyse case studies**: Then, we identified, selected, and analysed case studies. A summary of this work is included in Chapter 3:
    a. **Selection criteria**: We specified the criteria to select relevant case studies; and
    b. **Long-list**: We created a long-list of candidates for case studies using a public consultation[5] and desk research.
    c. **Guide to conducting the case studies**: We outlined a guide to help carrying out the case studies in a consistent and harmonised manner (see Annex I). For each case study two interviews were to be conducted: one with the LOGD supplier and, where appropriate, one with an LOGD reuser.
    d. **Collect metrics**, **conduct desk research and interviews:** We collected metrics, conducted desk research and did interviews.
    e. **Quality control.** We invited the organisations participating in each case study to validate the gathered information and findings.
3. **Summarise findings**: On the basis of 14 case studies, we generalised and summarised our findings.
    a. **Findings:** in chapter 4, we present the findings in terms of the LOGD analysis framework presented in chapter 2
    b. **Enablers and roadblocks:** in chapter 5, we list the enablers and roadblocks for LOGD that we encountered in the case studies.
    c. **Overall conclusions:** chapter 6 contains the overall conclusions of the study.

[5] https://joinup.ec.europa.eu/node/67268

1.4. Glossary

This section provides a number of common definitions used throughout the study.

1.4.1. Open Government Data (OGD)

The term Open Government Data (OGD) refers to data and information produced or commissioned by government or government controlled entities[6] and that is made available for reuse for private or commercial purposes, with minimal or no legal, technical or financial constraints[7]. According to the amendment to the PSI Directive 2013/37/EU, Open Data policies aim to promote the circulation of information not only for economic operators but also for the public, they can play an important role in kick-starting the development of new services based on novel ways to combine and make use of such information, stimulate economic growth and promote social engagement. Examples of OGD are public data on legislation, companies, transport, infrastructure, education, health, crime, environment, etc.

According to the "Open Data whitepaper - Unleashing the Potential" published by the UK Cabinet Office[8] OGD must be:

- **Accessible** (ideally via the Internet) at no more than the cost of reproduction, without limitations based on identity or intent;
- In a **digital, machine-readable format** for interoperation with other data; and
- **Free of restriction of use or redistribution** in its licensing conditions.

1.4.2. Linked Data

Linked data is a set of principles for publishing structured data so that it can be interlinked and become more usable. It builds upon standard Web technologies such as HTTP and URIs, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried[9,10].

Tim Berners-Lee defined four design principles (rules) of Linked Data[11]:

1. Use Uniform Resource Identifiers (URIs) to uniquely identify things (data entities);
2. Use HTTP URLs, corresponding to these URIs, so that information can be retrieved;
3. Provide metadata using open standards such as RDF;
4. Include links to related URIs, so that people can discover more things.

---

[6] OKFN, http://opendefinition.org/

[7] DIRECTIVE 2013/37/EU amending Directive 2003/98/EC on the reuse of public sector information http://eur-lex.europa.eu/JOHtml.do?uri=OJ:L:2013:175:SOM:EN:HTML

[8] http://www.cabinetoffice.gov.uk/resource-library/open-data-white-paper-unleashing-potential

9 Bizer, Heath & Berners-Lee, Linked Data—The Story So Far". International Journal on Semantic Web and Information Systems 5 (3): 1–22. doi:10.4018/jswis.2009081901. ISSN 15526283. Retrieved 2010-12-18.

10 http://www.w3.org/standards/semanticweb/data

11 http://www.w3.org/DesignIssues/LinkedData.html

The four principles of Linked Data referred to above were developed by Tim Berners-Lee in 2006. Those principles do **not** mention openness. The 'O' in Linked Open Data came from later projects, notably the W3C Linking Open Data project[12] and, more recently, LOD2[13]. These have given rise to the common view that linked data and linked open data are synonymous. However, this is not always the case.

In the financial sector, Garlik[14], part of Experian, makes extensive use of state-of-the-art triple stores - none of which are publicly accessible. Some companies, such as the BBC and Fujitsu European Labs, make extensive use of open data to augment their internal data[15]. On the other hand, pharmaceutical companies, such as AstraZeneca, are experimenting with the use of linked data to integrate internal data sources[16,17]. There is a growing recognition in the open data world that there is a need to engage with and interact positively with 'closed data.'

1.4.3. Linked Open Government Data (LOGD)

Open Government Data published according to the Linked Data design principles.

1.4.4. URI Policy

With URI Policy we mean the explicit or implicit rules that an organisation has for guaranteeing the long-term persistence, resolvability, and uniformity of Web identifiers (HTTP URIs).

[12] http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
[13] http://lod2.eu/
[14] http://www.garlik.com/
[15] http://www.w3.org/2013/04/odw/report#closed
[16] http://semanticweb.com/sindicetech-helps-enterprises-build-private-linked-data-clouds_b30454
[17] http://www.slideshare.net/kerfors/linked-data-in-pharma

Persistent and resolvable identifiers

The promise of open data is that it can help to improve transparency and efficiency within government, and to stimulate new economic growth. One important aspect of this is the ability to mix datasets created by one department with those created by another. Only by providing data in a manner that is cross-border and cross-sector can the work of one part of government be reused by another or an outside agency. Fundamental to this concept of reuse is the provision of common identifiers for buildings, roads, departments, places etc.

There exist may types of identifier schemes, typically developed within a particular environment for a particular purpose. Taken out of context such identifiers are meaningless. Equally problematic are identifiers that change over time.

The need therefore is for a set of identifiers that are:

- applicable in any context and preserve their meaning;
- unaffected by future changes in the name and structure of the organisation that creates them;
- likely to be unaffected by future changes in technology.

These are hard criteria to meet but, if designed and managed with persistence in mind, attainable using URLs.

A URL identifies a location on the Web. An HTTP URI, that is, one that begins http://… is identical in structure but may identify anything, including real-world objects. HTTP URIs (henceforth simply called URIs) have several distinct advantages over other identifier types.

First and foremost they can be de-referenced, that is, looked up. Put a URI in a Web browser and, all being well, you get back the thing it identifies. Where the identified thing is a real-world object then you'll be redirected to a *description of that thing*. Either way, you get back data. This standard Web architecture is at the heart of linked data and uses no novel technology. Importantly, a server can return data in multiple formats through a process called content negotiation. If you dereference a URI using a Web browser it will return HTML but if some other application dereferences the same URI it may declare as part of its request that it wants data back in XML, or RDF, or JSON, or any other technology and receive that instead. This is as future proof as it's possible to be since new formats can be added to the server at any time.

Not only is this future proof in terms of the data that can be returned, the whole server back end can be replaced with new technology should it be necessary - as one day it surely will - but the URI itself can survive that unchanged. Not all applications will need to dereference URIs - that's fine. They work as strings too!

Persistent URIs are the technology-neutral key to cross-border, cross-sector data exchange. They enable data matching in any format including linked data.

A study on the provision of persistent URIs that includes 12 cases and one counter case is available on Joinup [PURI].

2. LOGD ECOSYSTEM

In this chapter, we analyse the LOGD ecosystem and set the theoretical framework that is used to analyse the case studies.

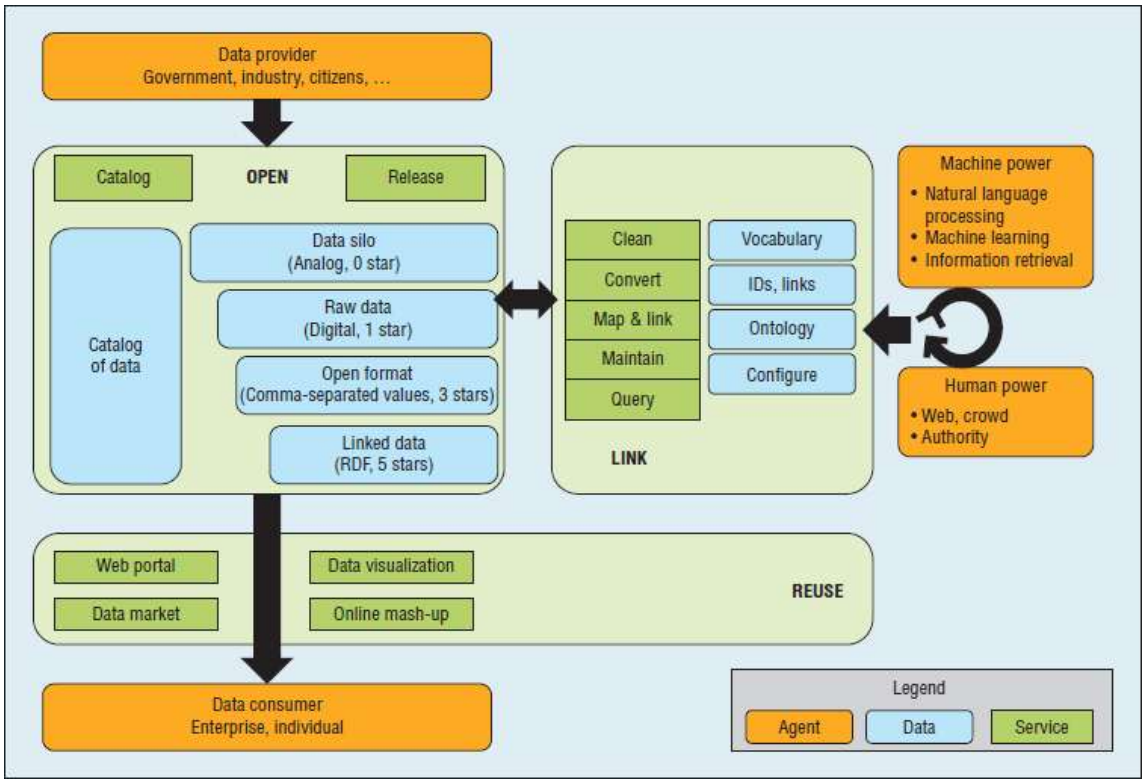


Figure 1 Roadmap of linked open government data [Ding]

Extending the work of [Ding] (see Figure 1), we distinguish four main actors that partake and interact in the LOGD ecosystem:

- **Data providers**, namely public administrations that open up their data and deliver it as Linked Open Government Data (performing the operations that are part of the 'Link' step in Figure 1). In this work, we investigate the motivation and the business models that drive LOGD provisioning.
- **Data consumers**, namely citizens, entrepreneurs, businesses and public administrations that reuse available LOGD in the context of value-added applications and services. Noticeably, the distinction between data providers and consumers is not crisp. An organisation that provides data may at the same time also consume the data of another entity. This intertwining is particularly common in the case of LOGD.
- **Data brokers**, namely third-party organisations, either private or public, that run data catalogues and marketplaces which facilitate the access to available LOGD. In some cases, data brokers also offer additional services, such as advanced querying, data visualisation, exporters of data in different formats,
- **Regulatory entities**, namely national/regional/local public administrations and transnational institutions, such as the European Commission, which regulate LOGD provisioning through policies, laws and directives.

In the remainder of this chapter the LOGD ecosystem is further analysed according to the nine areas in the Business Model Canvas [Osterwalder]:

- **value proposition**: an overall view of an organisation's bundle of products and services that are of value to the customer;
- **key resources**: the arrangement of activities and resources that are necessary to create value for the customer;
- **key partners**: a voluntarily initiated cooperative agreement between two or more organisations in order to create value for the customer;
- **key activities**: the ability to execute a repeatable pattern of actions that is necessary in order to create value for the customer;
- **cost structure**: the representation in money of all the means employed in the business model;
- **customer segments**: a segment of customers an organisation wants to offer value to;
- **revenue streams**: the way an organisation makes money through a variety of revenue flows;
- **channels**: a means of getting in touch with the customer; and
- **customer relationships**: the kind of link a company establishes between itself and the customer.

2.1. Value proposition

Table 2 provides the elements according to which the value proposition of LOGD can be classified and analysed.

Table 2 Value proposition of Linked Open Government Data

Value proposition	Description
Flexible data integration	LOGD facilitates data integration and enables the interconnection of previously disparate government datasets.
Increase in data quality	The increased (re)use of LOGD triggers increasing demands to improve data quality. Through crowd-sourcing and self-service mechanisms errors are progressively corrected.
New services	The availability of LOGD gives rise to new services offered by the public and/or private sector.
Cost reduction	The reuse of LOGD in eGovernment applications leads to considerable cost reductions.

Linked Data as a Service

Linked Data entails that data is provided as a service on the Web.

Linked Data as a Service (LDaaS) supports the on-demand access to data regardless of the physical, geographical or organisational separation between the Linked Data provider and the consumer. It can help decouple the application from the data and consequently contributes to the reusability of the latter in different contexts and applications.

Instead of downloading a whole dataset, LDaaS allows a data consumer to retrieve at real-time a specific piece of data by resolving its URIs in different formats, ready to be integrated and meshed-up with other data. This is actually expected to spark the development of value-added applications and services that capitalise on the on-demand reuse of data.

A high-quality, trusted LDaaS offering is based on the following pillars:

- **Persistent and resolvable URIs** for the data resources and for the accompanying metadata;
- **Commonly-agreed data models and controlled vocabularies** for representing both the data and its metadata; and
- **A scalable** (possibly Cloud-based) **infrastructure** that comprises both data warehousing and querying aspects, provides RESTful access to data and delivers it in different formats.

LDaaS approach delivers the following benefits:

- **Agility and timeliness**: Reusers can access the data they need when they need it. There is no longer a need to download whole datasets and databases as data can be consulted online.
- **Cost-effectiveness** (both for data providers and reusers): On the one hand, different data providers can make use of the same LDaaS infrastructure for serving their data as Linked Data. This can lead to significant economies of scale and cost savings. On the other hand, reusers can save on data storage, hosting and maintenance costs, as the Linked Data stays with the provider and does not have to be stored locally in order to be reused and integrated with other data.
- **Improved data quality**: Linked Data has three aspects which have a positive impact on data quality:
    - Data is assigned persistent and resolvable URIs, thus making it possible to resolve the ambiguity of things, in particular real-world entities, but also to create links between the alternative identifiers of the same thing.
    - Data is fresh. Using LOGD, reusers can have access to the most up-to-date version of the data at any time. A single version of the truth is hence guaranteed, as opposed to the proliferation of different versions of a dataset downloaded over a period of time.

The positive network effects of Linked Open Government Data

Linked Data, or the practice of linking data, has the potential of dramatically reducing the semantic interoperability conflicts that can arise when data is exchanged between economic agents or when data from different sources has to be integrated. Put positively, Linked Data can be seen as a technology for flexible data integration that can increase the number of interconnections between information systems of public administrations and businesses. The **network value** can be correlated to the number of possible connections to other datasets in a network.

This can be illustrated by a case where N information systems are using N different reference datasets (e.g. country codes) to be integrated. When organisations need to exchange data, mappings need to be created between these N reference datasets. The number of required bi-directional mappings in such a network of N reference datasets is N*(N-1). Take now the case of having one single, commonly agreed dataset that all public administrations could reuse. In this case, the number of mappings would add up to N. In such a setting the network value of LOGD increases more than proportionally to the number of datasets to interconnect. This is illustrated in the figure below.

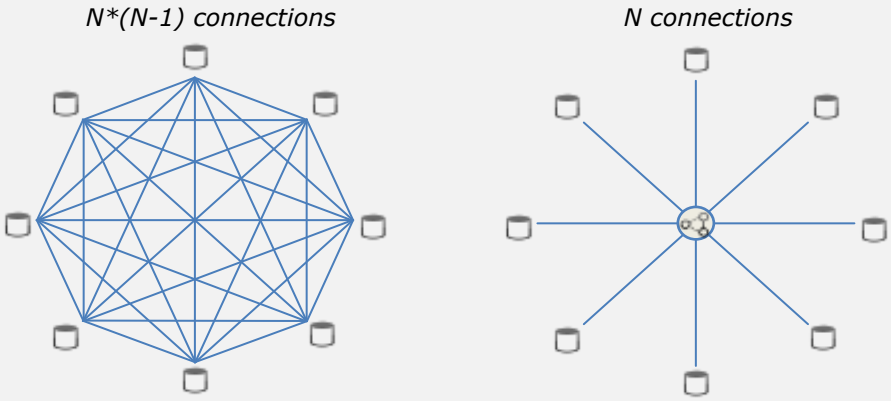


Figure 2 Number of interconnections

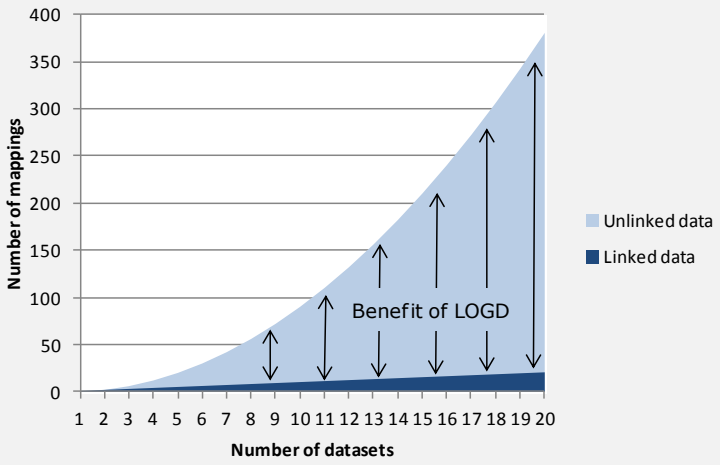


Figure 3 Benefit of LOGD

2.2. Key resources

This section provides the elements to analyse which key resources public administrations use for supplying LOGD.

Table 3 Key resources

Key resources	Description
Data	PSI domains [MEPSIR]: • **Geographic information** [MEPSIR]: including address information, aerial photos, buildings, cadastral information, geodetic networks, geology, hydrographical data and topographic information; • **Business information** [MEPSIR]: including Chamber of commerce information, official business registers, patent and trademark information and public tender databases; • **Legal information** [MEPSIR]: including decisions of national, foreign and international courts, national legislation and treaties; • **Meteorological information** [MEPSIR]: including climate data and models and weather forecasts; • **Social data** [MEPSIR]: including various types of statistics (economic, employment, health, population, public administration, social); • **Transport information** [MEPSIR]: including information on traffic congestion, work on roads, and public transport, and vehicle registration; • **Cultural heritage information**: including cultural heritage information from libraries, museums, and archives; • **Research and educational information:** including research and educational information from schools and universities.
URI policy	The policy that the public administrations has for guaranteeing persistence, resolvability, and uniformity of Web identifiers (HTTP URIs). The following options exist: • **No explicit URI policy.** • **Explicit URI policy.**
Linked data infrastructure	The technical Web infrastructure required to make data available as Linked Data.
Skills and competencies	The in-house or contracted competence of the public administration to develop and maintain linked data.

2.3. Key partners

This section provides the elements to analyse which key partnerships public administrations have for supplying LOGD.

Table 4 Partners

Partners	Description
Government:	• **Regulators;** • **eGov agencies; and** • **other public administrations.**
Businesses	• **ICT providers;** • **Data providers;** • **Data brokers.**
Non-governmental organisations	• **Lobbying groups, civic engagement communities, action groups.**

2.4. Key activities

This section provides the structure according to which the key activities that public administrations perform to deliver value to the consumers of LOGD.

Table 5 Key activities

Key activities	Description
Development	All activities that are required to identify, model, transform, harmonise, and publish and / or reuse LOGD. W3C's "Best practices for publishing Linked Data"[18] cite different approaches for modelling the lifecycle of Linked Data[19].
Maintenance	Normal activities to maintain LOGD, such as user support, server maintenance, etc.
Promotion	Activities related to the promotion of the use of LOGD by target consumers (e.g. other public administrations or businesses).

The Linked Data lifecycle

The lifecycle of Linked Data, in our case LOGD, comprises two sides; a supply and a demand one (see also Figure 4).

The supply of LOGD starts with the selection of datasets to be made available as LOGD. There is a close relationship between this step and the business model that the data provider has selected for provisioning LOGD. Several dimensions can be considered here, such as

[18] https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html

[19] http://www.w3.org/2011/gld/wiki/GLD_Life_cycle

- **Legal requirements**: Is there a law that makes open publication mandatory or is there no specific obligation?
- **Relation to public task:** Is the data the direct result of the primary public task of government or is it a product of a non-essential activity?
- **Current status of open publication:** Is the data already openly available or does it still need to be opened up?
- **Type of value:** Is the data useful for social engagement or does it have commercial value?
- **Audience:** Is the data primarily intended for the public or is it primarily aimed at back-office integration?

Once identified, data must be modelled, i.e. cleansed, structured, represented using common – standardised – vocabularies and assigned URIs. Finally, data providers must publish their data online and, wherever possible, link it with other data.

On the demand side, LOGD consumers can search for available LOGD on the portals of data providers or on data brokers, using each time the available SPARQL endpoint.

Once the desired data is found, it can be retrieved in the appropriate format, and reused in the context of a new (or existing) value-added service or application.

A very important aspect of the lifecycle, is the feedback loop between data consumers and data providers, which ensures that the expectations of the consumers are made known to the providers and that feedback is effectively communicated. As such, the feedback loop improves the quality of the LOGD provisioned by a particular data provider, but also the quality of the Linked Data service as a whole.

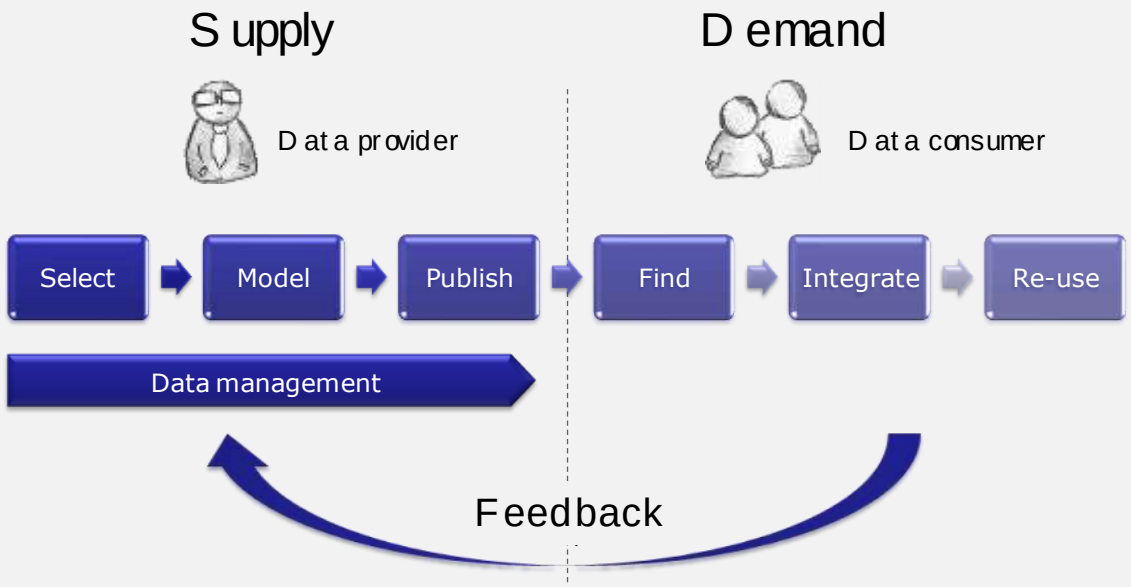


Figure 4 Linked Data lifecycle

For detailed information on how to publish LOGD, please refer to

- W3C Linked Data Cookbook, http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook
- Open Data Support[20] project, Training Module 2.1 'LOGD lifecycle', http://www.slideshare.net/OpenDataSupport/the-linked-open-government-data-lifecycle

2.5. Cost structures

The case studies will analyse the costs related to the publication, maintenance, and promotion of LOGD.

Table 6 Cost structures

Cost driver	Description
Development costs	The cost of all activities that are required to identify, model, transform, harmonise, publish and / or reuse LOGD
Maintenance costs	The cost of publishing updates of the data and / or maintenance costs of relevant infrastructure.
Promotion costs	The cost of promoting the availability of the data as linked data compared with the cost of promoting its availability through other means.

2.6. Customer segments

The customer segments can be analysed as listed in the table below.

Table 7 Customer segments

Customer segment
Government and NGOs: • public administrations; • NGOs.
Businesses: • Data brokers; • Service developers; • Media & data journalists.
Academia

[20] http://www.opendatasupport.eu

2.7. Revenue streams

The revenue systems for LGD can be analysed as indicated in Table 8.

Table 8 Revenue systems

Pricing model type	Description
Sources of revenue	• **Public funding:** LOGD is (partially) funded via dedicated government budgets. • **Usage fees:** LGD is (partially) funded via subscription fees. • **Advertisement:** LGD is (partially) funded through advertisement.
Pricing model	• **Free of charge** • **Freemium** • **Premium**
Price structure	• **Subscription fee** • **On-demand fee**
Licence types	• **Unknown licence:** The Linked Data service does not indicate under which licence the data may be used, modified, redistributed, etc. • **Attribution restrictions:** The licence requires giving appropriate credit to the author. • **Commercial use restrictions:** The licence restricts the reuse for commercial purposes. • **Share-a-like restrictions:** Share alike (copyleft) means that, in case of redistribution of the work, this (same) licence must be reused. • **Public domain:** the licence dedicates the work to the public domain by waiving all rights under copyright law, including all related and neighbouring rights, "to the extent allowed by law". • **Other:** (specify)

Charging for Government Data in Europe

In the LOGD ecosystem not all data is offered for free. For some data reusers need to pay in order to acquire it. According to Pollock there are three pricing policies for Public Sector Information:

1. **Profit-maximisation:** "Setting a price to maximise profit given the demand faced by the PSB. Where the product being supplied does not face competition then this will naturally result in monopoly pricing."
2. **Cost-recovery:** "Setting a price equal to average long-run costs (including, for example, all fixed costs related to data production)."
3. **Marginal costs and zero costs:** "Setting a price equal to the short run marginal cost of supplying data".

However, charging for access to Government Data is restricted by the PSI Directive [2003/98/EC]. Article 5 limits public sector bodies (PSBs) to generating "an income that should not exceed the cost of collection, production, reproduction and dissemination, together with a reasonable ROI" [POPSIS]. However the directive also urges Member States to encourage public sector bodies to make documents available at charges that do not exceed the marginal costs for reproducing and disseminating the documents."

In 2013 the PSI Directive was subjected to a revision. The result is a revised PSI Directive limiting the pricing of PSI, and thus all government data, by PSBs to the marginal cost for reproduction and dissemination as the default regime. The marginal cost is the cost of supplying data to an extra user. When thinking about providing digital information, this means that the marginal costs of Linked Data are virtually zero. This means that de-facto, LOGD will need to be free of charge when it falls within the scope of the revised PSI Directive. This means that PSBs will have to find other means to reach ROI.

The Pricing Of Public Sector Information Study (POPSIS)[21], conducted on behalf of the European Commission and published in October 2011, assessed different models for charging for government data through the analysis of 21 Public Sector Body (PSB) case studies and their impact on reuse and value creation. The study covered PSBs applying models from charging zero and marginal cost to partial and full cost-recovery regimes. The most important conclusion of the case study analysis is that the potential benefits of lowering charging for government data can be high and the cost for lowering these charges is relatively low.

[21] http://epsiplatform.eu/content/popsis-assessment-psi-charging-policies

Deciding whether or not to provide LOGD

To analyse the value proposition of Linked Data, economists consider it helpful to analyse how Linked Data fits into the classification of economic goods [Cobden]. In economics, a **public good** is a good that is both non-excludable and non-rivalrous in that individuals cannot be effectively excluded from use and where use by one individual does not reduce availability to others [Buchanan]. In contrast, Linked Data should be classified as a **club good**[22], as it is:

- **Excludable**. With Linked Closed Data, the consumption of Linked Data can be controlled via access control mechanisms.
- **Non-rivalrous**. To a certain extent, Linked Data can be provided in a scalable manner to a virtually infinite number of users.

Governments need to decide whether or not to provide LOGD services. The following argumentation may apply in favour of LOGD:

- **Nature of the data**: There are no restrictions (e.g. no personal data protection and/or privacy concerns).
- **Positive network effects**: the publication of Linked Data may significantly increase semantic interoperability, reduce the costs of resolving interoperability conflicts in information exchange and ease data integration. Governments may decide that making the LOGD available at zero costs is a welfare maximising strategy. This may be particularly the case for data models and reference data that is used by many in different contexts.
- **No other economic agent wants to/can offer it.** In case no other economic agent wants to or can offer some reference data as Linked Data, it may make sense for governments to assume this task, e.g. in the case of Base Registers. Otherwise, providing the Linked Data service would compete with the services of other market players.
- **Economies of scale**. Governments may already have the infrastructure in place to provide the service and can provide Linked Data with little additional costs.
- **Guarantees of stability and persistence**. Governments are in a good position to guarantee stability and persistence of the Linked Data service.

[22] http://www.econlib.org/library/Enc/PublicGoods.html

2.8. Channels

Table 9 provides a categorisation of different channels through which public administrations can deliver LOGD.

Table 9 Channels

Channel	Description
Web API	LOGD is accessible through a Web API (e.g. a SPARQL endpoint, a de-referenceable URI). Please note that the resolvability of a URI is a requirement for data to be classified as Linked Data.
Bulk Download	LOGD is part of a databank. For example, a location where an entire dataset can be downloaded.
Proprietary App	LOGD is accessible via an application on a proprietary platform such as iOS or Android.
Web App	LOGD is accessible via a Web application

2.9. Customer relationships

The customer relationships are the mechanisms used to establish a long-term relationship with the customer. The table below lists a number of possible mechanisms for LOGD.

Table 10 Customer relationships

Relationship mechanisms	Description
Branding	LOGD has a brand strategy that relates to its value proposition.
Advertisement	Consumers are informed of Linked Data Services via advertisement.
User support	A contact centre is used to support users in using LOGD.
User feedback mechanisms	User feedback mechanisms are in place to measure the user satisfaction and general feedback.

3. CASE STUDIES

This chapter outlines the selection and analysis of 14 case studies, and summarizes the main findings.

3.1. Selection criteria

We aimed at identifying a long list of case studies around the supply and reuse of Linked Open Government Data. Case studies were shortlisted based on criteria such as:

- **Geographic coverage**: we aimed at selecting case studies in the largest possible variety of Member States;
- **Covered PSI domains**: we aimed at selecting case studies covering the largest possible variety of PSI domains (geographical information, business information, legal information, meteorological information, social data, and transport information. Cultural/library information was added following the update to the PSI Directive).
- **Readiness**: we aimed at selecting case studies with organisations that have already deployed a publicly accessible Linked Data infrastructure and showed interest to participate in the study.

3.2. Long list of case studies

Following desk research and consultation[23] of the general public, more than 30 candidate case studies on LOGD were identified, as listed in Table 11. Applying the above listed selection criteria, 14 case studies were selected for further analysis.

Table 11 Long list of candidate case studies

CC	PSI domain(s)	Case study	Selected ?
AT	Geographical	AT - Renewable Energy and Energy Efficiency Partnership (REEEP) http://www.REEEP.org/	✓
AT	Geographical	AT - Austrian Geological Survey (GBA) http://www.geologie.ac.at/services/thesaurus/	
BE	Cultural/library	BE - Vlaams Theater Instituut – Travelogue http://vti.be/nl/linked-data	
DE	Cultural/library	DE - German National Library http://dnb.de/EN/lds	✓

[23] 2013-06-13, Call for case studies on Linked Open Government Data | Joinup - https://joinup.ec.europa.eu/node/67268

CC	PSI domain(s)	Case study	Selected ?
DK	Business, Geographical	DK - Danish Agency for Digitisation http://uk.fm.dk/publications/2012/good-basic-data-for-everyone/	
EU	Social data	EU - Buildings Performance Institute Europe http://www.buildingsdata.eu/	
EU	Cultural/library	EU – Europeana http://data.europeana.eu	✓
EU	Social data	EU - European Commission - Directorate General for Consumers & Health http://ec.europa.eu/semantic_webgate_acceptance/query/	✓
EU	Social data	EU - European Environment Agency (EEA) http://semantic.eea.europa.eu/	✓
EU	All	EU - European Union Data Portal http://open-data.europa.eu/	
EU	Social data	EU – Eurostat http://eurostat.linked-statistics.org/	
EU	Legal	EU - Publications Office of the European Union http://publications.europa.eu/mdr/	✓
EU	Social data	EU - Renewable Energy and Energy Efficiency Partnership http://data.reegle.info/ , http://api.reegle.info/	
ES	Meteorological	ES - AEMET – Spanish Meteorological Office http://aemet.linkeddata.es/	
IT	Business, Geographical	IT - Agenzia per l'Italia Digitale http://www.digitpa.gov.it/	✓
IT	Geographical	Regione Emilia-Romagna http://geoportale.regione.emilia-romagna.it/en	
IT	Geographical	IT - Trentino government linked open geo-data http://sgc.disi.unitn.it:8080/sgcmashup/	
NL	Geographical	NL - Amsterdam-Amstelland Fire Department http://netage.nl/en/	
NL	Geographical	NL - Building and address register http://lod.geodan.nl/, http://bag.vrom.nl/	
NL	Social data, business data	NL – Stelselcatalogus: linked metadata of Dutch base registers http://www.linkeddataoverheid.nl https://data.overheid.nl/linkeddata	
NO	Business	NO – Enhetsregisteret – Norwegian Company Register http://sws.ifi.uio.no/enhetsregisteret/, http://data.computas.com/	
SE	Cultural/library	SE - National Union Catalogue - Libris http://librisbloggen.kb.se/tag/linked-data/	
SE	Business	SE - Bolagsverke - Swedish Company Register http://www.skatteverket.se/	
SE	Cultural/library	SE - Swedish National Heritage Board - SOCH http://www. k**samsok.se**	✓

CC	PSI domain(s)	Case study	Selected ?
UK	Cultural/library	UK – BBC http://www.bbc.co.uk/blogs/internet/posts/Linked-Data-Connecting-together-the-BBCs-Online-Content	✓
UK	Business	UK - Companies House http://data.companieshouse.gov.uk/doc/company/03580655 (example)	✓
UK	Geographical	UK - Department of Environment, Food and Rural Affairs (DEFRA) http://data.gov.uk/location	
UK	social data	UK - Effective Service Delivery Toolkit (ESD-Toolkit) http://www.esd.org.uk/esdtoolkit	
UK	Legal	UK - National Archives http://legislation.gov.uk, http://www.gazettes-online.co.uk/	✓
UK	Business	UK – OpenCorporates http://opencorporates.com/	✓
UK	Geographical	UK - Ordnance Survey http://data.ordnancesurvey.co.uk/	✓
UN	Social data	UN - Food and Agriculture Organisation of the United Nations (FAO) http://aims.fao.org/standards/agrovoc/about	✓
US	Social data	US - Clinical quality information about US hospitals http://www.healthdata.gov/cqld	
US	Cultural/library	US - Library of Congress http://id.loc.gov/	
US	Cultural/library	US - Pacific Northwest National Library http://www.pnnl.gov/	
-	Cultural/library	OCLC http://www.oclc.org/data	

3.3. Overview of selected case studies

This section provides an overview of the collected metrics, findings, and enablers and roadblocks for each case study. More detailed information can be found in Annex II.

3.3.1. AT – Renewable Energy and Energy Efficiency Partnership (REEEP)

AT – Renewable Energy and Energy Efficiency Partnership (REEEP) http://www.REEEP.org/		
METRICS		
LOGD usage	#queries / hits #governmental reusers #commercial reusers	Not recorded separately from other server stats. Reegle has about 200,000 unique users per month
LOGD revenue	Subscription fees On-demand fees	None
LOGD costs	development cost maintenance costs promotion costs	The LOGD costs are not separated out from the cost of running Reegle. LOGD aspects estimated to be less than €1M over 10 years
LOGD Benefits	# integrated datasets	Reegle gets info from NREL (OpenEI), FAO, Eurostat, World Bank, DBpedia and the UN
	#derived applications	The reegle.info website, widgets, the Reegle tagging API, a WordPress plug-in
ANALYSIS		
Value proposition	**Flexible data integration**: The collection and sharing of information in the specific domain of renewable energy, particularly in developing countries.	
Key resources	**URI policy:** URI design policy **Data:** Substantial datasets aggregated from various sources, the ability to identify subject matter/key words in natural language documents automatically and collate human-readable information.	
Key partners	**Government agencies:** the US National Renewable Energy Laboratory, NREL.	
Key activities	**Maintenance:** The promotion of renewable energy, climate change management etc. through dissemination of relevant information and project work.	
Cost structures	Not separated from the cost of running the Reegle service as a whole. A service fee is paid to SWC.	
Customer segments	Governments, NGOs, energy companies - but all data is free.	
Revenue systems	Donations and project funding	
Channels	Web APIs (SPARQL endpoint, Tagging API), Web apps (widgets) Web site	
Customer relationships	Governments, donor organisations	
FINDINGS		
Business model	**Public funding.** Although REEEP is not a public body, much of its income is derived from the public sector through grants with remainder made up from donations.	

Enablers	**Efficiency gains in data integration** **Increased linking and integrated services** **Ease of navigation** REEEP can make a good argument about lowering the costs, reducing duplication and re-using data. This fits in with governments' desires to push freely available open data
Roadblocks	**Missing, restrictive or incompatible licences** **The inertia of the status quo** A lot of relevant data is held by organisations with a history of restrictive licences.

3.3.2. DE – German National Library (DNB)

DE – German National Library (DNB) **http://dnb.de/EN/lds**		
METRICS		
LOGD usage	#de-referenced URIs / queries	(not available)
	#governmental reusers	(not available)
	#commercial reusers	None
LOGD revenue	Subscription fees	None
	On-demand fees	None
LOGD costs	development cost	221 person days
	maintenance costs	Ca. 1 FTE
	promotion costs	(not available)
LOGD Benefits	# integrated datasets	(not available)
	#derived applications	(not available)
ANALYSIS		
Value proposition	**Flexible data integration:** Linked Data allows integration of third-party services and unique identification of persons, resources and locations in cultural heritage data.	
Key resources	**URI policy**: DNB has a URI policy that is in line with best practice. **Linked data infrastructure**: DNB uses a toolkit Metafacture that supports conversion of semi-structured data. It is used to export the internal data to RDF/XML. On-the-fly conversions are done for URI resolution. Full RDF dumps are prepared at irregular intervals. Effective from 2014, RDF dumps will be available three times per year (months 1, 5, and 9). **Competencies**: Skills are all in-house.	
Key partners	German library networks, Staatsbibliothek zu Berlin - Preußischer Kulturbesitz (responsible for the Deutsche ISIL-Agentur und Sigelstelle) and other national libraries.	
Key activities	• **Development** of mappings between internal database format and RDF vocabularies. • Implementation of data conversions • Standards work	
Cost structures	221 person days in initial development; 1 FTE for on-going work but many of the activities involved in supplying the service are part of the general bibliographic services which makes it impossible to specify the exact cost of the supply of linked data.	
Customer segments	Cultural heritage institutions.	
Revenue systems	Public funding	
Channels	URI resolution, bulk download.	
Customer relationships	Promotion through presentations, web page (http://dnb.de/EN/lds), articles in journals, participation at book fair and library and information related fairs.	
FINDINGS		
Business model	**Public funding:** DNB provides Linked Data in accordance with its public mandate to disseminate its data as widely as possible. Funding of the activities is through on-going public funding. The data is provided free of charge under CC0 licence.	

Enablers	**Efficiency gains in data integration** Linked Data allows easy integration with existing data when related resources are discovered.
Roadblocks	**Surfeit of standard vocabularies** Absence of widely agreed vocabularies and application profiles.

3.3.3. EU – Europeana

EU – Europeana **http://data.europeana.eu**		
METRICS		
LOGD usage	#de-referenced URIs / queries	4.000 per month
	#governmental reusers	(not available)
	#commercial reusers	(not available)
LOGD revenue	Subscription fees	None
	On-demand fees	None
LOGD costs	development cost	(not available)
	maintenance costs	(not available)
	promotion costs	(not available)
LOGD Benefits	# integrated datasets	807 datasets
	#derived applications	(not available)
ANALYSIS		
Value proposition	**Flexible data integration & new services:** Linked Data allows more expressiveness in the metadata in particular in expressing relationships between objects, parts of objects, structures and rights. This enables Europeana to link together information supplied by different data providers and thereby enhance the presentation of information to its users.	
Key resources	**URI policy**: Europeana's URI policy is described in the ISA study on Persistent URIs[24]. **Linked data infrastructure**: Europeana aggregates data from cultural heritage institutions and from regional and domain-related aggregators. Data is converted from the submissions and from the existing content (which is based on a simpler format) into Europeana Data Model-compliant data. **Skills and Competencies**: Mostly in-house staff, some contributions from other projects.	
Key partners	Cultural heritage organisations, national, libraries, The European Library	
Key activities	**Development:** Conversion of data based on Europeana Semantic Elements to Europeana Data Model; linking with other Linked Data collections (e.g. GeoNames, GEMET, local SKOS-based thesauri)	
Cost structures	Cost is integrated in normal project activities and not separately visible.	
Customer segments	Cultural heritage institutions.	
Revenue systems	Public funding	
Channels	Europeana API, SPARQL endpoint	
Customer relationships	Promotion through presentations	
FINDINGS		

[24] https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris

Business models	**Public funding**: Europeana is currently experimenting with the provision of Linked Data. The business model is based on recurrent public funding for the whole of its activities and free access to data, licensed under CC0 Public Domain Dedication. In the future, premium or freemium models may be considered.
Enablers	**Forward-looking strategies** **Open licensing and free access** **Enthusiasm from 'champions'** **Emerging best practice guidance** **Emerging best practice guidance** Linked Data contributes to changing perspectives of content owners towards more open models.
Roadblocks	**Necessary investments** In the initial phase, cost outweighs the immediate benefits.

3.3.4. EU – European Commission Directorate-General for Health and Consumers

EU – DG Health and Consumers http://ec.europa.eu/dgs/health_consumer/index_en.htm		
METRICS		
LOGD usage	#de-referenced URIs / queries	(not available)
	#governmental reusers	~ 15-20
	#commercial reusers	0
LOGD revenue	Subscription fees	None
	On-demand fees	None
LOGD costs	development cost maintenance costs	~110k per annum.
	promotion costs	0
LOGD Benefits	# integrated datasets	Between 10 and 20
	#derived applications	2
ANALYSIS		
Value proposition	**Flexible data integration**: serving both external reusers and also intra-organisational needs. **Increase in data quality:** Linked Data helps DG Health & Consumers identify quality problems in legacy data and act on it. **New services:** 2 Linked Data apps have already been developed, the RDFa maker and the Forest Reproduction Material client application to support the management of forest reproductive material.	
Key resources	**Data:** Mostly social data and also scientific data, e.g. taxonomies of pesticides and food additives. **URI policy**: DG Health & Consumers is following the EC-wide URI policy developed by the Publications Office. **Linked data infrastructure**: Data dictionary, Data source container, SANCO-LD Hub, SPARQL endpoint. **Skills and competencies**: Skills are all in-house. Some help also from external contractors.	
Key partners	**Government:** • EU Member States, • DG Agriculture and Rural Development (AGRI), • DG Maritime Affairs and Fisheries (MARE), • DG Environment (ENV), • DG Communications, Networks, Content and Technology (CONNECT), • Publications Office, • Eurostat, • European Maritime Agency, • European Centre for Disease Prevention and Control (ECDC), • European Food Safety Agency (EFSA), • European and Mediterranean Plant Protection Organization (EPPO), • Community Plant Variety Office (CPVO). **Businesses.** **Non-governmental organisations:** • UN/CEFACT, • Open Archives Initiative.	
Key activities	**Development and maintenance.** DG Health & Consumers follows three stages in the provision of LOGD. First they take data that is already publicly available, model it and publish it in open data formats. Then, they make it available also as Linked Data and	

	finally they use it internally to develop tools and facilitate access to the data and encourage others to use it in the same way. Data cleansing and harmonisation activities are one of the biggest challenges and require significant time and effort. DG Health & Consumers does not invest in promotion as such, but tries to promote the LOGD to the extent possible at relevant meetings and events.
Cost structures	**Development and maintenance costs**: an annual investment of 110k for publishing and managing LOGD. **No promotion costs**.
Customer segments	DG Health & Consumers is reusing internally its LOGD. **Government and NGOs**: Other DGs with an interest in the data (e.g. AGRI, ENV, ESTAT, MARE), European agencies (ECDC, EMA, EFSA), other European institutions, national administrations, international organisations, and citizens. **Businesses**: industry and market operators. **Citizens.**
Revenue systems	**Sources of revenue**: Public funding. **Pricing model**: Free of charge. **Licence type**: Public domain, EUPL v1.1.
Channels	**Web API:** A public SPARQL Endpoint accessible at http://ec.europa.eu/semantic_webgate_acceptance/query/ **Bulk download:** All open data published by DG Health & Consumers is accessible via the EU Open Data Portal at http://open-data.europa.eu/en/data/publisher/sanco or can be downloaded from CIRCABC via the links at http://ec.europa.eu/dgs/health_consumer/information_systems/ **Apps:** 2 Linked Data apps have already been developed. More value added apps and services are expected.
Customer relationships	**Branding and advertisement**: No formal branding/advertising strategy. However, provenance and version information are available when accessing the data, through source and URIs **User support and feedback mechanisms:** No formal support/feedback mechanism/channel set up. DG Health & Consumers has regular informal communications via phone/email with people that are interested in reusing its LOGD. Stakeholders (Member States, internal users) are involved in testing the apps and tools that are currently in development and their feedback is iteratively taken into account. In the future, feedback tools will be integrated in the LOGD apps that will be developed by DG Health & Consumers. The data is accompanied by the EU Disclaimer.
FINDINGS	
Business model	**Public funding:** DG Health & Consumers provides Linked Data in accordance with its public task, which is to empower consumers, protect and improve public health, ensure Europe's food is safe and wholesome, protest the health and welfare of farm animals and protect the health of crops and forests. Funding of the activities is through on-going public funding. The data is provided free of charge under EUPL v1.1.
Enablers	**Efficiency gains in data integration** **Emerging best practice guidance** **Open licensing and free access**

Roadblocks	**Lack of necessary competencies** **Surfeit of standard vocabularies** **The inertia of the status quo**

3.3.5. EU – European Environment Agency (EEA)

EU – European Environment Agency (EEA) **http://semantic.eea.europa.eu/**		
METRICS		
LOGD usage	#de-referenced URIs / queries	(not available)
	#governmental reusers	(not available)
	#commercial reusers	(not available)
LOGD revenue	Subscription fees	None
	On-demand fees	None
LOGD costs	development cost	Part of on-going development activities
	maintenance costs	Not measured separately from other costs
	promotion costs	(not available)
LOGD Benefits	# integrated datasets	(not available)
	#derived applications	(not available)
ANALYSIS		
Value proposition	**Flexible data integration**: Linked Data provides a way to make data integration more efficient.	
Key resources	**URI policy**: EEA has a URI policy based on Cool URIs. The intention is for URIs (as well as URL for HTML pages) to be persistent. **Linked data infrastructure**: Data from Reportnet in XML is converted to RDF using schema definitions, XSL style sheets and lookup tables. **Skills and Competencies**: Skills are all in-house. Some training from external experts.	
Key partners	Eionet members, Eurostat.	
Key activities	All renovation of data flows or websites take Linked Data into account.	
Cost structures	Investment in software tools would have been made in any case. Maintenance is part of normal operations and not separately costed.	
Customer segments	EEA itself is the main user.	
Revenue systems	Public Funding	
Channels	Web API (SPARQL endpoint), bulk download.	
Customer relationships	Feedback from institutional collaborations.	
FINDINGS		
Business model	**Public funding**: EEA operations are publicly funded. Linked Data technology is used to make internal operations more efficient. Availability of Linked Data for external users is a side-effect. Data is made available free of charge, mostly under CC-BY licence.	
Enablers	**Efficiency gains in data integration**	
Roadblocks	**Lack of necessary competencies** **Perceived lack of tools**	

3.3.6. EU – Publications Office of the European Union

EU – Publications Office of the European Union http://publications.europa.eu/mdr/		
METRICS		
LOGD usage	#queries / hits	(not available)
	#governmental reusers	(not available)
	#commercial reusers	(not available)
LOGD revenue	Subscription fees	LOGD is made available free of charge
	On-demand fees	
LOGD costs	development cost	(not available)
	maintenance costs	(not available)
	promotion costs	None
LOGD Benefits	# integrated datasets	(not available)
	#derived applications	(not available)
ANALYSIS		
Value proposition	**Flexible data integration**: declarative approach has increased integration of new types of documents. **Increase in data quality**: control based on the ontology (Common Data Model) has increased data quality. **New services**: new infrastructure enables creation of new services in particular in the domain of reuse. **Cost reductions**: expected as soon as initial implementation will be finished.	
Key resources	**Data**: all metadata published by the Publications Office is at the moment of publication also available as Linked Data. **URI policy**: URIs of resources in CELLAR follow the pattern http://publications.europa.eu/resource/{ps-name}/{ps-id} where ps-name identifies the production system and ps-id is the unique identifier for the resource in the context of the production system. These URIs will be persistent. **Linked Data infrastructure**: RDF store based on dedicated ontology (Common Data Model). **Skills and competencies**: both available in-house and through external contractors.	
Key partners	**Professional reusers**: legal information services, information brokers, etc. **Public entities** of the EU member states.	
Key activities	Publication of official documents of the EU Institutions.	
Cost structures	Not available.	
Customer segments	**Professional reusers:** legal information services, information brokers, etc. **Public entities** of the EU member states.	
Revenue systems	**Sources of revenue**: Public funding, EU budget. **Pricing model**: Free of charge. **Licence type**: Reuse is authorised with acknowledge of source; reuse policy of the European Commission is implemented by Commission Decision of 12 December 2011 on the reuse of Commission documents (2011/833/EU).	

Channels	**Web API** **Website**
Customer relationships	**Branding**: None **Advertisement**: None **User feedback mechanisms**: Foreseen in the scope of the Publications Office's Common Portal (under development).
FINDINGS	
Business model	**Public funding**: All Linked Data activities are funded from the regular budget. Linked Data is made available free of charge.
Enablers	**Availability of data**: CELLAR, Open Data Portal
Roadblocks	None

3.3.7. IT – Agenzia per l'Italia Digitale

IT – Agenzia per l'Italia Digitale http://www.digitpa.gov.it/		
METRICS		
LOGD usage	#queries / hits	Statistics are unavailable due to a server error but it is known that traffic is increasing.
	#governmental reusers	Not known
	#commercial reusers	Tiggit Software, makers of Thunderbird PEC and Pocket PEC - which makes it easy to use legally prescribed communication methods.
LOGD revenue	Subscription fees	None
	On-demand fees	
LOGD costs	development cost	This is (currently) a two-person operation. There are no additional costs. All software used is free/open source. There is some promotion through networking and teaching.
	maintenance costs	
	promotion costs	
LOGD Benefits	# integrated datasets	Data supplied by Italian public administrations
	#derived applications	External App, visualisations
ANALYSIS		
Value proposition	**New services**: The provision of high quality reference data that can readily be linked and aggregated.	
Key resources	**URI policy:** URI design policy **Skills and competencies:** Two expert staff, open source software	
Key partners	**Government:** AgID is a government agency that receives its data from a commercial third party that provides a suite of services to the government	
Key activities	**Development and maintenance:** Gathering and triplifying the data (using D2RQ)	
Cost structures	Staff costs only	
Customer segments	Government, NGOs, citizens (perhaps via intermediaries)	
Revenue systems	Public funding	
Channels	Web API (SPARQL), bulk download, proprietary Apps (PocketPEC), Web apps (visualisation)	
Customer relationships	Limited contact with app developer. The data is authoritative and up to date which makes its provision valuable to the developers of the PocketPEC application.	
FINDINGS		
Business model	**Public Funding:** AgID's task is to establish a data infrastructure on which others can build services.	
Enablers	**Enthusiasm from 'champions'** AgID benefits from the enthusiasm of the LD community, especially around Geo Data	

Roadblocks	**The inertia of the status quo** Semantic interoperability is seen as something for specialists, not government employees.

3.3.8. UK – BBC

UK – BBC http://www.bbc.co.uk		
METRICS		
LOGD usage	#queries / hits	(not available)
	#governmental reusers	
	#commercial reusers	
LOGD revenue	Subscription fees	Not applicable. The service is provided for the benefit of other BBC departments
	On-demand fees	
LOGD costs	development cost	(not available)
	maintenance costs	
	promotion costs	
LOGD Benefits	# integrated datasets	Examples include data about every athlete, discipline and event at London 2012; likewise every player, match and venue in the 2010 World Cup. Increasingly there is data about individuals and events.
	#derived applications	The World Cup and London 2012 websites
ANALYSIS		
Value proposition	• **Flexible data integration**: The BBC is creator and custodian of information about people, events, industry, science, the arts and more. The connections between those data points are highly diverse so linked data is the only technology to offer a means of managing it and making sense of it. • **New services:** publishing data is an invitation for others to join the search for innovative ways of combining this data with other sources and present it in an engaging manner. • **Cost reduction:** improve the efficiency of the internal systems rather than to provide data as a service for others to use. • **Linked Data as a public good:** anything the BBC creates is owned by the licence payer – so it must be made available.	
Key resources	• **URI policy:** URI design policy • **Skills and competencies:** Initially the implementation of the LD solution was contracted. Now the important thing is to have good Java coders knowing about SPARQL and the mechanics of RDF. BBC has 4 data architects across the organisation. • **Linked data infrastructure:** Hosted in the cloud. • We have APIs on the platform for conflict resolution and de-duplication.	
Key partners	• Businesses: Data providers	
Key activities	News, Education and Entertainment	
Cost structures	Not available	
Customer segments	• **Internal customer:** other BBC departments.	

Revenue systems	• **Sources of revenue: public funding:** funded by a licence fee payable by all households in the UK that own a television. • **Pricing model:** Free of charge although some premium content is only available in the UK. • **Licence types**: non-commercial use only
Channels	• **Web API** (SPARQL)
Customer relationships	• **Advertisement:** Connected Studio event • **User feedback mechanisms:** informal
FINDINGS	
Business model	The work in making data available through the Programmes pages and Wildlife finder is not driven by a business case, it's more of an experiment. The primary goal of the Linked Data Platform is to make sense of all the BBC's creative works and provide an API to allow the retrieval of any creative work about any 'thing', with the added benefit that we hold a semantic graph of data behind the 'things'. The BBC has an unusually large amount of high value audio, video, images and text content spanning nearly 90 years. Managing that content, managing the information within it and making sure that it is available to people throughout the organisation is a huge task. The development of the linked data platform is a response to this demand.
Enablers	**Efficiency gains in data integration** **Increased linking and integrated services** **Ease of model updates** **Ease of navigation** **Cost reductions**
Roadblocks	**Missing, restrictive, or incompatible licences** **The inertia of the status quo** Licences and data quality (BBC has a reputation for quality, not all internal data quality is up to scratch). **Perceived lack of tools** At the height of the London 2012 Olympics the platform was handling 2,000 SPARQL queries per second. The BBC had to work hard to be able to support that in a resilient environment.

3.3.9. UK – Companies House

UK – Companies House http://companieshouse.gov.uk/		
METRICS		
LOGD usage	#de-referenced URIs / queries	average 720,000 hits/week
	#governmental reusers	The new Gazettes service from the National Archives will use the Companies House URIs
	#commercial reusers	Commercial users tend to use the XML gateway service
LOGD revenue	Subscription fees	The URI service is free
	On-demand fees	
LOGD costs	development cost	2 person months
	maintenance costs	The service is not actively promoted and effectively costs nothing to maintain as it is an add-on to the Web Check service
	promotion costs	
LOGD Benefits	# integrated datasets	The UK Company Register
	#derived applications	OpenCorporates, Gazettes
ANALYSIS		
Value proposition	• **Flexible data integration**: Content negotiation is used to return data in one of multiple formats when a URI is dereferenced but Companies House itself does not integrate any other data sources • **Cost reduction**: N/A. The URI service is an add on to an existing service	
Key resources	• **Data** (the company register) • **Explicit URI Policy:** based on UK government URI guidelines. Contains commitment to persistence • **Linked data infrastructure***:* The linked data service is simply a serialisation of that data as RDF	
Key partners	Some advice was given by 'leading figures in the open data arena' (Jeni Tennison etc.) but the work was done in house.	
Key activities	Since the URI service was established, no further work has been necessary	
Cost structures	• **Development costs***:* 2 person months • **Maintenance costs***:* Minimal additional effort for providing the URI service	
Customer segments	Not available	
Revenue systems	• **Pricing model: free of charge** for basic company details offered as linked data • **Licence: UK Open Government Licence**	
Channels	• None other than the URI resolution service	
Customer relationships	• User support • User feedback mechanisms	
FINDINGS		

Business model	**Public funding**: The provision of stable, de-referenceable URIs that return basic information about companies is clearly a useful building block on Britain's information infrastructure but, beyond incremental improvements, it is unlikely that any significant further development will take place. The service is easy to run and costs are *de minimus*.
Enablers	**Forward-looking strategies** The 'URI Service' at Companies House was an easy add-on to their existing services.
Roadblocks	**None**. Only a little extra help was required by the CH team to put the service in place.

3.3.10. UK – Department of Environment, Food and Rural Affairs (DEFRA)

UK – Department of Environment, Food and Rural Affairs (DEFRA) http://data.gov.uk/location		
METRICS		
LOGD usage	#queries / hits	Not in public domain
	#governmental reusers	Widgets used by at least 10 local authorities
	#commercial reusers	Marine Conservation Society (charity), Beach Selecta App, interactive screen on north Cornwall beach
LOGD revenue	Subscription fees	None
	On-demand fees	
LOGD costs	development cost	'Less than equivalent traditional processes and the cost is going down as we become more experienced.'
	maintenance costs	Not in public domain
	promotion costs	None
LOGD Benefits	# integrated datasets	There are about 5 integrated datasets that make up the Bathing Water Quality Data
	#derived applications	The BWQ Explorer, Widgets, Beach Selecta App
ANALYSIS		
Value proposition	**Flexible data integration**, **Increase in data quality, Cost reduction,** Better management and reporting. For example: the way the Bathing Water Quality data was modelled meant that the history was there. People could refer to it, talk about the data – this was very helpful for policy makers who were able to shift policy mid-season. That would have taken a year to implement under traditional systems but was completed in 3 weeks at a cost of less than £10K.	
Key resources	**Data:** The regular sampling data collected by the Environment Agency (part of DEFRA) **URI Policy:** URI design policy Linked Data Infrastructure: largely the Amazon Web Services cloud, managed by Epimorphics, TSO and SWIRRL **Skills and competencies:** Largely outsourced	
Key partners	Epimorphics, TSO, SWIRRL	
Key activities	Collection and dissemination of environmental data	
Cost structures	Unknown	
Customer segments	Policy Makers, citizens	
Revenue systems	N/A	
Channels	Web API (SPARQL Endpoint, data API)	
Customer relationships	N/A	
FINDINGS		
Business model	**Public Funding:** EU Member States are required to provide bathing water quality information. This is the UK's response to that. 'The ethos here is: Demonstrate benefit and then kill off other systems. There's a minimum 5 year commitment.'	

Enablers	**Ease of model updates** **Enthusiasm from 'champions'** **Cost reductions** Support from senior management, input from John Sheridan and others (National Archives), high quality consultancy
Roadblocks	**The inertia of the status quo** 'The IT dept often had a focus on systems rather than a data centred focus of delivery.'

3.3.11. UK – National Archives

UK – National Archives http://www.legislation.gov.uk		
METRICS		
LOGD usage	#de-referenced URIs / queries	5M/week 2M monthly unique visitors
	#governmental reusers	There are many users of the information but no identifiable users of the LOGD as data. Services like iLegal, MobileLegislate and Longman Law Bespoke use the information, not the linked data.
	#commercial reusers	
LOGD revenue	Subscription fees	none
	On-demand fees	none
LOGD costs	development cost	1.5 FTE in house. 10-12 FTE contractors
	maintenance costs	0.5M/year
	promotion costs	none
LOGD Benefits	# integrated datasets	3
	#derived applications	none
ANALYSIS		
Value proposition	• **Flexible data integration**: No other technology comes close to providing what the National Archives needs: being able to provide data about legislation at the level of the Act itself down to individual paragraphs. • **Increase in data quality**: The use of open standards has led to a 30% improvement in services. • **New services***:* Linked Data gave rise to new services on provenance and authenticity of the data; as audit trails around the legislative process. At the time of writing, new services for the official gazettes are being procured that will also make use of linked data and create, among other things, a ***de-facto insolvency register*** based on official notices.	
Key resources	• **Data**: legislation • **URI policy**: UK government URI guidelines • **Linked data infrastructure**: Legislation uses XML for docs, RDF for data and process. • **Skills and competencies**: Largely contracted.	
Key partners	• **Government**: Parliament, EU Publications Office • **Businesses**: technical contractor	
Key activities	• **Maintenance:**	
Cost structures	• **Maintenance costs**: Legislation 1.5 FTE in house. Contractors ca 10 – 12.	
Customer segments	• **Internal customer**: The most prevalent use is the internal reuse. • **Unknown customers**: There are users of the information provided on legislation.gov.uk but no known users of the data as linked data.	
Revenue systems	• **Source of income**: public funding (legislation) • **Pricing mechanism**: free of charge • **Licensing**: UK Open Government Licence	
Channels	• Not applicable	
Customer relationships	• Not applicable	
FINDINGS		

Business model	• **Public funding**: Provision of information about legislation is a core task of the National Archives • **Concession**: The new service providing information about the official gazettes is a revenue source since the publication of official notices is mandatory and chargeable.
Enablers	**Efficiency gains in data integration** **Forward-looking strategies** The Companies House URI service is important to TNA. **Increased linking and integrated services** The flexibility of the data - reusers have a lot of choices **Ease of navigation**
Roadblocks	**Lack of necessary competencies** We need to realistic about the level of proficiency of data reusers; **Perceived lack of tools**

3.3.12. UK – OpenCorporates

UK – OpenCorporates http://opencorporates.com/		
METRICS		
LOGD usage	#queries / hits	Not in public domain
	#governmental reusers	Unknown
	#commercial reusers	
LOGD revenue	Subscription fees	
	On-demand fees	
LOGD costs	development cost	Not public domain
	maintenance costs	
	promotion costs	
LOGD Benefits	# integrated datasets	Company Registers from more than 75 jurisdictions but this is not linked data (OC does not use Companies House URI service for example)
	#derived applications	None using LD
ANALYSIS		
Value proposition	**New services**: The substantial value proposition of OpenCorporates (a huge register of information about companies with insights into their structures) applies to the service as a whole, not the linked data it publishes, which is an add on (in the same way it is for Companies House)	
Key resources	**Data:** acquired in various ways from business registers around the world either directly or, sometimes, using scraper scripts. **URI policy:** design policy	
Key partners	None	
Key activities	**Maintenance:** Collating more data, investigating corporate networks	
Cost structures	Not in public domain	
Customer segments	OpenCorporates has a small number of commercial customers who would rather pay than adhere to the share alike terms under which data is available for free.	
Revenue systems	Donations and grants, consultancy, commercial access to data (see above)	
Channels	OpenCorporates website including data visualisations, bulk download.	
Customer relationships	OpenCorporates is at the centre of a growing community but is not customer-centric.	
FINDINGS		
Business model	Data for free, service for a fee. If a customer wishes to combine OC data with their own and not share their data, a fee is charged. Grant funding is also sought.	
Enablers	**Enthusiasm from 'champions'** The support of the community is substantial	
Roadblocks	**The inertia of the status quo** Access to data is a big problem for OC.	

3.3.13. UK – Ordnance Survey

UK – Ordnance Survey http://data.ordnancesurvey.co.uk/		
METRICS		
LOGD usage	#queries / hits	The trend is upward although there seem to be few repeat users.
	#governmental reusers	DCLG, Environment Agency, NHS appears to be making more & more use of it.
	#commercial reusers	Not known
LOGD revenue	Subscription fees	None
	On-demand fees	
LOGD costs	development cost	Part of the contract with the Cabinet Office (the PSMA[25]) - impossible to break out the linked data cost
	maintenance costs	
	promotion costs	
LOGD Benefits	# integrated datasets	1:50000 Gazetteer, UK Post Codes, Admin Geography for Great Britain
	#derived applications	DCLG applications, Bathing Water Quality Explorer etc.
ANALYSIS		
Value proposition	OS made some of their data available openly under pressure and a contract from the Cabinet Office and as linked data under direct pressure from Tim Berners-Lee and Nigel Shadbolt. Now it's done, however, they see advantages internally it's helping to break out of silos. If more people did the same the benefits of cross-linking would be more apparent too.	
Key resources	**Data:** The data sets held by OS **URI policy:** design policy **Linked Data Infrastructure:** initially outsourced, now run on Amazon Web Services and managed directly by OS **Skills and competencies:** Initially skills were bought in but have now moved in house.	
Key partners	National and Local government, Royal Mail	
Key activities	Provision of authoritative mapping data for other branches of government and the public.	
Cost structures	Linked data activities not separated out from other activities at OS.	
Customer segments	Government, commerce (developers etc.), public	
Revenue systems	Linked data is free although OS operates a number of premium data services that are chargeable.	
Channels	**Web API** (SPARQL Endpoint, Linked Data API for URI resolution) **bulk download**.	
Customer relationships	Significant activity in building the community through GeoVation. No formal promotional activity for linked data. Personal Twitter accounts and OS blog act as main dissemination channels.	

[25] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/7519/1665146.pdf

FINDINGS	
Business model	**Public funding:** OS is funded by its customers. Its freely available open data, including linked data, is funded through the Public Service Mapping Agreement (i.e. a contract with the Cabinet Office).
Enablers	**Efficiency gains in data integration** **Forward-looking strategies** It's much easier to manage our data ourselves and for others, such as Transport for London, to manage their data. The use of LD by them (and others) would be beneficial to all. **Increased linking and integrated services** e.g. DCLG's Open Data Communities **Enthusiasm from 'champions'** Such as Steve Peters & Leigh Dodds
Roadblocks	**Lack of necessary competencies** A lot of developers don't like it, seeing it as too academic and hard to learn. Can't take people through the learning curve in a single hack day.

3.3.14. UN – Food and Agriculture Organization of the United Nations (FAO)

Global – Food and Agriculture Organization of the United Nations (FAO) http://aims.fao.org/standards/agrovoc/about		
METRICS		
LOGD usage	#de-referenced URIs / queries	(not available)
	#governmental reusers	(not available)
	#commercial reusers	(not available)
LOGD revenue	Subscription fees	None
	On-demand fees	None
LOGD costs	development cost	Around 100,000 Euro for setting up the Linked Data infrastructure
	maintenance costs	Part of normal operations
	promotion costs	(not available)
LOGD Benefits	# integrated datasets	(not available)
	#derived applications	(not available)
ANALYSIS		
Value proposition	Linked Data offers the possibilities to create links across disparate collections. Linking between AGROVOC and other Knowledge Organisation Systems allows seeing connections that were not visible before.	
Key resources	**URI policy**: FAO has a URI policy in line with best practice. Main consideration is stability. **Linked Data Infrastructure**: OpenAGRIS is a triple store parallel to the main AGRIS database. AGROVOC is implemented in two triple stores; one for updating and one for external access. **Skills and Competencies**: Skills are all in-house. External experts contracted for knowledge transfer and training.	
Key partners	On data side: agricultural research institutes. On the technical side: MIMOS in Malaysia, and several European projects.	
Key activities	Using the mappings between vocabularies, the linked data is produced automatically as much as possible. Checking of vocabulary mappings is done by experts.	
Cost structures	Since 2002 the team has spent around one million Euro within semantic web projects, the final LOD infrastructure has had costs of about 100,000 Euro; maintenance in a team of 15 people but not separately earmarked.	
Customer segments	European institutions, universities, libraries, research institutes.	
Revenue systems	None.	
Channels	Web services, SPARQL endpoint	
Customer relationships	Promotion through presentations, Webinars and articles	
FINDINGS		

Business model	**Public funding**: FAO provides Linked Data under its general mandate to make all of its data available and provide access to data elsewhere for a worldwide community. The provision of Linked Data is one of the channels that are being offered. AGRIS data is made available free of change under CC-BY; AGROVOC is licensed under a CC-BY-NC-SA licence.
Enablers	**Efficiency gains in data integration** Data owners publishing Linked Data
Roadblocks	**The inertia of the status quo** Institutional policies and capacities.

4. SUMMARY OF FINDINGS

This section summarizes the main findings for the business models analysed in the 14 case studies, along the 9 areas of the business model canvas.

4.1. Value proposition

4.1.1. Flexible data integration

In several of the case studies, the main value proposition of data providers is to streamline their own internal processes for data integration, as well as data sharing amongst different departments and organisations with which working relationships already exist. Linking reference collections (e.g. subject vocabularies, locations, people and events) allows for the discovery of relationships between data in different collections. Several providers mention that there is no other technology that comes close to providing the possibilities that Linked Data offers in terms of flexible linking.

Problems mentioned include a lack of institutional capacities, a lack of mature tools and a paucity of suitable training material.

Relevant case studies: REEEP, EEA, FAO

4.1.2. Increase in data quality

Crowd-sourcing and self-service mechanisms are not much used by data providers at this time to improve the data quality. This is perhaps unsurprising given the use cases considered, all of which are in some way themselves the source of authoritative data against which other data may be judged. However, several providers mentioned anecdotally that data quality increased from the internal use of Linked Open Government Data and the use of open standards.

Relevant case studies: DG SANCO, Publications Office, UK National Archives

4.1.3. New services

The use of Linked Open Government Data for data integration by both providers and reusers allows them to improve the data they offer through service interfaces. For example, these might show related information about people, places, and subjects. Some new services are developed from the LOGD, in particular based on cross-referencing datasets. The surfacing of links that always existed but were not readily identifiable provides exciting possibilities for new service.

Relevant case studies: UK National Archives, DNB

4.1.4. Cost reduction

The notion of cost reduction assumes that a new technique is used to carry out an existing task but at lower cost. In that limited view, neither providers nor reusers of LOGD report a reduction in cost when developing new systems. However, the benefit is an increase of efficiency (doing more for the same money). Stark examples of this are the FAO who report a 50% increase in the use of data because of the enrichments that were possible using Linked Data, and the UK National Archives which points out the strong association between Linked Data and open standards. This leads to greater contestability and portability that not only saves money but, in their case, lead to a 30% improvement in services (this is

documented in the National Archives' response to the Cabinet Office consultation on Open Standards[26]).

Actual cost reductions show themselves in greatly reduced costs of future development, i.e. linking to further data sets, amending the data model etc. An inherent advantage of Linked Data over relational databases is the ease with which data models can be extended and adapted; several of the case studies mention the improvement in navigation across their data as a result of using Linked Data, and applications can make use of disparate data sets that are maintained and updated in real time by the relevant publisher without the need to download and process data dumps.

The investments being made vary from the establishment of highly resilient, production-grade systems to very low cost additions to existing infrastructure to make data available in the Linked Data format alongside others. At the same time, where Linked Data is used as an additional integration tool in an environment where data was already exchanged (e.g. using more traditional XML methods), there is a need to maintain several exchange technologies in parallel which increases cost, at least for an interim period.

When making comparisons between the costs of different Linked Open Government Data services it is important to be clear about what costs are included so that the comparison is genuine and useful.

Relevant case studies: FAO, UK National Archives, DEFRA, BBC

4.2. Key resources

4.2.1. Data

The area where Linked Open Government Data is applied most successfully at this point in time is in **reference data**, e.g. authority data for people, organisations, places, languages, formats and controlled vocabularies for concepts and subjects.

Relevant case studies: Europeana, DNB, EEA, FAO, DigitPA.

4.2.2. URI policy

Most data providers have a URI policy that follows best practice in terms of URI format and such best practice includes designing URIs for persistence. Therefore by following best practice, organisations are preparing for the long term. However, almost none of them give any formal guarantees of persistence. In reality, the only guarantee of persistence of the provision of a service is the persistent demand for it.

Relevant case studies: all

4.2.3. Linked Data infrastructure

Many Linked Open Government Data providers have developed their own tools as an add-on to their existing infrastructure. In some cases, a parallel database is provided for external access. Infrastructure is hosted by the organisation itself or hosted on services like Amazon Web Services. Open source solutions are commonplace but not universal. Tools and systems developed and cited by the LOD2 project are particularly common.

Almost all interviewees cited a need for better tooling and some highlighted the need for more resilient production-grade systems. None of the cases studied provided a service level agreement for external users.

[26] https://www.gov.uk/government/consultations/open-standards-open-opportunities-flexibility-and-efficiency-in-government-it

Relevant case studies: all

4.2.4. Skills and competences

It is notable that where Linked Data is used, there is often an individual champion who, through their own enthusiasm, is already knowledgeable about Linked Data. This enthusiasm is transmitted to colleagues but it's often necessary to use consultancy services or external service providers to build complete systems. In a short space of time, however, in-house skills increase either through training or new hiring so that before long, all necessary skills are in-house.

Many interviewees highlighted a need for better training materials. Although there are many existing training materials, the level of existing knowledge they assume is often too high so that even for an experienced developer or IT professional, the learning curve can appear too steep.

Relevant case studies: Europeana, AgID, DEFRA, OpenCorporates, Ordnance Survey/DCLG.

4.3. Key partners

4.3.1. Government

Most providers work together with their usual peers in existing networks.

Relevant case studies: FAO, EEA, Europeana, DNB.

4.3.2. Businesses and non-governmental organisations

This study focuses on Linked Open Government Data and so the relevant businesses are those that supply services to government. An exception is OpenCorporates which is a for-profit company. There is a growing number of Linked Data/Semantic Web consultancy services across Europe. Several of the case studies are of NGOs. Beyond the immediate scope of this study, experiments with Linked Data are known to be being carried out by sections of the retail industry.

Relevant case studies: BBC, UK National Archives, OpenCorporates, DG SANCO.

4.4. Key activities

4.4.1. Development

To provide Linked Data, some development has to be done, e.g. to generate mappings from existing non-RDF data, and to add links to external sources. Some providers see this as a normal activity to upgrade their systems. Some of this also requires human intervention (finding the related resources, data cleansing).

Relevant case studies: DNB, FAO, Europeana.

4.4.2. Maintenance

Many providers see the maintenance of LOGD as a normal part of the operation, especially if they use Linked Data as the internal data approach.

Relevant case studies: REEEP, Europeana, EEA, Publications Office, Ordnance Survey.

4.4.3. Promotion

Not much is done in general in terms of promotion - Linked Data is like the plumbing of a house. Promotion is done on the level of the resulting service, not on the inner workings.

Relevant case studies: all.

4.5. Cost structures

4.5.1. Development cost

There was a very wide variety in development cost indicated by the interviewees. This ranged from practically zero cost to amounts in the order of a million Euro, while in some cases the development of Linked Data services was considered to be part of normal systems evolution and costs were not separately available. Differences in cost may also be related to the development approach; for example if the development involves complete redesign of the system the cost will be higher, while providing an add-on service based on conversion of existing data may be cheaper. The wide variety makes it impossible to derive specific conclusions.

Relevant case studies: all**.**

4.5.2. Maintenance cost

This is often included in normal maintenance. In some cases there is mention of a small number of FTEs (2-5) involved in the maintenance, with exception of UK National Archives who report between 10 and 15 people. Like in section 4.5.1 on development cost, the cost for maintenance is highly dependent on the approach taken.

Relevant case studies: all.

4.5.3. Promotion cost

Most providers do not spend funds on promotion of Linked Data, other than presentations at conferences. Some do, however, organise hack days and similar events.

Relevant case studies: Reegle and Ordnance Survey.

4.6. Customer segments

4.6.1. Government and NGOs

Most Linked Data is used internally or used by other public sector organisations and NGOs.

Relevant case studies: REEEP, FAO, EEA, BBC, DNB, Europeana, DG SANCO, Publications Office, UK National Archives.

4.6.2. Businesses

There is as yet no substantial usage by commercial reusers as reported in the cases investigated.

Relevant case studies: Publications Office, OpenCorporates, DG SANCO.

4.6.3. Academia

Some providers reported usage of their Linked Data by universities and research institutes.

Relevant case studies: FAO.

4.7. Revenue streams

4.7.1. Sources of revenue

In all of the case studies, the provision of Linked Data is financed through public funding. Many of the providers that participated in the study provide LOGD as part of their public role and do not charge for their data. Our finding cannot be generalised towards a conclusion that there are no other possible funding models for Linked Data; it is just that the organisations participating in the study do not seem to need other sources of revenue at this time.

We have found no evidence that existing or future consumers of government data would be willing to pay for Linked Open Government Data as a Service (LDaaS). Indeed, in the majority of the case studies, the *external* demand for LOGD is virtually non-existent. In all case studies, the supply of LOGD is currently financed through public funding. European Directives on Public Sector Information also put further restrictions on the pricing of Government Data.

Future outlook: LOGD providers often field requests for bulk downloads of their data, which most of them offer but don't promote. This suggests that the data is seen as useful by at least some third parties who wish to integrate the data but within their own internal network, not as part of an open ecosystem. The increasing supply of LOGD might reach a tipping point where the use of LDaaS by external parties takes off and generates positive network effects. Such a development would prompt the development of standards for access control that are currently absent from the Linked Data technology stack.

Relevant case studies: all.

4.7.2. Pricing model

All Linked Open Government Data providers offer their data for free. Some providers indicated that they may consider a freemium model with free basic information and payment for more detailed information.

Relevant case studies: Europeana, Companies House.

4.7.3. Price structure

As currently all LOGD is provided for free, there are no price structures.

Relevant case studies: all.

4.7.4. Licence types

Typical licence types are either CC0, UK Open Government Licence, with some services restricting further use with CC-BY-NC licences.

Relevant case studies: DNB, Europeana, FAO.

4.8. Channels

4.8.1. Web API

Many services offer access through direct URI resolution. SPARQL endpoints are also offered.

Relevant case studies: REEEP, DG SANCO, EEA, Publications Office, DigitPA, BBC, DEFRA, Ordnance Survey.

4.8.2. Bulk download

Bulk downloads are available from several providers. However, in some cases the download is generated from the non-RDF database at irregular intervals, which means that the data is not as recent as the live database. Where offered, the bulk download option is rarely promoted as the provider prefer people to use the live APIs.

Relevant case studies: DNB, DG Health, EEA, DigitPA, OpenCorporates, Ordnance Survey.

4.8.3. Proprietary app

Just two proprietary apps have been identified in the course of this study (PocketPEC and Beach Selecta), both of which are available for free.

Relevant case studies: DigitPA, DEFRA.

4.8.4. Web app

Many services use the LOGD in a portal site where the data is integrated in the search service. The Bathing Water Quality Explorer applications created under the Open Data Communities umbrella are notable Web applications that use LOGD.

Relevant case studies: DEFRA

4.9. Customer relationships

4.9.1. Branding

There is little branding of Linked Open Government Data services, as this is mostly seen as plumbing behind the user services that are provided.

Relevant case studies: all.

4.9.2. Advertisement

Providers do not advertise the availability of Linked Data in a structured way. Many indicate that they do make others aware of their provision use of Linked Data in presentations at conferences and other events. Blogs and tweets are a primary dissemination channel.

Relevant case studies: all.

4.9.3. User support

There is little user support. Many Linked Open Government Data providers do not know who their external users are.

Relevant case studies: all.

4.9.4. User feedback mechanisms

Very few providers offer feedback mechanisms. If they do, feedback is typically through informal communications as part of institutional collaborations, comments on blogs, replies to Tweets etc.

Relevant case studies: all.

5. SUMMARY OF ENABLERS AND ROADBLOCKS

From the case studies carried out in this study, a number of enablers and roadblocks to the provision and reuse of Linked Open Government Data can be identified.

5.1. Enablers

5.1.1. Efficiency gains in data integration – the network effect

In a large number of the case studies, public sector organisations see the benefits of implementing Linked Data-based solutions to support their internal data integration activities, or to facilitate data exchange within already existing collaboration structures. Each organisation is responsible only for maintaining its own data and does not need to download and integrate others' in order to add context and meaning to it. This cooperative methodology often renders the distinction between data publisher and data user largely meaningless. Rather there is a network effect[27] so that each individual organisation's data is more valuable by virtue of the ease of connection with the others'.

That said, a side effect of the approach is that data is available to external third parties who purely consume data or who use it to enrich their own data that is not public. For example, Fujitsu uses Linked Open Government Data to enrich its (sensitive) health sensor data[28].

This is a crucial part of the value proposition for Linked Open Government Data. It eases both data integration and management, and benefits from the inherent responsiveness of the modelling as more data comes on stream. The efficiency gains often enable organisations to meet their public task in new and better ways than was previously possible. The case studies in this report provide proof that the value proposition has been realised in many instances and nothing enables success quite like previous success.

Future outlook: Many of the organisations in the study express the expectation that the use of Linked Data will increase as the benefits in increased efficiency and flexibility are and become more obvious.

Relevant case studies: REEEP, DNB, DG SANCO, EEA, BBC, TNA, OS, DCLG, FAO.

5.1.2. Forward-looking strategies

Some of the providers of LOGD engage in these activities from a background of their mandate to disseminate their information as widely as possible. They may already offer data in a range of formats, for example human-readable webpages, XML files or other domain-specific data formats like MARC21 for libraries, and consider the provision of data as Linked Data as part of a forward-looking strategy. Where these providers are already engaged in collaborative structures and data sharing arrangements, Linked Data is seen to offer more efficient mechanisms to create better integration across collections.

Future outlook: As the benefits that organisations that have such forward-looking strategies (the early adopters) become visible, other organisations will follow their lead.

Relevant case studies: Europeana, Companies House, TNA, OS.

5.1.3. Increased linking and integrated services

Reusers often see the value of Linked Data in the ease with which links to one or more external sources can be provided as part of the user interface. For example, a user interface for data concerned with statistical information about different UK locations can easily link to

[27] http://en.wikipedia.org/wiki/Network_effect

[28] http://www.w3.org/2013/04/odw/odw13_submission_5.pdf

other information about the locations provided by the Ordnance Survey. While Linked Data is not strictly necessary to build such functions (e.g. some reusers also bring up information from Wikipedia), Linked Data offers a more standard way to access information from different sources without the need to build special software tailored to specific providers.

A prime example of this comes from the UK National Archives which, at the time of writing, is about to launch a new online service for the official gazettes. The London[29], Edinburgh and Belfast Gazettes act as the public record of a variety of announcements including notices of insolvency. All notices will be encoded in HTML5 with the data embedded as RDFa. By collating this data and making it available via a standard interface, TNA is creating a *de facto* insolvency register, one that is able to unambiguously identify the relevant business using the Companies House URI service.

No specialist software or system is required, indeed, there is no insolvency register to build. It emerges as an artifact of the Linked Open Government Data ecosystem.

Future outlook: Availability of more Linked Open Government Data, especially under open licences, will enable more enhanced and integrated services to become available.

Relevant Case studies: REEEP, BBC, TNA, DCLG (Ordnance Survey reuser).

5.1.4. Ease of model updates

Information systems based on traditional relational databases are designed with a specific data model in mind. Tables of data have a defined structure and are linked via primary keys. This architecture offers a number of benefits but it is often very difficult and costly to make changes to the data model and to add new data. 'Adding a new column' to the database is not a trivial task.

A big enabler offered by the Linked Data approach is that changes to the data model are almost trivial and do not entail any change in the existing architecture. As the name suggests, a triple store always has three columns[30]. No matter how complex the data model in use, the same software, based on the same set of open standards, can still be used.

Future outlook: Like any technical product, the commercial software that implements Linked Data standards is constantly improving. The standards are evolving too with the W3C's maturing 'Linked Data Platform' standard being particularly relevant in this context[31].

Relevant Case studies: BBC, DEFRA.

5.1.5. Ease of navigation

A mantra in the Linked data world is 'follow your nose,' meaning that when a piece of data includes a URI it should be looked up and this may in turn reveal further information. This path-like aspect of the technology can be taken all the way through to the user interface to provide better navigation through even complex data. Both the BBC and DNB – public service providers of rich and diverse information, highlight this aspect.

Future outlook: Developments within the media, cultural heritage and retail industry[32] are likely to complement government Linked Data initiatives and improve users' ability to navigate an increasingly complex information space.

Relevant Case studies: Reegle, BBC, TNA, DNB.

---

[29] http://www.london-gazette.co.uk/

[30] In fairness, it's usually 4, to support the concept of Named Graphs, but the essential point holds.

[31] http://www.w3.org/TR/ldp/

[32] http://www.w3.org/2013/04/odw/GS1-LinkedDataPresentation-ODI-April2013.pdf

5.1.6. Open licensing and free access

Where the access to and reuse of data is made explicit, open licences are generally used, including Creative Commons Zero Public Domain Dedication, Creative Commons Attribution and the UK Open Government[33] licences. In general, the data is provided free of charge.

Future outlook: There is a general tendency towards removal of restrictions to the reuse of public sector information. These approaches will enable future, wider usage of those data, also as LOGD.

Relevant case studies: Europeana, DG SANCO.

5.1.7. Enthusiasm from 'champions'

Many cases show that an important role is played by the knowledge and enthusiasm of an individual. Such individuals are often part of the enthusiastic community that exists around Linked Data. The crucial role they play is to create awareness of the possibilities and the potential benefits. Progress is especially rapid where that individual is close to the decision processes in the organisation.

Future outlook: The role of 'champions' and other thought leaders will continue to be important in raising awareness and driving organisational change.

Relevant case studies: Europeana, AgID, DEFRA, OpenCorporates, OS, DCLG.

5.1.8. Emerging best practice guidance

Other enablers include wide adoption of best practices such as developed by W3C[34] and experts in the field[35]. Following guidance on URI design [PURI] ensures that data will persist beyond the life time of a project, a technology or the institution that created it. These and other emerging best practice guidelines make it easier for both providers and reusers to apply common techniques and tools.

Future outlook: More activities in developing best practice guidelines are expected, for example in the proposed Data on the Web Best Practices Working Group[36] at W3C as well as in domain-specific activities.

Relevant case studies: Europeana, DG SANCO, National Archives, BBC.

5.2. Roadblocks

5.2.1. Necessary investments

As with all new technologies, the provision and reuse of Linked Data requires organisations to make investments in infrastructure and software. The level of investment varies enormously across the case studies. Where the supply of Linked Open Government Data is merely an additional format added to an existing service, such as that offered by Companies House, the investment is no more than nominal. Running the URI service is such a small addition to the existing costs that it can't be measured and would be outweighed by the cost of administering any charging scheme. Entirely new or replacement services obviously entail significantly more investment. The FAO report a very substantial investment, for example,

[33] http://nationalarchives.gov.uk/doc/open-government-licence/version/2/

[34] W3C. Best Practices for Publishing Linked Data. W3C Note. 06 June 2013. Editor's draft. https://dvcs.w3.org/hg/gld/raw-file/default/bp/index.html

[35] http://linkeddatabook.com/

[36] http://www.w3.org/2013/05/odbp-charter.html

but this must be compared with the costs of establishing a similar service using alternative technologies. It is all too easy to compare apples with pears in this context.

It is noteworthy that DEFRA highlighted the short time span and low cost of amending their data model (3 weeks, £10,000) compared with what, anecdotally, it would have cost under a traditional relational database system (1 year, 10 times more). Few LOGD providers were able to accurately state the additional investment costs required to provide Linked Data as a Service.

Future outlook: As organisations increasingly start seeing the benefits of Linked Data for their internal operations, investments in the new technology will be considered part of the normal development budgets, rather than as additional expenditure, especially when considering long term flexibility.

5.2.2. Lack of necessary competencies

As the cases studied necessarily involve organisations that are already involved in Linked Data activities and have made the necessary investments in technical infrastructure and staff competencies, several of the people interviewed mention that other organisations, either data providers or potential reusers, are currently less advanced both technically as well as in terms of human competencies. As a result, the take-up of Linked Open Government Data is not as rapid as it could be.

While online tutorials are available, these may assume a level of knowledge in some areas that is not common. People who understand Linked Data find it simple; those that don't sometimes find it impenetrable.

Future outlook: More activities in developing training materials are expected, particularly materials tailored to a particular community, such as the training materials[37] developed by the Open Data Support project[38] for publication of datasets by governments in Europe.

Relevant case studies: DG SANCO, EEA, TNA, OS.

5.2.3. Perceived lack of tools

Many of the providers and reusers of Linked Open Government Data develop their own tools. While these are appropriate for the use in a particular context, in a more general sense, the necessary resilient, production-grade tools that large organisations can use to deliver their services in a way that enhances their reputation are perceived to be not readily available.

There is a lack of suitable tools, or at least a perceived lack, for creating, manipulating and converting Linked Data, particularly in terms of automated procedures. This is perhaps surprising given that Oracle, IBM and YarcData (part of Cray) are already among the companies offering high specification Linked Data systems. The performance of Linked Data servers is a long way short of relational databases which are now highly optimised, resilient, production-grade systems. The latter should not be a barrier either, as Linked Data services can run on top relational database environments.

Future outlook: As more large organisations are implementing Linked Data approaches, the market for Linked Data-based tools will grow, which will encourage both Open Source communities as well as commercial vendors to start offering a wider range of production-grade tools.

Relevant case studies: EEA, BBC, TNA, DG SANCO.

[37] http://www.slideshare.net/OpenDataSupport

[38] http://www.opendatasupport.eu/

5.2.4. Lack of service level guarantees

The reuse of LOGD services by external third parties is hindered as providers do not yet give explicit service level guarantees (SLGs). The case studies show that this is largely because the use of Linked Data is first and foremost for the publisher's own benefit and the availability of the data for third parties is a side effect. SLGs are therefore lacking for, among other things:

- the availability of the Linked Data service;
- the long-term persistence of the URIs;
- the integrity of the resolved data;
- the available formats and quality of the data served at each URI;
- the latter point applies particularly where data served by one organisation depends on another.

Service Level Agreements, i.e. agreements between named parties cf. guarantees that are made by the provider only, do exist though in cases where the provision of the infrastructure is outsourced, as in the cases of REEEP and the National Archives.

Future outlook: Increased use will lead to increased reliance on the Linked Open Government Data that is provided. Providers will get more and immediate complaints if data is not available or if serious errors are identified. This will force providers to make sure that data is available and of high quality in line with their reputation. Governmental organisations are well placed to guarantee long-term persistence of the URIs.

Relevant case studies: REEEP, BBC, DG SANCO.

5.2.5. Missing, restrictive, or incompatible licences

Many organisations that offer Linked Open Government Data do not make explicit information available about the licence under which the data can be reused. In such cases, the legal default position is that a potential reuser will need to contact the data provider to know what can be done with the data. This seriously hinders the wide usage of such data.

In addition, if organisations opt for restrictive licences, it makes the reuse more difficult, for example if a reuser wants to merge data from various sources. In such cases, complex provenance relations may need to be maintained to comply with various restrictions on the data.

Future outlook: Increasing awareness of the importance of clear licences is already visible. In combination with an increased focus on open access to public sector information, a more coherent licensing landscape can be expected in the years to come. Clearly stated but incompatible licenses are as much of a roadblock to mixing data as an absence of licence information so licences with as few restrictions as possible are needed.

Relevant case studies: REEEP, BBC.

5.2.6. Surfeit of standard vocabularies

One characteristic that is often considered a feature of Linked Data, namely that anyone can choose their favourite vocabulary to describe data, can also be a problem that requires additional guidance. While it is obvious to many implementers that it is good practice to use Dublin Core to provide the title, author, description and publication date of any published work there is no obligation to do this. The development of vocabularies such as the Data Cube vocabulary, the Organisation Ontology, the ISA Programme Core Vocabularies and schema.org are all highly useful but initially confusing to many. On top of this there are several different application profiles based on these standard vocabularies for similar applications. Although Linked Open Government Data is, by its very nature, easy to integrate, when different institutions describe the same things in different ways, creating that Linked Data is still a challenge.

Future outlook: Further development of domain-specific application profiles, such as the DCAT application profile for data portals in Europe[39] or the Europeana Data Model[40] can be expected to underpin data integration in collaborative networks.

Relevant case studies: DNB, DG SANCO.

5.2.7. The inertia of the status quo

Workflows and practices in the public sector have evolved over a long period of time and, in general, change is only accomplished slowly. This is the way we do it because this is the way it's always been done. Sharing data openly creates understandable fears of exposure and scrutiny. One artefact of this is the quality of data used. In a closed world, incomplete or inaccurate data may be seen as the norm within that environment. Making the data open exposes those errors to people who may not take the full context into account.

Future outlook: "The status quo can best be overcome through dedicated reforms through successful LOGD programmes coupled with careful change management to meet the scepticism and inertia that can otherwise slow the adoption of new policies.

Relevant case studies: REEEP, DG SANCO, AgID, BBC, DEFRA, OpenCorporates, FAO.

[39] https://joinup.ec.europa.eu/asset/dcat_application_profile/description

[40] http://pro.europeana.eu/edm-documentation

6. CONCLUSION

This study has identified a total of 37 cases where LOGD is already provided by public administrations. This demonstrates that LOGD is moving forward in the Technology Adoption Lifecycle. In the Library domain, the adoption may have reached the Early Adopter phase given that the large players (National Libraries, and the main library data host OCLC) are the main providers of Linked Data. LOGD is becoming particularly important in the provision and management of reference data (information about people, organisations, places, controlled subject vocabularies).

Provision of LOGD to external reusers is in almost all cases not the first objective of the organisations that create Linked Data. More often, it is used to increase efficiency of internal data integration, or to support data exchange in existing collaborations. Many providers do not have a clear view of the consumers of their data: in general, they do not monitor usage and do not offer feedback mechanisms, nor do they give guarantees to external parties about the availability or the quality of their services.

In the public sector, the provision of Linked Data is essentially seen as part of the public task, and therefore the prevalent business model is one where investments and maintenance are funded from the on-going public funding with some help from occasional grants, and where the data is made available free of charge.

In the study, we have not seen a wide reuse by third parties that take data from various providers and create new services from such mash-ups. This may be because providers do not yet provide operational guarantees, because Linked Data requires acquisition of new skills on the part of the reusers, or because there is a perceived lack of tools that can be used to produce and consume Linked Data. As such, reuse by third parties is still very much in the Innovator phase with few examples of new services.

However, as main data providers are in the process of producing massive amounts of Linked Open Government Data, and given the opportunities for more efficient data integration that Linked Data technology offers, it can be foreseen that in the near future more reusers will find their way to the data and that this will provide a fertile environment for innovation.

7. FURTHER WORK

As the title makes clear, this study set out to discover alternative business models for Linked Open Government Data - but found only one in operation. Furthermore it proved difficult or impossible to collect many of the metrics that were identified as being important (see page 73). In particular it was not possible to quantify any change in the corrections requested for data or for the cost of data integration. There are a number of reasons for this:

1. Where the provision of LOGD sits on top of an existing service (that remains unchanged) the cost increases although by a marginal amount (e.g. Companies House). The usage statistics of these additional services are not called out in the general monitoring of Web site access.
2. Where Linked Data is used to replace an existing service, the cost of the change in technology needs to be compared with what it would have cost to use a different approach to go from the same starting point to the same end point. Each specific case will differ according to the initial starting conditions and so meaningful comparisons are all-but impossible to draw.
3. Where LOGD is used, it is primarily for the benefit of the organisation that creates and manages the data, perhaps in consultation with its partners, and not for third parties. Therefore no effort is made to monitor external usage, which is seen as an added bonus, not a core function.

Taking these factors into account, a more detailed and quantitative breakdown of the costs and benefits of LOGD can only be ascertained by comparing closely related case studies. For example: the French equivalent of the British Bathing Water Quality Explorer (page 152), provides a similar service but does not use LOGD to achieve it. A comparative study of this and other examples where the starting conditions and eventual service are closely aligned, but where one uses LOGD and another doesn't, would be illuminating.

To assess the value of LOGD for internal users, it will be necessary to find cases where similar or identical data is integrated for similar or identical purposes, again, with one system using LOGD and another using a different technology. The Open Data Communities work could be the basis of such an investigation (see page 184).

ACKNOWLEDGEMENTS

Specific acknowledgement is due to:

Person	Organisation
Vassilios Peristeras	European Commission, Directorate-General for Informatics
Florian Bauer	Renewable Energy and Energy Efficiency Partnership
Lars G. Svensson	German National Library
Stefan Rohde-Enslin	Museum-digital
Ruth Ní Bhraonáin	European Commission, Directorate-General for Health & Consumers
Giorgos Georgiannakis	European Commission, Directorate-General for Health & Consumers
Jill Cousins	Europeana Foundation
Antoine Isaac	Europeana Foundation
Stefan Jensen	European Environment Agency
Søren Roug	European Environment Agency
Giorgia Lodi	Agenzia per l'Italia Digitale
Antonio Maccioni	Agenzia per l'Italia Digitale
Dave Rogers	British Broadcasting Corporation (BBC)
Oli Bartlett	British Broadcasting Corporation (BBC)
Mark Fairhurst	Companies House
Chris Smith	Companies House
Stacey Smith	Companies House
Alex Coley	Department of Environment, Food and Rural Affairs (DEFRA) - UK
Ewan Peters	Arup
John Sheridan	National Archives - UK
Chris Taggart	OpenCorporates
John Goodwin	Ordnance Survey Research
Steve Peters	Department for Communities and Local Government - UK
Johannes Keizer	Food and Agricultural Organisation of the United Nations
Peter Schmitz	Publications Office of the European Union
Marc Küster	Publications Office of the European Union
Magnus Kolsjö	PwC Sweden

REFERENCES

[2003/98/EC] Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the reuse of public sector information http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003L0098:en:NOT

[ADMS.SW] ADMS.SW Metadata Vocabulary and Software Taxonomies – Licence types. https://joinup.ec.europa.eu/svn/adms_foss/amds_sw_taxonomies/ADMS_SW_v1-00_Taxonomies.html#

[Buchanan] Buchanan, James; Wm. Craig Stubblebine (November 1962). "Externality". Economica 29 (116): 371–384.

[Cobden] Marcus Cobden, Jennifer Black, Nicholas Gibbins, Les Carr, Nigel R. Shadbolt: A Research Agenda for Linked Closed Dataset. http://eprints.soton.ac.uk/272711/

[Ferro] Enrico Ferro, Michele Osella: Business Models for PSI Reuse: A Multidimensional Framework. http://www.w3.org/2012/06/pmod/pmod2012_submission_16.pdf

[Ding2012] Li Ding, Vassilios Peristeras, Michael Hausenblas: Linked Open Government Data [Guest editors' introduction]. IEEE Intelligent Systems 27(3): 11-15 (2012) Roadmap of linked open government data from our editorial in IEEE IS [1]. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6237454

[CoolURIs] Cool URIs for the Semantic Web, 2006 http://www.w3.org/TR/cooluris/

[Shadbolt] Nigel Shadbolt, Kieron O'Hara, Tim Berners-Lee, Nicholas Gibbins, Hugh Glaser, Wendy Hall, m. c. schraefel: Linked Open Government Data: Lessons from Data.gov.uk. IEEE Intelligent Systems 27(3): 16-24 (2012) http://eprints.soton.ac.uk/340564/

[MEPSIR] Measuring European Public Sector Information Resources http://ec.europa.eu/information_society/policy/psi/mepsir/index_en.htm

[Norway-LOD] Dumitru Roman, David Norheim: An Overview of Norwegian Linked Open Data. http://www.semicolon.no/wp-content/uploads/2013/02/Norwegian-LOD.pdf

[CompaniesHouse-URIs] Uniform Resource Identifiers (URI) Customer Guide http://www.companieshouse.gov.uk/about/pdf/uniformResourceIdentifiersCustomerGuide.pdf

[Vafopoulos-LD] A framework for Linked Data business models

[Dodds] Thoughts on Linked Data Business Models, http://blog.ldodds.com/2010/01/10/thoughts-on-linked-data-business-models/

[ODI]	How to make a business case for open data? http://www.theodi.org/guide/how-make-business-case-open-data
[Osterwalder-Book]	Osterwalder, A., & Pigneur, Y. (2009). Business Model Generation.
[Osterwalder-PhD]	Osterwalder, A., "The Business Model Ontology – A proposition in a design science approach", PhD thesis, Université de Lausanne, 2004. http://www.hec.unil.ch/aosterwa/phd/osterwalder_phd_bm_ontology.pdf
[Pollock]	Pollock, Rufus, (2009), The Economics of Public Sector Information, Cambridge Working Papers in Economics, Faculty of Economics, University of Cambridge, http://EconPapers.repec.org/RePEc:cam:camdae:0920.
[PURI]	Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris
[POPSIS]	POPSIS - Pricing Of Public Sector Information Study http://epsiplatform.eu/content/popsis-assessment-psi-charging-policies
[TimBernersLee]	Linked Data Design Issues. 2006. http://www.w3.org/DesignIssues/LinkedData.html

Annex I. GUIDE TO CONDUCTING THE CASE STUDIES

This annex outlines the approach to conducting the case studies. It guarantees that each case study is carried out in a consistent way.

I.1 Conduct preparatory desk research

For each case study, at least the following information must be collected via desk research and included in Annex II:

- **Mission of the LOGD supplier / reuser**: a summary of a charter explaining in simple words the public task and mission of the LOGD supplier / reuser.
- **Supply and reuse of linked data**: a summary of the provided linked data, including links, documentation, used vocabularies, and description of linked data services.
- **Pricing of LOGD**: a summary of the applied pricing and documentation;
- **Licensing of LOGD**: a summary of the licensing conditions; and
- **Market analysis:** a summary of relevant market players, suppliers and users of LOGD.

The sources of information can be:

- Website of public administrations;
- Annual reports; and
- Previous PSI studies.

I.2 Collect metrics

For each case study, the following metrics must be collected – where available.

Metric
Usage • Number of de-referenced URIs / queries; • Number of governmental reusers of LOGD; and • Number of commercial reusers of LOGD.
Revenue and other sources of income • Public funding; • LOGD revenue in subscription fees / on demand fees per year ; • LOGD price evolution (subscription / on demand fees) per year; • Advertisement.
Cost • **Development cost**: the cost of all activities that were required to identify, model, transform, harmonise, publish and / or reuse LOGD; • **Maintenance cost per year**: the cost of publishing updates of the data, maintenance costs of relevant infrastructure; and • **Promotion costs per year**: the cost of promoting the availability of the data as linked data.
Benefits • Number of linked datasets (outgoing and incoming links); • Number of derived applications; • Number (and increase) of corrections requested; and • Cost reductions on information integration.

I.3 Conduct the interviews

For each case study at least two interviews must be conducted:

- Interview 1: LOGD provider; and
- Interview 2: LOGD reuser.

In both cases, the interviewees must play a coordinating role, for example have the profile of a manager or director. Summaries of interviews must be sent for validation to interviewees afterwards.

CC: project officer

Subject: Invitation to participate in a study on Linked Open Government Data

Dear $Name,

We work as contractors of the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission (DG Informatics). We have been commissioned to undertake a study to identify viable business models for Linked Open Government Data (LOGD). The responsible project officer for this study is Dr. Vassilios Peristeras <Vassilios.Peristeras@ec.europa.eu>.

To this end, we will undertake in-depth case studies looking at the costs, value proposition and revenues related to LOGD. For each case study, we intend to collect information from both public administrations that supply LOGD and its reusers.

We believe that your organisation's involvement in LOGD can provide a number of very relevant findings to our study and we would like to learn more about it.

$Further specify why, demonstrating an initial understanding of the candidate interviewee.

We would expect this interview to take about one hour. We will call you to schedule a meeting at a moment of your convenience.

You may also find attached the metrics on LOGD we wish to collect and the interview questions that we will ask you. What is in it for you? We believe that our study offers you the following:

- visibility: the best case studies will be presented at the SEMIC 2014 conference.

- insight and benchmarking of LOGD efforts

- learn from other organisations that participate in our study

We look forward to hearing from you.

Best regards,

I.4 Interview questions LOGD provider / consumer

The below interview questions apply both to LOGD providers and consumers. The interviewer must share these questions with the interviews before the interview. The interviewer will use the 'LOGD analysis framework' listed in Chapter 2 to give further structure to the interview. A written summary of each interview must be sent for validation to the interviewee. The validated interview summary must be included in appendix.

Organisation

1. How does the supply / reuse of LOGD relate to the public task of your organisation?
2. Does your organisation have a "business case" document that motivates your investment in supplying / re-using LOGD? Can you share it with us?
3. What are your future plans? Do you plan to expand or abolish the supply / reuse of LOGD?

Value proposition

4. Did the supply / reuse of LOGD give rise to new opportunities for flexible data integration? Increase in data quality? New services? Cost reductions?
5. What are the main enablers / inhibitors for LOGD to deliver value for its reusers?
6. Do you supply / require service level statements for LOGD?

Key resources

7. Does your organisation have a URI policy? Can we have a copy?
8. Can you describe the Linked Data infrastructure of your organisation?
9. Which skills and competencies did you need in order to supply / use LOGD? Do you have them in-house or contracted?

Key partners

10. Which organisations are key partners in the supply / reuse of LOGD?

Key activities

11. Which activities do you carry out to supply / reuse LOGD?

Cost structure

12. Which investments has your organisation made to enable it to supply / reuse LOGD?
13. How many FTEs in your organisation are involved in provisioning the supply / reuse of LOGD?
14. Which costs have you incurred to publish LOGD, maintain, and promote it? What are the trends?

Customer segments

15. Who are the main users of your LOGD services? Is LOGD only used by external customers or also for back-office exchange (e.g. with other agencies)?
16. How often is LOGD used? What are the trends?

Revenue systems

17. How is the provisioning of LOGD funded?
18. Which pricing mechanism or other source of income exists?
19. Under which licence is LOGD made available for reuse? Can we have a copy?

Channels

20. Which channels are predominantly used to consult LOGD: Web API? Web site? (mobile) App? Data market?

Customer relationships

21. Do you have a branding strategy for LOGD?
22. Do you invest in advertisement for LOGD?
23. Do you provide / make use of any mechanisms on user feedback or evaluation?

Annex II. DETAILED INFORMATION GATHERED FROM THE CASE STUDIES

This annex contains the detailed information that was gathered for the 14 case studies.

II.1 AT - Renewable Energy and Energy Efficiency Partnership (REEEP)

II.1.1 Desk research

Working with the Renewable Energy Network for the 21st Century[41], the Renewable Energy and Energy Efficiency Partnership, REEEP[42], curates and publishes data through its reegle.info Web site[43]. REEEP is not a public sector body, rather it is a non-profit organisation that includes 45 governments among its 385 partner organisations. Its work is largely project-based, using renewable technologies to make improvements in the developing world. Data available through reegle.info is gathered from several sources including the World Bank, Open Energy Info (a wiki dedicated to renewable energy), the CIA Factbook, FAO and more. The Open Energy Info wiki[44] is a key source of data and REEEP is a partner in the running of that service.

An important tool is the Reegle Tagging API. This automatically recognises key words and phrases related to energy and energy efficiency in natural text and so offers a scalable way to generate metadata about documents. The Tagging API handles documents in many different formats including PDF. It is this ability to curate and classify natural language texts that is perhaps the real value-add offered by Reegle over and above that made available through Open Energy Info. Reegle's thesaurus for climate compatible development, renewables, efficiency, green growth and other sectors is offered as RDF (as a SKOS Concept Scheme) and via SPARQL endpoint. The Tagging API is itself based on the Thesaurus.

Reegle is based entirely on linked data technologies and there is a strong partnership between REEEP and the Semantic Web Company[45] in providing the service. The data is available a static RDF files (in RDF/XML or N3 formats) and via a SPARQL endpoint. Data is available about several topics such as 'actors' (details of organisations involved with and related to renewable energy) and project outputs (brief details of relevant projects, the technologies used, countries covered etc.). Perhaps the most interesting feature of reegle.info is country profiles. These are available as human readable Web pages but the data behind them is available as LOD (via content negotiation) and via a SPARQL endpoint.

[41] http://www.ren21.net/

[42] http://www.reeep.org/

[43] http://www.reegle.info/

[44] http://openei.org/

[45] http://www.semantic-web.at/

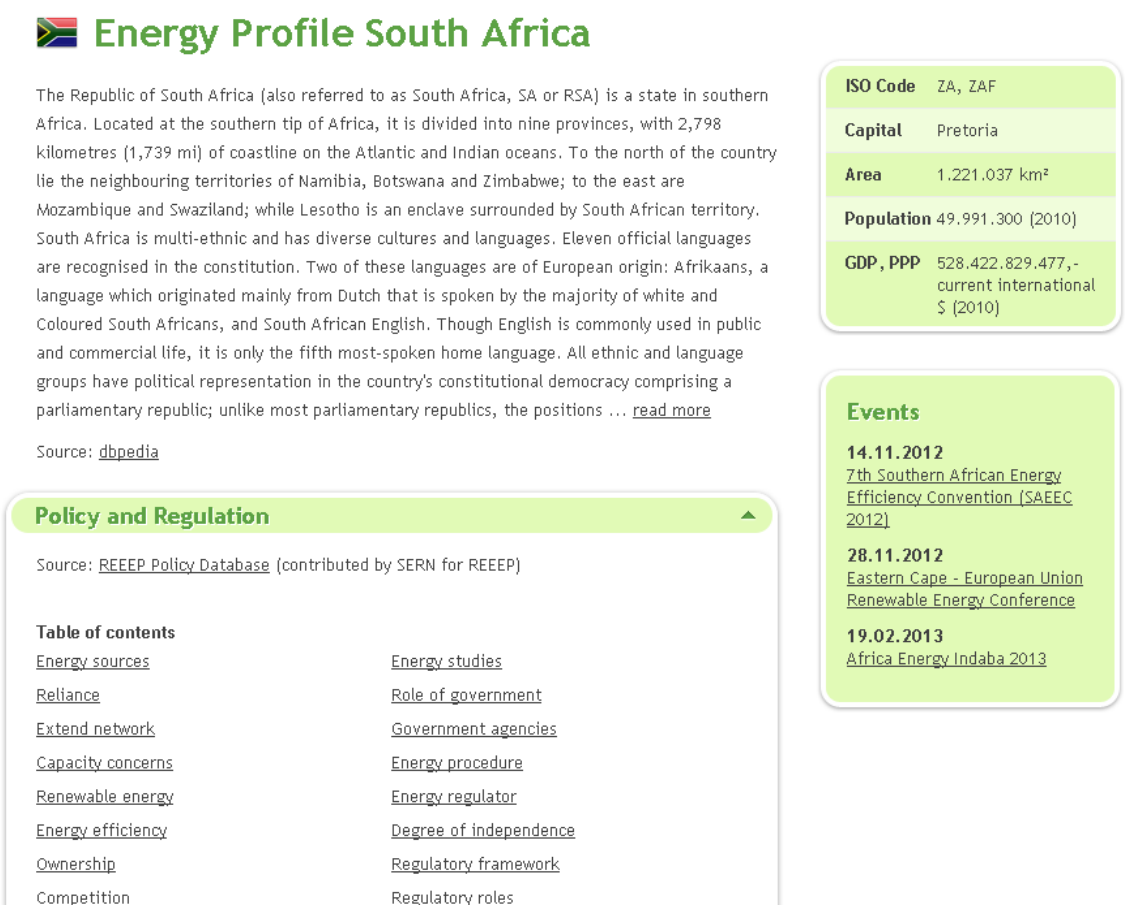


Figure 5 Partial screenshot of the Reegle country profile of South Africa. Notice the 'Extended network' link, the data behind which is shown in Figure 6)

The country profiles presented to users (Figure 5) include links to the available data so that the page shown in Figure 5 is a compilation of available data that is composed at run time.

The Reegle project makes efforts to ensure that its data is readily accessible via other means too. A Word Press plug in provides the Tagging API functionality to users of that blogging platform and among the widgets[46] there's one that enables anyone to add the thesaurus to their Web site.

All data is available for free and is published under the UK Open Government Licence.

[46] http://www.reegle.info/add-reegle-to-your-site

Figure 6 Partial screenshot of the HTML view of some of the data collated in the country profile (Figure 5)

II.1.2 Interview

Interview date	16/7/13
Interviewee	Florian Bauer, REEEP Operations and IT Director
Interviewer	Phil Archer

Q: How does the supply of Linked Data relate to the public task of your organisation?

A: reegle.info is a pillar of the work of REEEP. It's a central tool for sharing information which is one of our main activities. Most of our funding comes from governments and sharing data is a major argument when we propose projects. It's a USP of REEEP that we combine project management with a set of tools to communicate outcomes and achievements.

Q: Does your organisation have a "business case" document that motivates your investment in supplying Linked Data? Can you share it with us?

A: Our Strategy Overview document 2012-2015[47] makes it clear that supplying data is a central part of what we do. See, for example page 7:

REEEP supports the acquisition and distribution of knowledge in two ways, by actively managing its friends, partners, and thematic networks, and by building up information and knowledge services through its web portal reegle.info and its pivotal role in the Linked Open Data movement. In both cases, REEEP acts as a rapid, lean and flexible service provider helping other organisations to enhance their work and link people, Information, and knowledge.

Linked Data is a new field for our funders. They like the principle but it's hard to explain in a non-tech way. A second challenge we have is that some funders don't 'get' the open data idea. We're considering adding 'closed linked data' in future projects - we can't necessarily open everything.

REEEP didn't have a legacy infrastructure. Project results were published but not as open data. Now everything we do is published as a matter of policy.

Bear in mind that REEEP doesn't deal with data in the sense of properties and values, it publishes text and links to documents. The common issues of data quality therefore don't apply in the same way as for other LOD publishers.

We link our thesaurus to others and that helps to improve the quality of our won data.

Q: What are the main enablers / inhibitors for Linked Data to deliver value for its reusers?

A: The biggest inhibitor for us is licensing. Lots of organisations have a history of holding on to their data or publishing it under very restrictive licences. Publishing combined data can invoke a whole chain of licences.

Also we need a good way to define schemas. How do we know which schemas are available and usable and therefore when we need to define our own?

Enablers - we can make a good argument about lowering the costs, reducing duplication and re-using data. This fits in with governments' desires to push freely available open data.

Q: Do you supply service level statements for Linked Data?

A: No. We would need a business case for that, i.e. we'd need to charge for our data in order to be able to make guarantees about its availability.

Q: Does your organisation have a URI policy?

A: No. For now our scale is such that it's manageable since more or less everything comes through me but as we grow and more people become involved a URI policy is likely to become more needed.

[47] http://www.reeep.org/sites/default/files/REEEP%20Strategy%20Overview%202012-2015.pdf

Q: Can you describe the Linked Data infrastructure of your organisation?

We have our own infrastructure hosted on managed servers (in Germany). All software is licensed by us. We use Virtuoso, Pool Party etc. All managed for us by the Semantic Web Company.

Q: Which skills and competencies did you need in order to supply Linked Data? Do you have them in-house or contracted?

I was brought in to begin the publishing side of things and it was the Semantic Web Company that showed me the value of the approach (we'd already been in touch with them about earlier projects). SWC is an important part of the story. The number of staff involved with Reegle.info has grown from 1 to 2 and will soon be 3 but we depend on SWC for implementation. They run the full service which includes our Web site and enterprise knowledge management system too.

Q: Which organisations are key partners in the supply of Linked Data?

A: Our closes link is with the US National Renewable Energy Laboratory, NREL, which is the organisation behind OpenEI - the Open Energy Information wiki. For example, they have lots of maps of renewable energy potential that we link to and they use our data. It's a two way street.

Q: So why do you need both OpenEI and Reegle.info?

A: Because they're different in target audiences (OpenEI is mostly US focussed, Reegle is developing country-focussed). They look and feel different.

As well as OpenEI we also get data from FAO, Eurostat, the World Bank, the UN and DBpedia.

All that data comes to use as CSV (in the best cases) and we have to triplify that which we have the tools to do (LODMS[48], an output of the LOD2 project). This means that we can effectively run a SPARQL query against a CSV file. We developed this tool for our own use under one of our German government funded projects.

Q. Can we talk about costs…

A: The problem is separating out the relevant costs. Reegle costs a few hundred K per year and over the past 10 years has probably cost several million all told. The linked data aspects have come to less than €1M.

Q: Other than yourselves, who are the main users of your Linked Data services?

A: The most prominent user of our data is OpenEI but there are plenty of Web sites that use our policy and regulation data, such as CI Grasp[49] and Climate Tech Wiki[50]. The latter uses our RDF data about the actors in the renewable energy field. In all

[48] http://www.semantic-web.at/linked-open-data-management-suite-lodms

[49] http://cigrasp.pik-potsdam.de/countries/833900607/energy_profile

[50] http://climatetechwiki.org/

we know of about 20 organisations that use our data, of which about 5 use the linked aspect.

We're creating an App as well.

Q: Examples?

A: I don't know - we don't track other users. I'd have to go into the server logs and do some digging to find out. It's really not important to us. We use linked data to achieve our goals, not to power other people's applications. They're welcome to use the data, of course, and we know that the total number of Reegle users is going up but there is no need from our point of view to track who those users are any more than we track who visits the Web site. We don't record the number of SPARQL queries executed separately from other Web site stats.

It's hard to explain the value of open data to our funders. They're more concerned, as we are, with the number of people who use our site and that's about 200,000 unique users per month. The linked data we have is about servicing those users.

Q: Can people download all your data in one go?

A: Yes, but we don't promote it and it's not used much. If people do want our data we'd rather they used the SPARQL endpoint as that's always got the up to date information. Remember that it's mostly text and links to text we provide, not 'data' in the traditional sense.

Q: How is the provisioning of your linked data funded?

A: Project by project. It's hard if not impossible to get funding for continuing to do what you already do so projects are sought to add new features or tools. That's what gets funded.

Q: Do you have a branding strategy for linked data? Do you invest in promotion?

A: We run LOD workshops and do some capacity building with other providers. The workshops promote the concept of LOD. The first was in Abu Dhabi[51] where we launched our linked data guide[52]. The event attracted huge interest, and so we repeated it in Washington[53], Bonn[54] and Bangkok.

We now include events like this in our project proposals. They typically cost between €10-€20K.

Conversations and feedback from events like these often lead to new project ideas. The feedback is always positive. For example, Bernadette Hyland of 3 Round Stones (co-chair of the W3C Government Linked Data WG) was very enthusiastic about the Washington event which was the first dedicated LD event she'd seen in the USA.

[51] http://www.reeep.org/news/linked-open-data-guide-launched-masdar-workshop

[52] http://www.reeep.org/sites/default/files/LOD-the-Essentials_1.pdf

[53] http://www.reeep.org/news/linked-open-data-lifts-washington

[54] http://www.reeep.org/events/workshop-increasing-access-climate-and-energy-data-linked-open-data

Figure 7 Participants in REEEP's first linked data event in Abu Dhabi

Interview Summary

Reegle.info is primarily a service for humans who visit the site for information about renewable energy and energy efficiency. To make the service as effective as possible, REEEP uses linked data technologies as that's the most suitable to their needs, particularly in terms of combining different data sets. The data itself is mostly textual or metadata about documents. The nature of the organisation is such that they want to share the knowledge and information as widely as possible and LOD helps with that. REEEP believes that we all should try to make all info and data that we produce as widely known as possible and as easy to reuse as possible. This is used as an important argument when seeking public funding money funded organisation.

The tools and services are used by several other organisations for whom it is easy to include data drawn from Reegle.info, in particular its country profiles. The continuation of the service depends on the development of new projects of the sort that governments are willing to fund.

II.2 DE – German National Library (DNB)

II.2.1 Desk research

Source: http://www.dnb.de/EN/wir

The German National Library (DNB) is entrusted with the task of collecting, permanently archiving, bibliographically classifying and making available to the general public all German and German-language publications from 1913 onwards, sheet music and sound recordings, foreign publications about Germany, translations of German works, and the works of German-speaking emigrants published abroad between 1933 and 1945. The German National Library maintains co-operative relations on the national and international level. It is, for instance, the leading partner in developing and maintaining rules and standards in Germany and plays a significant role in the development of international standards.

It is a federal institution with legal capacity under public law. The annual funds provided from the budget of the Minister of State for Culture and the Media currently amount to roughly EUR 45 million.

http://www.dnb.de/SharedDocs/Downloads/EN/DNB/service/linkedDataModellierungTiteldaten.pdf?__blob=publicationFile

DNB started to publish its authority data as Linked Data in 2010. Bibliographic data was added to the DNB's existing Linked Data service in January 2012. Since then the DNB data has been available in RDF format under Creative Commons Zero[55] licence.

The first data to be incorporated was the bibliographic data of the library's main holdings (excluding printed music and the holdings of the German Exile Collections) and the serial publications (journals, newspapers and periodicals in the German Union Catalogue of Serials (ZDB)[56]). This documentation describes the modelling of both sets of holdings.

The purpose of the procedure for selecting and modelling the data is to reference the bibliographic data of the DNB and the ZDB in RDF. The bibliographic records are not represented in their full complexity in RDF, rather specific elements have been selected for linked data representation which are required for identification of the resource.

The modelling follows the core element set recommendations of the DINI WG KIM Bibliographic Data group3 V.1.0[57]. This working group consists of representatives of the library networks and a handful of large libraries in the German-speaking countries. It has set itself the target of harmonising the RDF representations of bibliographic data in the German-speaking countries. The DNB is playing a leading role in this process.

Source: http://dnb.de/EN/lds

[55] http://creativecommons.org/publicdomain/zero/1.0/

[56] http://www.zeitschriftendatenbank.de/

[57] https://wiki.dnb.de/pages/viewpage.action?pageId=68060017

Developments in the area of the semantic web are aimed at improving the usefulness and accessibility of data. The idea of the Semantic web also allows links to be created between data from heterogeneous sources, leading in turn to the establishment of new services. As a result of the linked-data movement, many providers (mostly non-profit organisations, universities or public institutions) are already offering their data in a form which is semantic-web-compatible. Above all, this includes data which is of general use within the public domain. Examples include geographical information, thesauri, encyclopaedias as well as bibliographic and authority data.

Libraries, too, have recognised the great potential offered by this form of data publication. The first institutions are already actively offering their information as linked data, or are planning to do so. The German National Library is committed to making a significant contribution to ensuring the stability and reliability of the "linked-data-cloud" by providing data which has largely been generated and maintained by trained professionals. The German National Library with its high quality data intends to become one of the mainstays of the semantic web.

In the long term the German National Library is planning to offer a linked data service which will permit the semantic web community to use the entire stock of its national bibliographic data, including all authority data. A suitable data service needs to be created to distribute the new data format alongside the already established access channels (OAI[58], SRU[59] etc.).

The German National Library is endeavouring to make a significant contribution to the global information infrastructure with its new data service by laying the foundations for modern commercial and non-commercial web services.

The German National Library collaborates with the Bibliographic Framework Transition Initiative[60], which aims to determine a transition path for the MARC 21[61] exchange format in order to reap the benefits of newer technology while preserving a robust data exchange that has supported resource sharing and cataloguing cost savings in recent decades.

The Linked Data Service is already incorporated in the web portal and therefore publicly accessible. There are different methods of accessing data in RDF/XML: by accessing the appropriate URIs/URLs, though SRU and OAI, or by downloading an FTP data dump.

II.2.2 Collected metrics

Usage

No metrics are available for the number of de-referenced URIs and queries. There is no information about the numbers of governmental and commercial reusers. The main users are currently in the public sector, but it is expected that in the future

[58] Deutsche Nationalbibliothek. OAI Interface – Overview. http://www.dnb.de/EN/oai

[59] Deutsche Nationalbibliothek. SRU Interface – Overview. http://www.dnb.de/EN/sru

[60] Library of Congress. Bibliographic Framework Transition Initiative. http://www.loc.gov/bibframe/

[61] Library of Congress. MARC 21 Format for Bibliographic Data. http://www.loc.gov/marc/bibliographic/

also commercial reusers will want to use the data, e.g. search engines, publishers and other information services.

From the reuser side, over 8,000 GND URIs are used by Museum-digital.

Revenue and other sources of income

Funding of the linked data activities is from the general funding of the library. There is no revenue from subscription fees, on-demand fees or advertisement.

Cost

Development cost is estimated at approximately 221 person days.

Up to June 2012, the service was a project, from July 2012 and onwards, it is a product. Costs are considered to be part of the general bibliographic services and general product maintenance.

Benefits

Linked datasets include:

- Authority data: Wikipedia, DBpedia and the STW Thesaurus for Economics[62] with more upcoming.
- Bibliographic data: ZDB[63] and culturegraph.org[64] with links to the British Library upcoming.

There are several prototypes that use the Linked Data. One operational service that uses DNB data is Museum Digital (http://www.museum-digital.de/).

The number (and increase) of corrections requested cannot be measured separately for the linked data application.

As DNB does not build on the Linked Data infrastructure internally, there are no cost reductions on information integration.

II.2.3 Provider interview: Lars G. Svensson

Interview date	6 August 2013
Interviewee	Lars G. Svensson (Advisor for Knowledge Networking)
Interviewer	Makx Dekkers

Q: How does the supply of Linked Data relate to the public task of your organisation?

A: The bibliographic descriptions are created from the German National Bibliography, from which data is published in several formats, e.g. MARC 21 for libraries.

[62] ZBW - Leibniz Information Centre for Economics. STW Thesaurus for Economics. http://zbw.eu/stw/

[63] Staatsbibliothek zu Berlin Preussischer Kulturbesitz. Zeitschriftendatenbank (ZDB). http://www.zeitschriftendatenbank.de/

[64] Culturegraph.org – Plattform für Wissensvernetzung. http://www.culturegraph.org/

From a strategic perspective, the national bibliography has the objective to make the data available as widely as possible. As outlined in the paper "The Deutsche Nationalbibliographie as linked open data: Applications and opportunities"[65] a national bibliography needs to be accessible on the Web, and linked data is the emerging technology for data access on the Web. Publishing this data as linked data is a means to reach out to audiences outside the traditional library sphere. Generally spoken, RDF is just another serialisation of the data that is provided anyway.

Q: Does your organisation have a "business case" document that motivates your investment in supplying Linked Data? Can you share it with us?

A: There is no specific business case document. The linked data activities are part of the general aim to reach out to communities outside the library sector, encouraging them to build applications on data from DNB, thereby promoting government data reuse.

Q: What are your future plans? Do you plan to expand or abolish the supply of Linked Data?

A: The provision of linked data is considered a strategic priority. It is expected that investment will increase.

Q: Did the supply of Linked Data give rise to new opportunities for flexible data integration within your organisation and with partner organisations? Increase in data quality? New services? Cost reductions?

A: Supply of Linked Data created better opportunities to reference data for integration in third party services, e.g. for services like museum-digital.de[66]. Also the use of URIs enables unique identification of e.g. persons from the cultural heritage sector. This enables better reuse of data and provides a pivot point for information about a specific entity (e.g. by differentiating persons or places having the same lexical string as their name; Paris <France> vs. Paris <Texas> vs. Paris <Prince of Troy>)

Q: What are the main enablers / inhibitors for Linked Data to deliver value for its reusers?

A: The main enabler is the ease with which relationships can be added to existing data when related resources are discovered. An inhibitor is the fact that there are several vocabularies to choose from to express bibliographic information but the absence of a widely agreed set of vocabularies and application profiles. Furthermore, there is a lack of tools to support integration of linked data into library workflows. Vendors of library management systems need to work with the community to provide appropriate, production-grade tools.

Q: Do you supply service level statements for Linked Data?

A: Service Level Agreements are not supplied.

---

[65] Jürgen Kett, Sarah Beyer, Mathias Manecke, Yvonne Jahns, Lars G. Svensson. The Deutsche Nationalbibliographie as linked open data: Applications and opportunities. http://nbn-resolving.de/urn:nbn:de:101-2012052306

[66] Museum-digital.de. http://museum-digital.de/

Q: Does your organisation have a URI policy? Can we have a copy?

A: All linked data is supplied through the domain d-nb.info. URI patterns are:

http://d-nb.info/(internal reference number) for bibliographic data

http://d-nb.info/gnd/(authority record identifier) for authority data

http://d-nb.info/standards/elementset/(term) for ontologies

http://d-nb.info/standards/vocab/(term) for value vocabularies

Q: Can you describe the Linked Data infrastructure of your organisation?

A: At the heart of DNB's Linked Data infrastructure is an open source library and toolkit (Metafacture[67]) which is developed as a community initiative and managed by DNB, providing a programming language for transformation of semi-structured data called Metamorph. All conversions from the internal data representation (in the internal PICA+ format) to RDF/XML are specified in Metamorph.

On-line access to the RDF/XML representation of the data relies on an on-the-fly conversion of the record/resource requested. This is achieved by including Metafacture as a library in the Linked Data Service web application.

Complete RDF/XML dumps of the data sets are created regularly by processing a dump of the PICA+ data with Metafacture/Metamorph. Effective from 2014, RDF dumps will be available three times per year (months 1, 5, and 9).

DNB is currently in the process of improving the infrastructure and tools for maintaining information about relations between different data sets which is included for instance as same-as relations in RDF/XML data.

Q: Which skills and competencies did you need in order to supply Linked Data? Do you have them in-house or contracted?

A: Skills are all in-house, from modelling to technical infrastructure. The following expertise was needed:

- Knowledge of the library data structure and support from cataloguing specialists
- Knowledge of relevant RDF vocabularies for description of bibliographic and authority data
- Expertise in mapping internal database structure to RDF vocabularies
- Software development skill in order to integrate provision of RDF data into the other services.

Q: Which organisations are key partners in the supply of Linked Data?

A: DNB's data supply is very much self-contained in that they create all data as a part of its public tasks. Regarding standardization, DNB co-operates with the German library networks, Staatsbibliothek zu Berlin - Preußischer Kulturbesitz (responsible for the Deutsche ISIL-Agentur und Sigelstelle) and with other national libraries, e.g. the British Library, the Bibliothèque nationale de France and the Swedish national library.

[67] Metafacture. https://github.com/culturegraph/metafacture-core/wiki

Q: Which activities do you carry out to supply Linked Data?

A: Development of mappings between internal database format and RDF vocabularies, implementation of data conversions and standardisation work.

Q: Which investments has your organisation made to enable it to supply Linked Data?

A: Four projects totalling 221 person days for implementation of basic infrastructure (conceptual models, data mappings, software development etc.).

Q: How many FTEs in your organisation are involved in provisioning the supply of Linked Data?

A: Approximately 1 FTE is involved at DNB in the provision of Linked Data, but many of the activities involved in supplying the service are part of the general bibliographic services which makes it impossible to specify the exact cost of the supply of linked data.

Q: Which costs have you incurred to publish Linked Data, maintain, and promote it? What are the trends?

A: The cost of developing and running the Linked Data services is associated with the staff mentioned in the previous two answers.

Q: Who are the main users of your Linked Data services? Is Linked Data only used by external customers or also for back-office exchange (e.g. with other agencies)?

A: Other cultural heritage institutions use DNB's Linked Data, particularly referencing the authority data in the GND[68]. It is currently not used for data exchange between libraries.

Q: How often is Linked Data used? What are the trends?

A: There are currently no metrics for the use of the Linked Data services; those are planned, but not yet implemented.

Q: How is the provisioning of Linked Data funded?

A: Funding of the Linked Data activities is from public funding (part of everyday duties).

Q: Which pricing mechanism or other source of income exists?

A: Access to and reuse of the Linked Data is free of charge.

Under which licence is Linked Data made available for reuse? Can we have a copy?

All Linked Data is made available under the Creative Commons Zero Public Domain Dedication licence[69]. Initially DNB created their own licence but finally decided to follow the same approach as the Europeana Data Exchange Agreement. An

[68] Deutsche Nationalbibliothek. Gemeinsame Normdatei (GND). http://www.dnb.de/EN/gnd

[69] CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. http://creativecommons.org/publicdomain/zero/1.0/

overview of the history behind this decision is contained in the document "Licensing Library and Authority Data under CC0: The DNB Experience"[70].

Q: Which channels are predominantly used to consult Linked Data: Web API? Web site? (mobile) App? Data market?

A: Access to the Linked Data is through URI dereferencing and data dumps. Resolution of a URI identifying bibliographic or authority data results in access to the actual data in the database; on the way in, the URI is mapped to the internal reference while on the way out the internal format is converted to RDF on-the-fly. Data dumps are created through a bulk conversion process. Almost all of the bibliographic records and almost all data elements from the bibliographic records are made available as linked data (excluding some data that are just relevant for library applications). All authority records and all information from those records is made available as linked data.

Q: Do you have a branding strategy for Linked Data?

A: DNB does not have a branding strategy for Linked Data.

Q: Do you invest in advertisement for Linked Data?

A: There is no investment in advertising, except for PR work done at conferences, through mailing lists, the web page at http://dnb.de/EN/lds, articles in journals, participation at book fair and library and information related fairs.

Q: Do you provide / make use of any mechanisms on user feedback or evaluation?

A: There is a mailing list for feedback, http://lists.d-nb.de/mailman/listinfo/lds.

II.2.4 Reuser interview: Stefan Rohde-Enslin

Interview date	15 August 2013
Interviewee	Stefan Rohde-Enslin (Museum-digital)
Interviewer	Makx Dekkers

Q: How does the reuse of Linked Data relate to the public task of your organisation?

A: Museum-digital is not an established organisation but a community initiative. Its objective is to help museums to make their collections visible on the Web. It involves several regional museum associations in Germany and contains description of the collections of 260 museums representing some 30,000 museum objects. Linked Data is a tool that enables the harmonisation of metadata across collections.

Q: Does your organisation have a "business case" document that motivates your investment in re-using Linked Data? Can you share it with us?

[70] Lars G. Svensson. Licensing Library and Authority Data Under CC0: The DNB Experience. http://www.w3.org/2013/04/odw/odw13_submission_57.pdf

A: The use of Linked Data grew out of the possibilities that were offered by the availability of Linked Data reference collections, e.g. GND, DBpedia, GeoNames etc.

Q: What are your future plans? Do you plan to expand or abolish the reuse of Linked Data?

A: As the Linked Data reference collections become more stable, museum-digital will increase the usage, e.g. to start making connections between people (including family relations) to support navigation.

Q: Did the reuse of Linked Data give rise to new opportunities for flexible data integration within your organisation and with partner organisations? Increase in data quality? New services? Cost reductions?

A: The Linked Data is used to enable harmonisation of metadata so that searching based on people, places and subjects becomes more efficient. The objective of museum-digital is to provide discovery services that did not exist before.

Q: What are the main enablers / inhibitors for Linked Data to deliver value for its reusers?

A: Linked Data enables the harmonisation of museum metadata, which helps smaller museums to contribute to Europeana and also helps them to create inventories of their own collection.

A problem in using Linked Data is that many of the systems that provide data do not yet have production-grade reliability. The data is not always stable and sometimes not available due to downtime. It would be helpful if owners of those collections found some way to inform their reusers of downtime and major revisions.

Q: Do your suppliers provide service level statements for Linked Data? What happens if the data is (temporarily) unavailable?

A: No, there are no guarantees. If reference data (e.g. GND or DBpedia) is unavailable, only the data that the museum has provided will be visible in the user interface.

Q: Do your suppliers have a URI persistence policy? If not, how do you protect yourself from broken links?

A: Even if the URI stays the same, the structure and content of the data that the URI resolves to may change which then requires a rewrite of the system that reuses the data.

Q: Can you describe the Linked Data infrastructure of your organisation?

A: Museums contribute descriptions of their objects and use tools from museum-digital to link data to the reference data. Museum-digital enhances the data that is then indexed in the portal.

Q: Which skills and competencies did you need in order to reuse Linked Data? Do you have them in-house or contracted?

A: No particular skills other than software development.

Q: Which organisations are key partners in the reuse of Linked Data? E.g. suppliers of Linked Data that you are re-using, technical partners?

A: Museum-digital uses Linked Data from the German National Library (GND), DBpedia, GeoNames, Library of Congress (LCSH) and others.

Q: Which activities do you carry out to reuse Linked Data?

A: When a museum creates a description for inclusion in museum-digital, the cataloguer gets suggestions for named entities (persons, places). If the entity is not yet in the system, the cataloguer enters the name. Museum-digital staff then searches (at the push of a button) for the person or place in GND, Wikipedia and then includes the URI of the entity in the description with some descriptive text to allow indexing.

Q: Which investments has your organisation made to enable it to reuse Linked Data?

A: No major investments were made. Some activities, e.g. data cleansing, were funded through small project grants.

Q: How many FTEs in your organisation are involved in the reuse of Linked Data?

A: Overall, about 8-10 people are involved in providing, preparing and enriching the data across the participating organisations.

Q: Which costs have you incurred to reuse Linked Data, maintain, and promote it? What are the trends?

A: Programming time, mostly provided as volunteer contribution. One person responsible for data cleansing and linking.

Q: Who are the main users of your Linked Data services? Do you provide services based on the reused Linked Data or do use it for internal purposes (e.g. for data integration)?

A: The data is used to enhance the services on the portal and as a tool for the cataloguers in the museums.

Q: How often is Linked Data used? What are the trends?

A: Linked Data is used for all object descriptions.

Q: How are your Linked Data activities funded?

A: The activities are mostly based on voluntary effort and staff at the participating organisations with some small Project grants.

Q: Which pricing mechanism or other source of income exists?

A: The use of Linked Data is free of change.

Q: Under which licence do you get the Linked Data? Under which licence do you give access to the Linked Data that you reuse?

A: It is not always clear what the reuse conditions are. In the case of GND, Dr, Rohde called the National Library asking if he could use the data. The linked data that Museum-digital creates is made available under a CC-BY-NC-SA licence.

Q: Which channels are predominantly used to import Linked Data? E.g. direct URI resolution, SPARQL endpoint, data dump?

A: URI resolution only; data dumps are usually outdated.

Q: Do you have a branding strategy for Linked Data?

A: No. This may be something to do if the infrastructure becomes more reliable.

Q: Do you invest in advertisement for Linked Data?

A: No, other than giving presentations about it.

Q: Do you make use of any mechanisms for user feedback or evaluation?

A: No. It would be useful if providers of reference data had feedback channels.

II.3 EU – Directorate General for Health & Consumers

II.3.1 Desk research

DG Health & Consumers is a DG of the European Commission which aims to empower consumers, protect and improve public health, ensure Europe's food is safe and wholesome, protect the health and welfare of farm animals and protect the health of crops and forests[71].

Value proposition: Data made available for free

DG Health & Consumers is a pioneer when it comes to the provisioning of LOGD in its field, e.g. reference datasets of pesticides, food enzymes and additives. DG Health & Consumers makes available its Linked Data for free to the public, including potential reusers such as other DGs, EU agencies, private companies and citizens.

The LOGD provided by DG Health & Consumers are licensed under the "European Union Public Licence[72]".

EUPL is a software licence that has been created and approved by the European Commission. It is a free software licence.

This licence was originally intended to be used for the distribution of software. However, its generic scope makes it also suitable for licensing data. The main goal of EUPL is its focusing on being consistent with the copyright law in the 28 Member States of the European Union, while retaining compatibility with popular open-source software licences such as the GNU General Public License.

EUPL v1.2 is currently being drafted.

Key findings

- **Linked Data, free of charge**: All LOGD of DG Health and Consumers is provided free of charge.

II.3.2 Collected metrics

Metric
Usage • **Number of de-referenced URIs / queries:** N/A • **Number of governmental reusers of LOGD:** Estimated around 15-20 • **Number of commercial reusers of LOGD:** Not available as the LOGD is not yet in production.
Revenue and other sources of income • LOGD is funded by public funding.

[71] http://ec.europa.eu/dgs/health_consumer/about_us/who_we_are_en.htm

[72] http://joinup.ec.europa.eu/software/page/eupl

Metric
Cost • **Development and maintenance cost**: ~110k per annum. • **Promotion costs per year**: 0
Benefits • **Number of linked datasets (outgoing and incoming links):** 10-20 linked datasets to start with. • **Number of derived applications:** currently one but more are expected.

II.3.3 Interview with R. Ní Bhraonáin & Giorgos Georgiannakis

Interview date	2013-07-05
Interviewee	Ruth Ní Bhraonáin and Giorgos Georgiannakis, IT Project Officers at DG Health & Consumers
Interviewer	Nikolaos Loutas

Q: How does the supply of LOGD relate to the public task of your organisation?

A: First, we would like to clarify that the DG Health & Consumers public task fits into that of a wider organisation, i.e. the European Commission. The specific mission statement of DG Health & Consumers is to make Europe a healthier, safer place, where consumers can be confident that their interests are protected. In this context, the DG is subject to obligations under legislation to collect and share information between Member States and stakeholders.

The supply of LOGD relates to the public task of the DG, which is to empower consumers, protect and improve public health, ensure Europe's food is safe and wholesome, protest the health and welfare of farm animals and protect the health of crops and forests[73]. We will achieve our goals by watching, listening to people's concerns and acting.

- Monitoring - once the EU has passed laws on food and product safety, consumer rights or public health, it is up to national, regional and local governments to apply those laws - to ensure traders, manufacturers and food producers stick to the rules. Part of our job is to check that this is really happening.
- Listening - to be effective, our policies must take account of related EU policies on trade, competitiveness and the environment for example, and the concerns of our stakeholders. Through broad consultation, we want to hear from all interested parties.
- Acting - where EU action is needed, we make proposals using a mixture of laws, support for projects and other measures. We also support national or regional authorities where they are better placed to act.

[73] http://ec.europa.eu/dgs/health_consumer/about_us/who_we_are_en.htm

In the context of public sector information and in particular with regards to open data, the Commission and DG Communications, Networks, Content & Technology (CONNECT)[74], the responsible policy DG, aims to advance the Commission's open data policy by means of:

- The Revised PSI Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the reuse of public sector information[75]; and
- The Communication on Open Data[76].

Consequently, DG Health & Consumers has already published 12 of its public datasets in open data format and made them accessible via the European Union Open Data Portal[77] managed by the Publications Office[78]. Going one step further and publishing its data as LOGD allows DG Health & Consumers to formally encode in the open data links and dependencies between the data that have already been observed but could not until now be made explicit. Hence, linked data helps DG Health & Consumers break down the information silos that still exist.

Linking data is a key enabler for achieving the key objective of DG Health & Consumers, i.e. "make Europe a healthier place". The policy officers of DG Health & Consumers, the Member States national administrations and stakeholders need access to high-quality data and knowledge. Linked data facilitates the exchange and sharing of this data and knowledge.

DG Health & Consumers has demand from external organisations asking for complex information in different formats, linked data being one of them. This was an additional driver for moving towards linked data.

Q: Does your organisation have a "business case" document that motivates your investment in supplying LOGD? Can you share it with us?

A: We already explained, the supply of LOGD by DG Health & Consumers is a direct consequence of its obligation to conform to the revised PSI Directive and open up legally public data.

In the case of DG Health & Consumers, the legal obligation is complemented by a strong business requirement, as a number of companies, organisations and associations from different communities contact us with ideas of reusing our data, mostly visualising it in different ways to highlight different aspects of it.

Additionally, linked data can serve intra-organisational purposes. In order to reduce and manage the risk for consumers, DG Health & Consumers needs to take decisions based on data coming from national administrations, for example DG

[74] http://ec.europa.eu/dgs/connect/en/content/dg-connect

[75] Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the reuse of public sector information. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF

[76] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Open data - An engine for innovation, growth and transparent governance, 2012, http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive_proposal/2012/open_data.pdf

[77] http://open-data.europa.eu/en/data/publisher/sanco

[78] http://publications.europa.eu/index_en.htm

Health & Consumers' auditors in the field need access to aggregated data across diverse areas. In such cases, linked data helps DG Health & Consumers relate its own data with data coming from the Member States, but also national datasets with each other.

Hence, we see that often DG Health & Consumers acts as a disseminator of data and a registry of consolidated knowledge coming from all Member States. The links between the data span across national borders, policy areas and sectors. For example, all EU Member States are obliged to collect certain data on an annual basis, e.g. metrics with regards to residues or pesticides. Linked data allows for traceability over different reporting periods, the establishment of trends, linking data to legislation and makes the implementation of the DGs mission statement based on clear evidence.

Without linked data technologies, this massive data integration exercise is more error-prone and demands more time and resources, as the end-user has to consult different country registers and consolidate the data herself.

Q: What are your future plans? Do you plan to expand or abolish the supply LOGD?

A: We plan to continue making our public data available in open data format and to expand the number of publicly available datasets presented in open data format. In the near future, we are releasing for test the first applications built using our LOGD. We plan to use the open linked data technology and tools developed to manage intra-organisational linked closed data as well, e.g. our Forest Reproduction Material application, which directly serves the purposes of the Directive on the marketing of forest reproductive material[79]. Rapid alert systems, where only members can have access, are another case where we see a potential application of linked closed data.

During a work-flow, some data may be initially private but at a certain stage, e.g. after approval (establishments, feed additives) may be published as public data. Therefore, applying open data technology to link data from the beginning can lead to benefits in the future. For example, in the case of the authorisation of plant protection products, all information about the authorisation is kept private throughout the process. However, once an authorisation is granted, all information about it is made public.

In other cases, information on a particular subject coming from a Member State may be private, e.g. information about animals with a particular disease from a specific farm, but the aggregate data on that subject from the 28 MSs would be public and presented in open data format, which can then be shared for further scientific analysis with agencies or external stakeholders.

In addition to focusing on data, we are extending our work on data models and reference data that can be used for aligning and homogenising data coming from different sources, e.g. DG Health & Consumers systems and systems in the Member

[79] COUNCIL DIRECTIVE 1999/105/EC of 22 December 1999 on the marketing of forest reproductive material, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2000:011:0017:0040:EN:PDF

States. We are also ensuring that our models and reference data reuse and link to widely-used vocabularies in the field, such as Agrovoc.

It is clear that we are ourselves reusers of linked data published by other organisations.

This year, we collaborated with the ISA Programme on a linked data pilot that integrated plant protection product data from eight Member States that highlighted among others the importance of common models and reference data, and aligned identifiers, especially when linked data is used as a data integration technology[80].

Q: Did the supply of LOGD give rise to new opportunities for flexible data integration? Increase in data quality? New services? Cost reductions?

A: Our LOGD is not yet in production, it is still in beta mode. However, we can confirm that there is strong interest and demand from the Member States, from industry and from EU agencies in accessing it and reusing it.

Additionally, as discussed earlier, we are expecting linked closed data to serve intra-organisational needs, for example, in the fields of investigations and consumer protection, where integration of confidential data is required.

The use of URIs is a core element of linked open data developments in DG Health & Consumers and is a fundamental part of any linked data solution. We anticipate it will solve referential integrity problems in the data.

Finally, the inherent ability of linked data to expose data in different machine-readable formats, e.g. CSV, JSON, XML, RDF, and to work well together with other technologies, such as XML, is expected to turn linked data into a viable approach for integrating systems. This is in fact something that we explored in the context of the plant protection product LOGD pilot that we co-developed with the ISA Programme.

Q: What are the main enablers / inhibitors for LOGD to deliver value for its reusers?

A: We believe that the first thing to be done for creating value from LOGD is to raise awareness of the capabilities, the anticipated benefits and the potential uses of LOGD. In this direction, DG Health & Consumers has created two videos featuring two LOGD applications, the RDFa maker and the Forest Reproduction Material client application. Additionally, DG Health & Consumers is informing Chief Health Officers in the MSs on the value offering of LOGD.

Common data models and reference data are enablers for LOGD. Reusing commonly-agreed data models and standardised reference data from authoritative sources facilitates the linking of data coming from different systems, organisations, sectors and countries. DG Health & Consumers is actively working towards this by developing common data models, such as the plant protection product ontology, and reference data, such as taxonomies of plants, pests and substances. Wherever possible, DG Health & Consumers is reusing existing models and reference datasets.

[80]Pilot on Linking data about applications and decisions for authorisation of plant protection products http://health.testproject.eu/PPP/

An enterprise-wide strategy on persistent URI design and management is also key enabler and should not be overseen. Currently, DG Health & Consumers is taking direction from DG OP on URIs and working with them to address any issues.

Clear provenance information, showing that the data comes from a trusted provider is also fundamental for promoting the reuse of LOGD, as it guarantees persistence and quality. Reusers will otherwise invest in building apps with data from sources that are not persistent or trusted.

In our process of publishing our data as LOGD, we have also identified a number of roadblocks including:

- Poor data quality of legacy data;
- Multiple (conflicting/overlapping) classifications from different domains of practice on the same subject/thing, e.g. one taxonomy may classify a particular plant as a tree and another as a bush.
- Mapping between different data models and reference datasets is not easy and requires both domain expertise and technical skills.
- Cultural barriers, for example, people are often afraid of losing ownership of their data or resist to any change in their current way of working (see tools referred to above where the data provider publishes the links to their own data but remains the host and owner of the actual dataset).

Q: Do you supply / require service level statements for LOGD?

A: All the data that we make publicly available is accompanied by an EU disclaimer[81], which clarifies among others that the data is:

- of a general nature only and is not intended to address the specific circumstances of any particular individual or entity;
- not necessarily comprehensive, complete, accurate or up to date;
- sometimes linked to external sites over which the Commission services have no control and for which the Commission assumes no responsibility;
- not professional or legal advice (if you need specific advice, you should always consult a suitably qualified professional).

The open data of DG Health & Consumers is hosted on Circa-BC[82] and is covered by the SLAs of that system. However, there is no SLA in place yet for the LOGD infrastructure.

Q: Does your organisation have a URI policy? Can we have a copy?

A: DG Health & Consumers is following the EC-wide URI policy developed by the Publications Office.

Q: Can you describe the Linked Data infrastructure of your organisation?

A: We have put in place the Linked Data infrastructure of Figure 8 and a SPARQL endpoint accessible at http://ec.europa.eu/semantic_webgate_acceptance/query/

[81] EU disclaimer – legal notice, http://ec.europa.eu/geninfo/legal_notices_en.htm

[82] https://circabc.europa.eu/faces/jsp/extension/wai/navigation/container.jsp

We make our LOGD available in JSON, XML and RDF.

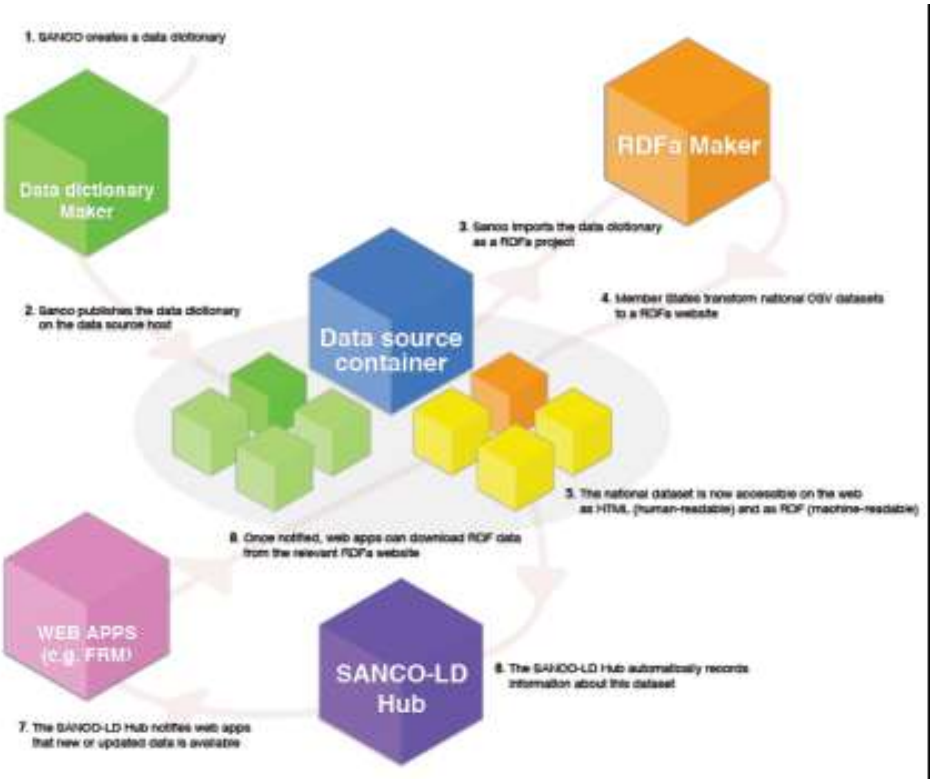


Figure 8 DG Health & Consumers Linked Data infrastructure

Q: Which skills and competencies did you need in order to supply / reuse LOGD? Do you have them in-house or contracted?

A: We are building our LOGD offering mostly using in-house expertise. Our team comprises experienced business analysts, who provide the use cases, and skilled IT professionals, with solid technical backgrounds. We are always trying to keep up with state-of-the-art technologies and favour the implementation of best practice and industry widespread standard solutions. Due to the limited resources that we have, we are following a stepwise approach and we are learning by doing.

Q: Which organisations are key partners in the supply / reuse of LOGD?

A: We are collaborating with the following stakeholders in order to expand the supply of LOGD and promote its reuse:

- EU Member States,
- DG Agriculture and Rural Development (AGRI),
- DG Maritime Affairs and Fisheries (MARE),
- DG Environment (ENV),
- DG Communications, Networks, Content and Technology (CONNECT),
- Publications Office,
- Eurostat,
- European Maritime Agency,

- European Centre for Disease Prevention and Control (ECDC),
- European Food Safety Agency (EFSA),
- European and Mediterranean Plant Protection Organization (EPPO),
- Community Plant Variety Office (CPVO),
- UN/CEFACT,
- Open Archives Initiative,
- All international organisations and stakeholders in our business domains.

We are also trying to engage market operators and citizens.

Q: Which activities do you carry out to supply / reuse LOGD?

A: Generally speaking we see three main stages in the provision of LOGD. First we take data that is already publicly available and publish it in open data formats, we then make it available also as Linked Data and finally we use it internally to develop tools and facilitate access to the data and encourage others to use it in the same way.

As said earlier our focus is on development and maintenance. Data cleansing and harmonisation activities are one of the biggest challenges and require significant time and effort.

We do not invest in promotion as such, but we do try to promote our LOGD to the extent possible at relevant meetings and events.

Q: Which investments has your organisation made to enable it to supply LOGD? Which costs have you incurred to publish LOGD, maintain, and promote it? What are the trends?

A: DG Health & Consumers follows a low-cost strategy. We invest mostly in publishing and hosting LOGD, and on making sure that our people have the required skills and competencies to work with LOGD.

We estimate an annual investment of 110k for publishing and managing LOGD.

We do not have a specific awareness-raising/promotion plan, other than providing links and keywords in the metadata of all published open data and then further publicising the open data where possible, for example at the European Data Forum, as part of the EU Open Data Portal, at SEMIC 2013 and at various other events..

Q: How many FTEs in your organisation are involved in provisioning the supply of LOGD?

A: 1 FTE is involved (and 1 external – contractor).

Q: Who are the main users of your LOGD services? Is LOGD only used by external customers or also for back-office exchange (e.g. with other agencies)?

A: The LOGD provisioned by DG Health & Consumers is expected to be used by both internal and external customers.

Internally, our middle and senior management have expressed their interest to use the data and reports on the data or apps that could help them make better informed decisions.

Our first LOGD app, to support the management of forest reproductive material responding to the requirements of Directive 1999/EC105 is planned for launch in Q4 2013.

DG Health & Consumers also has external requests from organisations asking for complex information in different formats, open format linked data being one of them.

Q: How often is LOGD used? What are the trends?

A: Our LOGD is not in production yet, so we do not have usage statistics.

Q: How is the provisioning of LOGD funded?

A: The LOGD provisioned by DG Health & Consumers is funded by public funds.

Q: Which pricing mechanism or other source of income exists?

A: Following the Commissions Open Data policy, i.e. the principles set in the revised PSI Directive, we provide our LOGD free of charge. No pricing mechanisms exist.

Q: Under which licence is LOGD made available for reuse? Can we have a copy?

A: All our open data, including LOGD, are licensed under the EUPL v1.1 and through the licensing managed by DG OP as part of the EU Open Data Portal.

Q: Which channels are predominantly used to consult LOGD: Web API? Web site? (mobile) App? Data market?

A: DG Health & Consumers LOGD can be queried and retrieved via a public SPARQL Endpoint accessible at http://ec.europa.eu/semantic_webgate_acceptance/query/

All open data published by DG Health & Consumers is accessible via the EU Open Data Portal at http://open-data.europa.eu/en/data/publisher/sanco or can be downloaded from CIRCABC via the links at http://ec.europa.eu/dgs/health_consumer/information_systems/

Once the LOGD goes into production, we expect it to be also made available via apps and data-driven services, such as the Forest Reproduction Material application.

Q: Do you have a branding strategy for LOGD?

A: We do not have yet a formal branding strategy. However, provenance and version information are available when accessing the data, through source and URIs and the data is accompanied by the EU Disclaimer.

As you see, branding the LOGD means for us to clearly declare its origin (provenance). This acts as an enabler of its reuse, as potential reusers can see that the LOGD comes from a trusted source.

Q: Do you invest in advertisement for LOGD?

A: As discussed earlier, we do not have a specific awareness-raising/promotion plan for our LOGD but see earlier answer.

Q: Do you provide / make use of any mechanisms on user feedback or evaluation?

A: Listening to the requirements of the reusers of our LOGD is very important for DG Health & Consumers. Although we do not have yet a formal feedback mechanism/channel set up, we do have regular informal communication via phone/email with people that are interested in reusing our data. Stakeholders (Member States, internal users) are involved in testing the apps and tools that are currently in development and their feedback is iteratively taken into account. In the future, we would like to integrate interactive feedback tools in the LOGD apps that will be developed by DG Health & Consumers. We have already some basic such mechanisms in place in the forest reproduction material application.

II.4 EU – Europeana

II.4.1 Desk research[83]

Source: http://pro.europeana.eu/documents/900548/6e60b1c8-bd7d-4e7b-86bb-cf421ef09341

Europeana is an innovative web portal that opens a doorway to the digital resources of Europe's museums, libraries, archives and audio-visual collections. Visitors can discover, share in, reuse and be inspired by the rich diversity of Europe's cultural and scientific heritage. Books and manuscripts, photos and paintings, television and film, sculpture and crafts, sheet music and recordings and much more. From The Girl with the Pearl Earring to Newton's Laws of Motion, from the music of Mozart to the TV news of times gone by – you can find it all in Europeana (www.europeana.eu).

Europeana is funded by the European Commission and Ministries of Culture in 21 member states. Europeana teams Europe's cultural and scientific heritage with technological innovation. It presents Europe's rich cultural and scientific history online, on tablets, smartphones and via application interfaces (APIs) in ways that are relevant to today's user. It opens new doors to learning and creativity - personal and professional - and as a result has an important contribution to make to Europe's creative and digital economy.

Europeana's strategic activities are outlined in its Strategic Plan (http://pro.europeana.eu/c/document_library/get_file?uuid=c4f19464-7504-44db-ac1e-3ddb78c922d7&groupId=10602) as

- Aggregate: Build the open trusted source for European cultural heritage content
- Facilitate: Support the cultural heritage sector through knowledge transfer, innovation and advocacy
- Distribute: Make their heritage available to users wherever they are, whenever they want it
- Engage: Cultivate new ways for users to participate in their cultural heritage

Europeana concentrates on aggregating metadata from cultural heritage institutions, directly or through regional or domain-specific aggregators, the majority of which is again supported by the European Union or national funding schemes. The metadata that these cultural heritage institutions share with Europeana are made available under a Creative Commons CC Zero Public Domain Dedication licence enabling the metadata to be used, and reused by anyone for any purpose.

The collection at Europeana includes metadata of about 30 million items, from every domain of cultural heritage, from all EU Member States plus additional countries. The material types with the largest volume in the collection are images

[83] http://pro.europeana.eu/documents/900548/6e60b1c8-bd7d-4e7b-86bb-cf421ef09341

and texts, but there are also smaller amounts of sounds, videos and 3D objects present in the collection. The portal interface is available in 31 languages.

Providing access to the metadata is primarily through the portal at Europeana.eu, and through the Europeana API (http://pro.europeana.eu/api). This API is a web service that provides remote access to the metadata in Europeana. It allows the building of applications, websites and mash-ups that include a customised view of Europeana content. The API user can decide how much information is shown, where, and in what format. Access to the API requires prior registration to obtain an API key. The API uses the standard technology of REST calls over HTTP. Responses are returned in the JSON format.

In the Strategic Plan 2011-2015 (http://pro.europeana.eu/c/document_library/get_file?uuid=c4f19464-7504-44db-ac1e-3ddb78c922d7&groupId=10602), the strategy towards Linked Data is formulated as follows (p.12):

> ***Our aim is to accumulate digitised content, standardise the data that describes it, apply linked data techniques to enrich it and promote persistent identifiers to locate it in the long-term.***

And under the heading of "Improve the quality of the metadata" on p.13:

> ***The full potential of interoperability will be unlocked by the implementation in 2011 of the Europeana Data Model [EDM], a new way of structuring data. EDM will enable the use of Semantic Web technology, support Linked Open Data, maintain more domain-specific rich information and allow digital objects from providers to be shown alongside authoritative and curated information from other domains. The change will benefit not only Europeana but also our providers, who will be able to use the enriched data to upgrade services to their own users.***

Furthermore, on p.15, under the heading of "Strengthen Europeana's advocacy role":

> ***We actively advocate across a range of topics that contribute to sustainable access, including open business models, improved access through Linked Data applications, the importance of persistent identifiers, the need for better data, the removal of barriers to access, increased user participation and the responsible reuse of content.***

Linked Data is only a small and currently experimental part of their activities.

An experimental SPARQL endpoint to http://data.europeana.eu/ is hosted by Ontotext AD, http://europeana.ontotext.com/. The data is described using the Resource Description Framework (RDF) format, and structured using the Europeana Data Model (EDM). The technical details of the structure of the data are provided at http://pro.europeana.eu/tech-details.

All metadata is provided free of charge. Production of Linked Data is paid for from general project resources, and supported by activities of network partners, the majority of which are also co-funded by the European Commission or national public funding schemes.

All metadata is provided under the CC Zero 1.0 Public Domain Dedication (http://creativecommons.org/publicdomain/zero/1.0/).

The market for the Linked Data from Europeana encompasses all players in the field of cultural heritage, including individual users, cultural heritage institutions themselves, not-for-profit and commercial reusers.

II.4.2 Collected metrics

Usage

Website statistics for http://data.europeana.eu are available at http://eumunin.isti.cnr.it/stat/.

Revenue and other sources of income

Europeana is currently fully funded from public sources.

Cost

The cost of Linked Data provision is not separately accounted.

Benefits

There are currently 807 datasets (OAI-PMH sets as decided by the providers) with descriptions of approximately 20 million objects (see: http://data.europeana.eu/download/2.0/datasets/rdf/) available in the Linked Data pilot.

While the provision of Linked Data does not have any benefits as there is no reuse yet, the provision of Linked Data URIs by Europeana's data providers leads to a reduction in resources necessary for ingestion and enrichment of metadata. (See answers to question 4).

II.4.3 Interviews: Jill Cousins, Antoine Isaac

Interview date	17/7/13 & 23/7/13
Interviewee	Jill Cousins, Executive Director Europeana Foundation Antoine Isaac Scientific Coordinator, Europeana
Interviewer	Makx Dekkers

Q: How does the supply of Linked Data relate to the public task of your organisation?

A: Linked Open Data promises to be a tool enabling increased interoperability across metadata from various domains to facilitate discovery and reuse which is the central objective of Europeana. Because of its central position in the landscape of cultural heritage information, Europeana has an important role to play in making sure that there is no market failure around the provision of data related to cultural heritage, and Linked Open Data as a global standardised approach could play an important role in supporting that objective.

Q: Does your organisation have a "business case" document that motivates your investment in supplying Linked Data? Can you share it with us?

The Strategic Plan 2011-2015[84] outlines some of the strategic perspectives on Linked Data but there is no separate business plan. Currently, the activities around Linked Data are in an experimental stage.

A: What are your future plans? Do you plan to expand or abolish the supply of Linked Data?

One of the questions for European is what the relation between Linked Data and Google's schema.org is and how this will develop in the future. Europeana does not want to treat Linked Data as a black box but needs to understand the risks and benefits of possible future scenarios.

Q: Did the supply of Linked Data give rise to new opportunities for flexible data integration within your organisation and with partner organisations? Increase in data quality? New services? Cost reductions?

A: For the time being, provision of Linked Data by Europeana does not provide new opportunities, but mainly adds cost. However, the theory behind Linked Data does promise those opportunities but it is not easy to make it work and it takes more time than expected. There is also the issue that not all of the content providers to Europeana are on an equal level from a technological perspective so there is also a need to explain the approach and the benefits. Reuse requires effort on the side of the reuser; just providing an API is not enough to get data reused. External usage also requires knowledge on the part of the reuser about how the data is organised.

On the other hand, Europeana is also a reuser of Linked Data that is provided by other organisations; for example they use GeoNames and GEMET URIs in the data, and some of their data providers contribute data that contains URIs (GeoNames and locally maintained SKOS-based thesauri). Such URIs do make the job of ingesting and enriching metadata easier and therefore benefits Europeana by reducing the time needed to process such contributions. Europeana can harvest any SKOS-based data.

Q: What are the main enablers / inhibitors for Linked Data to deliver value for its reusers?

A: There are real benefits in the Linked Data approach as opposed to the more closed XML-based approach that Europeana has been based on until now. In comparison with the use of the Europeana Semantic Elements schema, the Europeana Data Model allows more expressiveness (relationships between objects and parts of object, structures, rights). However, currently the cost outweighs the benefits. There is a lot of potential but that potential still needs to be realised.

One thing that is important in the communications between Europeana and its data providers is that Linked Data is changing people's approaches to data modelling, away from closed-world approaches that work for a particular project or partnership towards a more open view.

[84] http://pro.europeana.eu/c/document_library/get_file?uuid=c4f19464-7504-44db-ac1e-3ddb78c922d7&groupId=10602

Q: Do you supply service level statements for Linked Data?

A: As the Linked Data activities are currently in an experimental phase, no guarantees are given. In the future, when production services are provided, this needs to be considered.

Q: Does your organisation have a URI policy? Can we have a copy?

A: The URI policy is described in the ISA study on Persistent URIs[85].

Q: Can you describe the Linked Data infrastructure of your organisation?

A: The infrastructure is described in several documents linked from a dedicated Linked Open Data page[86] on the Europeana PRO Website.

A recent blog post at LODLAM[87] gives an overview of what Europeana does with Linked Open Data.

Q: Which skills and competencies did you need in order to supply Linked Data? Do you have them in-house or contracted?

A: The activities are partly done in-house, and partly by other projects around Europeana.

Q: Which organisations are key partners in the supply of Linked Data?

A: Some of the larger national libraries (UK, France, Germany) provide Linked Data already. An important partner organisation is TEL (The European Library)[88] that is now also in the process to release LOD datasets which was one of the requirements from the research libraries that are now part of TEL. Contacts with Google are on-going. The library domain in particular is becoming very active in this space.

Q: Which activities do you carry out to supply Linked Data?

A: Currently, the Linked Data pilot gets its data from a conversion of the existing ESE-based data which involves some manual intervention. About 20 million objects (of the total of 27 million) have been loaded in the Linked Data pilot.

There are plans to upgrade the Europeana API[89] which now only delivers JSON to also deliver RDF/XML and possibly JSON-LD.

Q: Which investments has your organisation made to enable it to supply Linked Data?

A: There is no separate costing of Linked Data activities. The experiment is part of the R&D work in the current Europeana V2.0 project, and is part of a wider research agenda that also includes multilingualism, data modelling and semantic enrichment.

[85] https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris
[86] http://pro.europeana.eu/linked-open-data
[87] http://lodlam.net/2013/06/18/what-is-europeana-doing-with-sw-and-lod/
[88] http://www.theeuropeanlibrary.org/tel4/
[89] http://pro.europeana.eu/web/guest/api

Q: How many FTEs in your organisation are involved in provisioning the supply of Linked Data?

A: Overall about 1.5 to 2 FTE, spread out over a small team (scientific coordinator, interoperability specialist, developers). Apart from the pilot, this also includes the data modelling and the supply of schema.org data to Google.

Q: Which costs have you incurred to publish Linked Data, maintain, and promote it? What are the trends?

A: The cost of publication of Linked Data is not separately accounted. However, when the API is able to deliver RDF/XML, there is no additional cost to providing Linked Data.

Q: Who are the main users of your Linked Data services? Is Linked Data only used by external customers or also for back-office exchange (e.g. with other agencies)?

A: There are no regular users of Linked Data yet. To promote use, it may be necessary to provide usage information and guidelines, in the way the BnF does (e.g. http://data.bnf.fr/about). The potential users include app developers, but they need first to understand what data they can get. Also, organisations that provide metadata to Europeana can benefit from receiving their (enriched) data back, but there are not many cultural heritage organisations that have the technical means to handle Linked Data yet.

Q: How often is Linked Data used? What are the trends?

A: Europeana already receives Linked Data from its providers. The amount of this EDM-based data is increasing.

Q: How is the provisioning of Linked Data funded?

A: The work is funded as part of the project budget.

Q: Which pricing mechanism or other source of income exists?

A: The Linked Data is provided for free. Charging for access to the data would be a step backwards from the openness that is the very objective of Europeana. However, in the future premium services or freemium models may be considered.

Q: Under which licence is Linked Data made available for reuse? Can we have a copy?

A: Europeana's metadata is made available for access and reuse under CC Zero Public Domain Dedication and Linked Data will be no exception.

Q: Which channels are predominantly used to consult Linked Data: Web API? Web site? (mobile) App? Data market?

A: Access to Europeana's Linked Data will be through the upgraded API. In addition, the Europeana Creative project is investigating how to bring the Ontotext's SPARQL endpoint for data.europeana.eu a bit closer to production status.

In the future, main access channels are expected to be:

1. traditional XML and JSON API, with RDF/XML and JSON-LD quite advanced in implementation

2. RDF data dumps (which will happen to be XML data dumps)
3. Linked Data access to individual object data (content negotiation and RDF API output)

Q: Do you have a branding strategy for Linked Data?

A: Not at the moment, but given the importance of Europeana in the cultural heritage sector, this may change in the future when Linked Data services move out of the experimental stage.

Q: Do you invest in advertisement for Linked Data?

A: The Linked Data activities have been regularly included in presentations but in decreasing levels as the data in data.europeana.eu gets outdated. A presentation of the (re)use of Linked Open Data and Semantic Web principles will be given at SWIB13[90], and Europeana sponsored the LODLAM Summit 2013[91]. The Linked Open Data videos at Vimeo[92] (in five languages) and YouTube[93] are still accessed frequently.

Q: Do you provide / make use of any mechanisms on user feedback or evaluation?

A: During the pilot phase, feedback was received by e-mail.

II.4.4 Europeana as a reuser

Europeana receives some of its data with GEMET URIs. In some cases, the provider also includes data for the concepts next to the data for the cultural objects, but sometimes it's not complete. In such cases, Europeana tries to dereference the URI by HTTP GET through a plug-in in the ingestion manager tool. If data is received for the URI, it is then mapped to the Europeana Data Model.

Such URIs do make the job of ingesting and enriching metadata easier and therefore benefits Europeana by reducing the time needed to process such contributions. Europeana can harvest any SKOS-based data.

[90] http://swib.org/swib13/
[91] http://summit2013.lodlam.net/
[92] http://vimeo.com/album/2072014/
[93] http://www.youtube.com/watch?v=uju4wT9uBIA

II.5 EU – European Environment Agency

II.5.1 Desk research

Source: http://www.eea.europa.eu/about-us

The European Environment Agency (EEA) is an agency of the European Union. The regulation establishing the EEA was adopted by the European Union in 1990. It came into force in late 1993 immediately after the decision was taken to locate the EEA in Copenhagen. Work started in earnest in 1994. The regulation also established the European environment information and observation network (Eionet)[94] which has 33 member countries, the 28 European Union Member States plus Iceland, Liechtenstein, Norway, Switzerland and Turkey.

EEA's mandate is:

- To help the Community and member countries make informed decisions about improving the environment, integrating environmental considerations into economic policies and moving towards sustainability
- To coordinate the European environment information and observation network (Eionet)

Source: http://www.eea.europa.eu/about-us/what/seis-initiatives/seis-initiatives

The Shared Environmental Information System (SEIS), an initiative led by EEA, aims to make information available for human consumption but also to make the data available for machine-to-machine communication via standard APIs and open data formats. Implementing and supporting SEIS-friendly technology is one of the core activities of the Agency.

For these reasons the Agency main portal has been extended with semantic web technology also known as Linked Data. The same technology is increasingly used to implement SEIS within Eionet and Reportnet, Eionet's infrastructure for supporting and improving data and information flows

Everything on EEA's website is harvestable via external systems and linked data spiders, so the data and information can be easily reused, integrated and re-distributed by to a wider network of users. As a practical example, organisations are now able to easily exchange their catalogues of datasets creating more complete federated dataset catalogues, also known as Open Data Catalogues. The technology makes it effortless for the Agency to contribute to the European Commission Open Data Portal.

Source: http://semantic.eea.europa.eu/

EEA's Semantic Data Service is an object-oriented search engine (the Content Registry) that enables searching for the content of data in Eionet. Being object-oriented means it understands what e.g. a measuring station is and can show what measurements it has made. Not all of the Eionet services are included, only those that have been specified by the administrators of this site.

[94] Eionet, the European environment information and observation network. http://www.eionet.europa.eu/

Source: http://semantic.eea.europa.eu/documentation/generalprinciples

The Content Registry harvests structured data from Member States (in XML, Excel) and Eurostat (in various formats) at regular intervals and creates an RDF triple store from the harvested data either using the data directly if it is in RDF or converting it to RDF using schema mappings and code list.

The purpose is to facilitate analysis of the data both from a quality assessment perspective and to produce European datasets and reports.

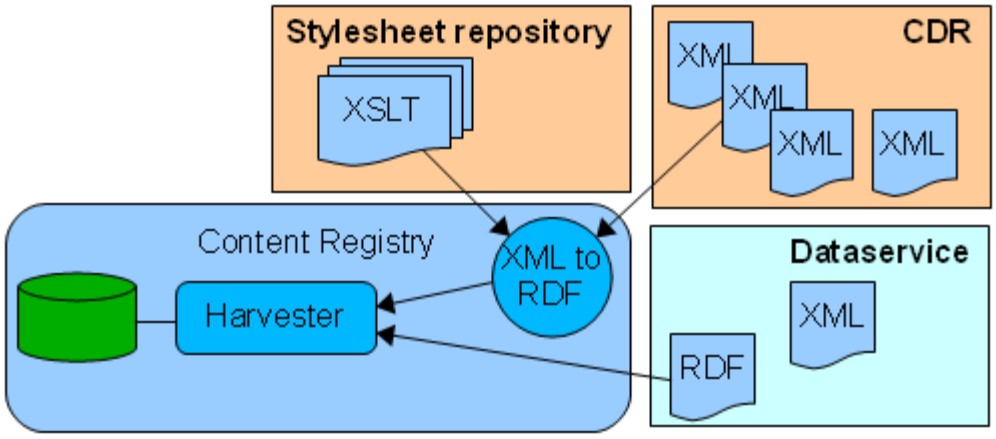


Figure 9: EEA Ingestion architecture

II.5.2 Collected metrics

Usage

Exact monitoring data about usage of the linked data is not available.

Revenue and other sources of income

Funding of all data-related activities at EEA is from public sources. There are no other sources of income.

Cost

The cost of the provision of Linked Data cannot be separately identified. Investments would have been made in any case. The Linked Data production and publication is considered an evolutionary add-on to EEA's normal activities.

Benefits

The Linked Data activities are mainly envisaged to enhance the data provision by EEA. The availability of the Linked Data to external users is a side-effect.

EEA Linked Data files contain an estimated 500 million triples.

II.5.3 Interview Stefan Jensen and Søren Roug

Interview date	8 August 2013

Interviewee	Stefan Jensen – Head of group, SEIS and spatial data infrastructure (SDI) Søren Roug – Head of group, IT networking and public systems
Interviewer	Makx Dekkers

Q: How does the supply of Linked Data relate to the public task of your organisation?

A: We supply the data in LOD format to promote a more cost-efficient way to share data between organisations. Some orgs create web service APIs. We also use LOD. The use of Linked Data intends to make internal processes more efficient as part of the general mandate to make environment-related information available.

Q: Does your organisation have a "business case" document that motivates your investment in supplying Linked Data? Can you share it with us?

A: Not really, it is an architectural principle that we employ to integrate our internal systems. We started with RSS and evolved into RDF. That the database dumps in RDF format can also be used by outsiders is a side-effect.

Q: What are your future plans? Do you plan to expand or abolish the supply of Linked Data?

A: We will provide more linked data. In a strategic sense, applications based on Linked Data become more and more central to EEA's operations.

Q: Did the supply of Linked Data give rise to new opportunities for flexible data integration within your organisation and with partner organisations? Increase in data quality? New services? Cost reductions?

A: The flexible data integration was the main motivation. At this moment, EEA needs to keep various routes open for content ingestion as the content owners are not on the same level technologically.

Q: What are the main enablers / inhibitors for Linked Data to deliver value for its reusers?

A: There is a lack of tools. We had to develop RDF exporters in several programming languages and a maintenance system for vocabularies. Furthermore, reuse of Linked Data faces hurdles as it requires a relatively advanced technological level on the part of the reusers. More work is needed in that area.

Q: Do you supply service level statements for Linked Data?

A: No. Reusers of Linked Data will complain to the data provider if things go wrong.

Q: Does your organisation have a URI policy? Can we have a copy?

A: Avoid question marks in the URI. The approach is based on W3C's Note "Cool URIs for the Semantic Web"[95]. The URIs tend to follow patterns used in the database.

Examples:

[95] W3C. Cool URIs for the Semantic Web. http://www.w3.org/TR/cooluris/

- http://www.eionet.europa.eu/gemet/concept/24
- http://www.eea.europa.eu/data-and-maps/indicators/eea32-heavy-metal-hm-emissions-1/assessment-2/
- http://www.eea.europa.eu/data-and-maps/figures/eu27-air-pollutant-emission-trends-3/

It certainly is our intention to have persistent URIs – not only for Linked Data but for HTML pages as well. We're still building tools and it isn't always possible to adhere to that ideal.

Q: Can you describe the Linked Data infrastructure of your organisation?

A: We have a system called Reportnet to handle deliveries of data in XML from national institutions. It handles the definition of schemas, style sheets and lookup tables as well. To be able to verify if a code was known already in our system we could either implement a REST API in the applicable database system or copy the data to a triple store and use SPARQL. By converting the deliveries to RDF and loading them into the triple store we can do analysis on the data as soon as we receive it. It only takes an XSL style sheet.

Q: Which skills and competencies did you need in order to supply Linked Data? Do you have them in-house or contracted?

A: We learned them as we evolved. We didn't contract a company that already had Linked Data experience. EEA received some training from the EU funded LOD project with a two-day course.

Q: Which organisations are key partners in the supply of Linked Data?

A: We are doing some capacity-building with our partners, but none of them have yet achieved a sufficient level of maturity yet.

Q: Which activities do you carry out to supply Linked Data?

A: It is taken into consideration whenever we renovate data flows or websites.

Q: Which investments has your organisation made to enable it to supply Linked Data?

A: We have made some software investments in tools, but we would have done them anyway as XML or Web-Service tools if we didn't use the Linked Data principles

Q: How many FTEs in your organisation are involved in provisioning the supply of Linked Data?

A: We consider it a side-effect of our normal activities.

Q: Which costs have you incurred to publish Linked Data, maintain, and promote it? What are the trends?

A: We haven't calculated this.

Q: Who are the main users of your Linked Data services? Is Linked Data only used by external customers or also for back-office exchange (e.g. with other agencies)?

A: We are the main users ourselves. We do not yet do statistics about external users.

Q: How often is Linked Data used? What are the trends?

A: We see it as a way to be able to handle heterogeneous data more efficiently. We are therefore looking at moving as much of our data as possible into triple stores.

Q: How is the provisioning of Linked Data funded?

A: All internal funding as part of normal software development operations.

Q: Which pricing mechanism or other source of income exists?

A: We don't charge for data.

Q: Under which licence is Linked Data made available for reuse? Can we have a copy?

A: Normally CC-BY, occasionally Open Database License.

Q: Which channels are predominantly used to consult Linked Data: Web API? Web site? (mobile) App? Data market?

A: The entire EEA website content is available via a SPARQL endpoint. RDF dumps are also available.

Q: Do you have a branding strategy for Linked Data?

A: No.

Q: Do you invest in advertisement for Linked Data?

A: No.

Q: Do you provide / make use of any mechanisms on user feedback or evaluation?

A: Feedback is given by content partners in the Eionet Forum, by INSPIRE partners and other cooperation partners including the European Commission DG CONNECT, the Publications Office of the EU and the US Environment Protection Agency.

II.6 EU – Publications Office of the European Union

II.6.1 Desk research

Source: http://publications.europa.eu/index_en.htm

The Publications Office of the European Union (Publications Office, PO) is an interinstitutional office whose task is to publish the publications of the institutions of the European Union on the basis of the Decision of the European Parliament, the Council, the Commission, the Court of Justice, the Court of Auditors, the European Economic and Social Committee and the Committee of the Regions of 26 June 2009 on the organisation and operation of the Publications Office of the European Union (2009/496/EC, Euratom[96]).

The Publications Office publishes the daily Official Journal of the European Union in 23 languages (24 when Irish is required) and produces (or co-produces) publicity for EU initiatives and activities. It publishes or co-publishes the publications in the context of the communication activities of the institutions.

Moreover, the Publications Office offers a number of online services giving free access to information on EU law (EUR-Lex[97]), EU publications (EU Bookshop[98]), public procurement (TED[99]), and EU research and development (CORDIS[100]).

Source: http://publications.europa.eu/mdr/

The Metadata Registry registers and maintains definition data (metadata elements, named authority lists, schemas, etc.) used by the different European Institutions involved in the legal decision making process gathered in the Interinstitutional Metadata Maintenance Committee (IMMC) and by the Publications Office of the EU in its production and dissemination process.

II.6.2 Collected metrics

Metric
Usage • **Number of de-referenced URIs / queries:** N/A • **Number of governmental reusers of LOGD:** N/A • **Number of commercial reusers of LOGD:** N/A
Revenue and other sources of income • LOGD is funded by public funding.
Cost • **Development and maintenance cost**: N/A • **Promotion costs per year**: N/A

[96] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:168:0041:01:EN:HTML
[97] http://publications.europa.eu/eur_lex/index_en.htm
[98] http://publications.europa.eu/eu_bookshop/index_en.htm
[99] http://publications.europa.eu/tenders/index_en.htm
[100] http://publications.europa.eu/cordis/index_en.htm

Metric
Benefits • **Number of linked datasets (outgoing and incoming links):** N/A • **Number of derived applications:** N/A

II.6.3 Interview

Interview date	18 September 2013
Interviewees	Peter Schmitz, Head of Unit, Postproduction Reception, Validation and Cellar Management Unit, Direction Core Business Services and Marc Küster – answers provided in writing
Interviewer	Makx Dekkers

Q: How does the supply of Linked Data relate to the public task of your organisation?

A: The Publications Office has implemented the backbone for its Internet dissemination based on semantic technology, the so-called CELLAR. The CELLAR is the Office's centralized common repository for all of its dissemination-related content and metadata. The metadata part is implemented by an RDF store based on a dedicated ontology called CDM (Common Data Model). In consequence, all metadata, which is published by the Publications Office in the scope of its public tasks, is at the moment of publication also available as Linked Data.

Q: Does your organisation have a "business case" document that motivates your investment in supplying Linked Data? Can you share it with us?

A: See answer on the previous point.

Q: What are your future plans? Do you plan to expand or abolish the supply of Linked Data?

The Publications Office will expand the supply of Linked Data.

Q: Did the supply of Linked Data give rise to new opportunities for flexible data integration within your organisation and with partner organisations? Increase in data quality? New services? Cost reductions?

A: The supply of Linked Data is a by-product of the choice for the technical implementation of the Publications Office's new dissemination backbone.

The declarative approach used for the implementation of the metadata repository has increased in a significant manner the integration of new types of documents. Control based on the ontology has increased the data quality. The new infrastructure enables the creation of new services in particular in the domain of "reuse". Cost reductions are expected as soon as the initial implementation will be finished.

Q: What are the main enablers / inhibitors for Linked Data to deliver value for its reusers?

A: Enablers:

- CELLAR
- Open Data Portal

Inhibitors:

- None

Q: Do you supply service level statements for Linked Data?

A: Not yet.

Q: Does your organisation have a URI policy? Can we have a copy?

A: Yes. URIs of resources in CELLAR follow the pattern `http://publications.europa.eu/resource/{ps-name}/{ps-id}` where `ps-name` identifies the production system and `ps-id` is the unique identifier for the resource in the context of the production system. These URIs will be persistent.

Q: Can you describe the Linked Data infrastructure of your organisation?

A: See the answer on the supply of Linked Data related to the public task above.

Q: Which skills and competencies did you need in order to supply Linked Data? Do you have them in-house or contracted?

A: Expertise in semantic technology. Skills and competencies are both available in-house and through external contractors

Q: Which organisations are key partners in the supply of Linked Data?

A: Professional reusers (legal information services, information brokers…), public entities of the EU member states.

Q: Which activities do you carry out to supply Linked Data?

A: Publication of official documents of the EU Institutions.

Q: Which investments has your organisation made to enable it to supply Linked Data?

A: Not available.

Q: How many FTEs in your organisation are involved in provisioning the supply of Linked Data?

A: Not available.

Q: Which costs have you incurred to publish Linked Data, maintain, and promote it? What are the trends?

A: Not available.

Q: Who are the main users of your Linked Data services? Is Linked Data only used by external customers or also for back-office exchange (e.g. with other agencies)?

A: These are the same organisations as the ones mentioned in the answer on key partners above.

Q: How often is Linked Data used? What are the trends?

A: The system is available for the public since 1st of July 2013; data loading is not yet completely finished. In consequence, it is too early to be able to answer this question.

Q: How is the provisioning of Linked Data funded?

A: EU budget.

Q: Which pricing mechanism or other source of income exists?

A: Linked data is available free of charge.

Q: Under which licence is Linked Data made available for reuse? Can we have a copy?

A: Commission Decision of 12 December 2011 on the reuse of Commission documents (2011/833/EU), http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:330:0039:0042:EN:PDF, Europa copyright notice (Commission) http://ec.europa.eu/geninfo/legal_notices_en.htm.

Q. Which channels are predominantly used to consult Linked Data: Web API? Web site? (mobile) App? Data market?

A: Web API, Web site.

Q: Do you have a branding strategy for Linked Data?

A: No. Provenance and authenticity are the major strategic domains for the Publications Office.

Q: Do you invest in advertisement for Linked Data?

A: No.

Q: Do you provide / make use of any mechanisms on user feedback or evaluation?

A: Foreseen in the scope of the Publications Office's Common Portal (under development).

II.7 IT – Agenzia per l'Italia Digitale

II.7.1 Desk research

Among other functions, the Agenzia per l'Italia Digitale[101], AgID, is responsible for the design, implementation and management of the Sistema Pubblico di Connettività (SPC) - Public Connectivity System. SPC. Under the governance of its Coordination Commission[102], SPC is the Italian interoperability framework and consists, among the others, of a set of technical regulations and provisions intended to federate the ICT infrastructure of public administrations. The aim is to integrate services through rules and shared services, allowing efficiency savings and improvements in end-user-centred services, avoiding repeated requests for data from administrations, and duplication of information.

An important SPC shared service is represented by SPCData that currently allows for the provision of three data sets as linked open data[103]:

- the index of public administrations;
- SPC contracts;
- data classifications.

The simplest of the data sets is the data classifications (Classificazione IPA delle PA) which simply declares a set of SKOS concepts, in particular central and local administrations. These definitions are based on ESA95[104] and are used as the value of the org:classification property for each of the public administrations in the index.

The full index of public administrations includes information on all Italian public administrations, their organisation structures and the services offered. This can be downloaded as individual files or as a complete set in one. Data is available as triples or in CSV.

The contract data is interesting. It shows the value of public administration procurement contracts broken down by region (or national government). As an example of what's possible, the following SPARQL query returns the top 10 highest value procurement contracts and the region where they were placed. The national government comes out as the biggest spender, understandably enough, with the Lazio Region as another 'big spender.'

```
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix pc:  <http://purl.org/procurement/public-contracts#>
prefix gr:  <http://purl.org/goodrelations/v1#>
prefix sc:  <http://spcdata.digitpa.gov.it/ServizioContratto/>
prefix pvc: <http://spcdata.digitpa.gov.it/PrezzoVoceContratto/>
prefix foaf: <http://xmlns.com/foaf/0.1/>
```

[101] http://www.agid.gov.it/

[102] http://www.agid.gov.it/spc/commissione-coordinamento

[103] http://spcdata.digitpa.gov.it/

[104] http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/Annexes/naga_a_esms_an1.pdf

```
prefix gn: <http://www.geonames.org/ontology#>

SELECT ?amt ?name ?placeName WHERE {

  ?a rdf:type gr:PriceSpecification .
  ?a gr:hasCurrencyValue ?amt .
  ?offering gr:hasPriceSpecification ?a .
  ?offering gr:includes ?service .
  ?service gr:name ?name .
  ?contract pc:item ?offering .
  ?contract pc:contractingAuthority ?admin .
  ?group foaf:member ?admin .
  ?group gn:locatedIn ?place .
  ?place rdfs:label ?placeName
}
ORDER By DESC(?amt)
LIMIT 10
```

Figure 10 Example SPARQL query suitable for the SPC contract data

amt	name	placeName
42838699.9	System management	Italia
42741160.2	Gestione posti di lavoro	Italia
15478487.46	Progettazione e realizzazione di siti web	Italia
8012048.04	Supporto tecnico alle attivita' di tipo redazionale e gestione dei contenuti di un sito web	Italia
7293510	System management	Italia
5947524.4	Accesso ad applicazioni in modalita' web	Italia
5779914.68	Hosting di siti web	Italia
5439980.22	Accesso ad applicazioni in modalita' web	Lazio
5439980.22	Accesso ad applicazioni in modalita' web	Lazio
4433811.58	Gestione posti di lavoro	Italia

Table 12 The results executing the query in Figure 10

II.7.2 Interview

Interview date	24/7/13
Interviewees	Giorgia Lodi & Antonio Maccioni
Interviewer	Phil Archer

Dr Giorgia Lodi and Antonio Maccioni comprise the (entire) linked data staff at AgID and were directly responsible for creating the SPC linked data (SPCData) described above. At the time of our discussion (24 July 2013) they are awaiting approval for their proposed agenda for future work. AgID is responsible for defining technical rules for the interoperability of the Italian base registries, as defined by the national master law of innovation named CAD – Codice dell'Amministrazione Digitale. It's these that, it is hoped, will all be available as linked data in the short to medium term.

It's also worth noting that an important set of guidelines concerning Italian public sector linked data, written by Giorgia and Antonio with the support of a variety of public administrations both central and local, was published at the beginning of August 2013.

To date, the data described in the previous section is the only linked data released by AgID. The proposed agenda includes the addition of further data sets by the end of 2013. The proposed additional data includes, for example, a classification of Italian public administrations according to COFOG - something that was highlighted in a recent ISA Programme pilot study[105].

The data on which SPCData depends - i.e. the data that identifies different ministries, their locations and contact points, is gathered by a commercial company that provides a number of infrastructure services to the Italian government. This data is available for free as a set of downloadable CSV files and can be obtained directly from the relational database in which it is held. AgID used the latter method, employing D2RQ[106] to obtain the original data for SPCData. The plan for the future is to regularly take copies of the CSV dumps and process those in a semi-automated method.

Q: So how does the commercial company get the data originally?

A: The public administrations have to send it to them

Q: Couldn't they send it to you directly and cut out the middle man? That seems like an obvious cost saving.

A: They do more than that including services at the desktop, it's a big contract, so it's hard to pick off individual bits. It's the way it is. Moreover, there are checks to be run as well - whether a given PA should be in the registry or not, for example.

Q: So how does your LD work relate to the public task?

A: Our agency promotes interoperability; we coordinate open data actions in Italy. We wanted to understand the standards and see how we can provide interoperable data, particularly for our base registers. We have a duty to open these which started in 2012 and was strengthened at the end of that year.

Q: Does your organisation have a "business case" document that motivates your investment in supplying Linked Data?

A: Not as such. The business case is as described. Our manager was already convinced by linked data. It wasn't a conscious decision to use LD - that's the technology we need to achieve our goal and it's the goal that forms the business case.

[105] http://joinup.ec.europa.eu/asset/core_business/document/organization-ontology-pilot-linking-public-sectors-organisational-data

[106] http://d2rq.org/

Q: Did the supply of Linked Data give rise to new opportunities for flexible data integration within your organisation and with partner organisations? Increase in data quality? New services? Cost reductions?

A: Other people have built services on top of the data. For example, an address book for public administrations (PocketPEC)[107]. That kind of application could have been built before but now it's a lot easier. The makers of PocketPEC wanted us to add in lat and long coordinates so they could put the different offices on a map. We have thought about this but haven't done it yet. (Editorial note - efforts to contact the makers of Pocket PEC have been unsuccessful).

Antonio, in collaboration with a friend of his, also built an Android App named ComunicaPA as a personal project:

ComunicaPA[108] allows citizens to easily communicate with the Italian Public Administration. ComunicaPA is one of the first existing attempts to join Semantic Web with the mobile world.

By nature, mobile applications cannot embed large datasets. In spite of being a database-less and server-less app, ComunicaPA shows how mobile apps can scale up over big datasets just by exploiting Semantic Web standards.

From a citizen point of view, ComunicaPA retrieves Linked Data from SPCData and then, integrating the internal mobile functionalities, is able to: send e-mails and certified e-mails (PEC in Italian), geolocalize on the map PA offices and call them, surf official websites, etc.

ComunicaPA takes also advantage of the fact that data are official and maintained by third parties. It is like the data management part is fully outsourced.

As a result, it highlights the benefits and positive impacts in the development of mobile applications by using available Linked Data sources on the Web.

ComunicaPA *does* show the location of public administration offices on a Google map but this is done by text search, not by finding the lat and long of the office.

We know of a company on France that uses our data too (in another application designed to make it easy to contact Italian Public Administrations). We weren't aware of them until one day our SPARQL endpoint went offline - they got in touch very quickly.

[107] http://www.pocketpec.it/

[108] http://play.google.com/store/apps/details?id=org.source.comunicapa

We don't have a service level commitments at this point - there's not sufficient demand yet.

Q: Is the SPCData for you or others?

A: It's part of the national agenda. One of the things we've proposed is that each PA should create their own organisation data (using the ORG Ontology) and then link it to SPCData. We've decided to limit the scope just to national government for now. We'll provide some support for PAs to do this - and the guidelines previously mentioned should help too. The guidelines include advice on choosing vocabularies, worked examples and so on. We're hoping to do some eLearning courses; again, this is all part of our obligation. That said, our experience is that creating organisation descriptions isn't *that* difficult.

Q: Do you have anything like the UK Gov Camp[109] and Tea Camp[110] events in Italy?

A: Yes. We had similar events in several Italian cities as part of Open Data Day (a worldwide event). There were lots of meetings and conversations including talking about SPCData. So yes, there is a small community within the Italian public sector around this. Of course there's another community that says that we should only publish raw data and that if anyone wants to convert it to linked data then they can.

Q: What are the main enablers / inhibitors for Linked Data to deliver value for its reusers?

A: The biggest obstacle is the culture. Semantic interoperability is still seen as something only doable by specialists - please don't bother me with it. Other PAs include people who are aware that LD can be seen as high quality data - this is encouraging - but they get scared when you start talking about vocabularies and ontologies.

The biggest enablers are the communities. The Geo data community is very active and helpful.

Q: Does your organisation have a URI policy?

A: Our technical guidelines include a section on persistent URIs. We contributed to and refer to the ISA Programme work on this[111].

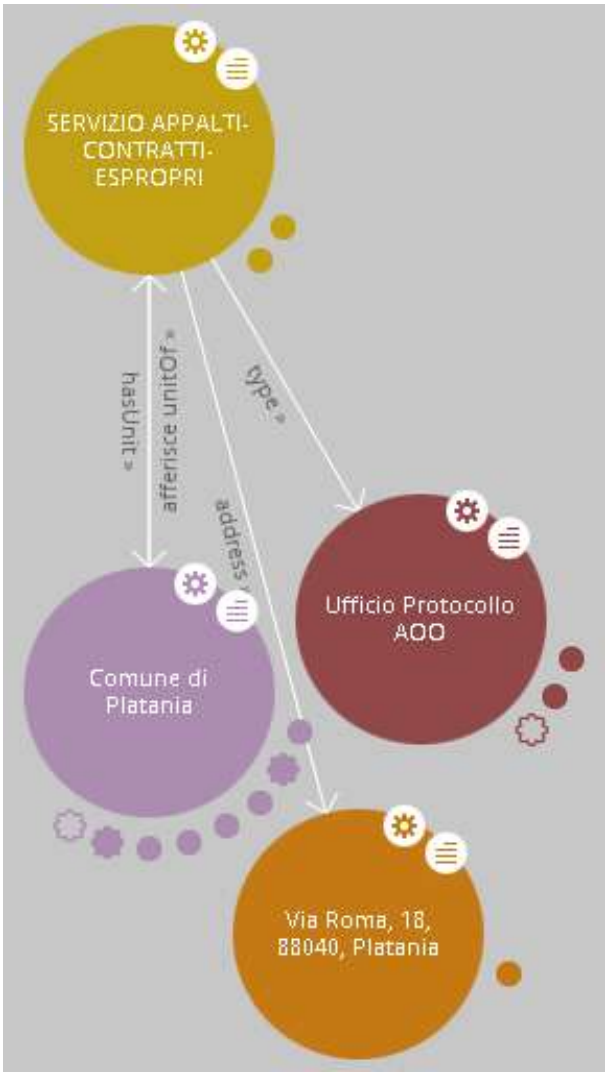
Q: Which skills and competencies did you need in order to supply Linked Data? Do you have them in-house or contracted?

A: GL - I knew about semantic technologies but hadn't used it. AM had more experience. We had some initial help from CNR to help us get up and running. The PAs and CNR are our key partners. The original modelling was all done on bits of paper with strange-looking graphs drawn on them.

[109] http://www.ukgovcamp.com/

[110] http://teacamp.co.uk/

[111] http://joinup.ec.europa.eu/sites/default/files/D7.1.3%20-%20Study%20on%20persistent%20URIs_0.pdf



Q: Which investments has your organisation made to enable it to supply Linked Data?

A: It's not possible to identify specific costs for the linked data work - it's all done as part of our broader work. There are no extra costs because we use linked data. On the contrary, we only use open source (free) software such as OpenLink Virtuoso and Tomcat.

We hope that the new agenda will be accepted and we'll see the activity grow. We may need to try and find more people within the agency to help us but we don't see it as something separate - this is how we fulfil this particular government task. We don't have a plan to hire new members of staff specifically to work on linked data.

Figure 11 Static screenshot of an example of the LODLive visualisation of SPC contract data http://spcdata.digitpa.gov.it/lodlive/llive.html?http://spcdata.digitpa.gov.it/UfficioProtocolloAOO/c_g734-4

Q: Can you point me to any other users of your data?

A: We know that there are some government users although they tend to use the CSVs rather than our SPARQL endpoint for now. We use two applications within the SPCData. Go to http://spcdata.digitpa.gov.it/data.html and for instance search for, say, "Appalti", and you see a list of contracts. You can navigate them using Pubby or LodLive (Pubby is the tool used by DBpedia[112]); LodLive gives you an interactive visualisation of the data (Figure 11). LodLive is entirely JavaScript (JQuery) driven and sends XHttpRequests to our SPARQL endpoint. The company that built that is also the company the built the linked data project for the Italian Parliament[113].

Q: How often is Linked Data used? What are the trends?

A: We had a problem collecting our server statistics but we know that traffic to the Web site is increasing.

[112] http://spcdata.digitpa.gov.it/browse/page/UfficioProtocolloAOO/c_g734-4

[113] http://dati.camera.it/it/video/World-e-Parliament-Conference-2012-session-A4.html

Q: Italy is going through a particularly difficult time. How confident are you of the future of the agency?

A: Very: the agency enjoys cross-party support. The only thing is that progress might be slow - a lot of people like what we do and want to take credit by being in control of AgID.

Q: Under which licence is Linked Data made available for reuse?

A: CC-BY-SA although we're thinking about moving it to just CC-BY.

Q: Do you invest in advertisement for Linked Data?

A: We do some promotion - but there are only two of us remember! We'll be able to do more once the guidelines are published and the agenda is agreed. GL taught a masters course at Sapienza University of Rome last year which I guess could be seen as a promotional activity.

Q: Do you provide / make use of any mechanisms on user feedback or evaluation?

A: Just an e-mail address which gets used a little.

Q: What other data sets do you link to?

A: We link to DBpedia, GeoNames, FAO and the LOD published by municipality of Firenze (Florence) and Region of Piemonte.

Interview Summary

Unlike the UK, the provision of linked data is part of a strategy that has been centralised in a single service. There are regions that are also providing linked data but AgID is acting as a central source of skill and infrastructure - based on two people. Sensibly and necessarily, they are taking a step by step approach, beginning with basic information about Italian public administrations. The same agency is also responsible for other base registers so the is an inherent centralisation at play.

The linked data so far released lends itself to easy integration into use apps. The Pocket PEC application uses the data to provide secure e-mail exchange with public administrations - that's a clear use case and it's repeated in the ComunicaPA application built by Antonio and his friend. It is noteworthy that AgID received urgent e-mails from an unknown developer in a different country when its SPARQL endpoint went down.

II.7.3 Interview II Tiggit Software (LOGD user)

Interview date	26/08/13
Interviewees	Marco, Pocket PEC developer
Interviewer	Phil Archer (by e-mail)

Q: Can you give me a brief history of the application? What was your original motivation for developing PocketPEC?

In 2005, an official (legally usable) electronic delivery service, named certified electronic mail (in Italian "PEC: Posta Elettronica Certificata") was defined by CNIPA

(today, AgID). PEC is based on S/MIME standards. It is a traditional e-mail system where, for any outgoing mail, the sending PEC provider is in charge of creating a transport envelope, containing the original outgoing message and transaction information, and signing the envelope before forwarding it to the destination user. The sending and the receiving PEC providers, also, generate signed emails, towards the sending user, certifying message acceptance and delivery. Details on the PEC system are available in RFC6109. When the PEC standard was published we, at Tiggit Software, felt it was a good opportunity to build on our expertise and investment by improving our own mobile email tools.

In order to simplify PEC usage, we provide two software packages:

ThunderPEC: a Mozilla Thunderbird add-on for desktop computers

PocketPEC: a mobile app for Android and BlackBerry devices.

ThunderPEC and PocketPEC[114] implement the following features:

- Directly display the contents of the original message
- Verifies the digital signature of the PEC message
- View transaction data associated with a PEC message
- Simplify the configuration of the PEC mailbox
- Receipts aggregated view for outgoing messages
- Display the contents of digitally-signed attachments and related signatures
- **PEC address search on AgID Open Data repository**

Q: What is the target market? Is the application successful in your view? What feedback have you had? It's made available for free so what's the business case?

The PEC system is mainly used as a replacement of the registered letter postal service in the dialogue between public institutions, citizens, professionals and companies. In order to spread PEC usage in the dialogue between citizens and public administrations, every Italian citizen could get a free-of-charge PEC mailbox (www.postacertificata.gov.it). A PEC address is needed to start a new company or change related data in public registries. All professionals (i.e. lawyers) have to provide a PEC address to their professional orders.

So, the target market is wide and we got positive feedback from our customers/users as we succeed in our main objective: simplify PEC usage.

ThunderPEC is provided as free software (under LGPL license) and is available on Mozilla Add-ons site (https://addons.mozilla.org/it/thunderbird/) while PocketPEC uses a subscription scheme and is available on both Google Play and BlackBerry App World.

[114] www.pocketpec.it

Q: Am I right that you use the linked data provided at http://spcdata.digitpa.gov.it/ **?**

Both applications interact with AgID Open Data repository (spcdata.digitpa.gov.it), in order to provide PEC address search functionality to the end-user.

Figure 12 ***PocketPEC Search Public Administration window***

Figure 13 ***Using Search PA result for sending a PEC message***

Q: Can you talk a bit about the technologies you use... do you use the SPARQL endpoint? What is your opinion of linked data? Would you prefer the data to be available as a CSV download instead? Why?

Both applications use the SPARQL endpoint to perform searches and JSON data representation for results.

Linked data provides a mechanism to access to up-to-date information and to gather easily additional data. As any live data source, it is important to update the stored information to reflect changes as soon as possible (a PEC address change or a new phone number, for example). Otherwise, there will be no added value in comparison with a CSV download.

Q: Do you use other data sources as well as the SPC data? If so, does linked data help you aggregate the data?

Currently, no. As the AgID Open Data repository deals only with public administrations, we would like to add other PEC directory services related to professionals and companies.

Q: AgID is likely to publish more data in future - do you think you'll be able to use that? What data would you most like to see published by the Italian government?

In the past, we proposed to add GPS information in order to add a drive-me-to facility to our software. Looking at mobile users, it would be useful to add pointers

to public administration services (i.e. the web page for certificate request) in order to avoid searches and speed up the access to information request.

Q: What are your future plans in this area?

Currently, PocketPEC and ThunderPEC retrieve only the PEC address information. In next releases, both software packages will provide access to other information in order to integrate the possibility to call the public administration, to drive the user to the related street address and to add a new contact in the device address-book.

II.7.4 Interview Summary

This is a rare case of a commercial application built on linked open government data. The driver here is the specific e-mail system used and mandated by Italian law more than the availability of the data, although the availability of the data does make the mandated system much easier to use. The value in the data is that it is authoritative and up to date. The fact that it provided via a SPARQL endpoint is not a key factor here as there is one source for one type of data. The added value of LOGD will show itself when Tiggit Software is able to execute its plans for future development of the application.

II.8 UK – BBC

II.8.1 Desk research

The British Broadcasting Corporation (BBC) is the UK's Public Service Broadcaster and operates under a Royal Charter[115], funded by a licence fee payable by all households in the UK that own a television. This funding model means that it is able to experiment with new ideas which it feels compelled to do, at least in part, as a response to constant attack from right wing commentators[116]. The size and age of the Corporation (it began in 1924) means that its archives form an important part of the UK's political, scientific and cultural history for which the Web and linked data have provided the means of access for the people who paid for its production.

The publication of linked data on the BBC Web site has evolved over several years[117] and the Wildlife site[118] is an early example of the Corporation's use of linked data to present archive material. The URL of pages such as http://www.bbc.co.uk/nature/life/Taxus_baccata can be appended with '.rdf' to see the underlying data. The result of the approach is that every page links to more relevant information - exploration is hard to resist.

Experience gained through early work such as this lead to the BBC adopting a linked data approach when preparing to cover the World Cup 2010[119] and London 2012 Olympic Games[120]. The motivation for adopting a linked data approach is given in a blog post by Jem Rayfield[121]:

> "RDF semantics improve navigation, content reuse, re-purposing, search engine rankings, journalist determined levels of automation […] and will in future support semantic advertisement placement for audiences outside of the UK. [It] facilitates multi-dimensional entry points and a richer navigation."

It's worth highlighting elements of this quote; in the BBC's experience, using linked data offers the following benefits:

- improved navigation;
- content reuse;
- search engine rankings;
- semantic placement of advertising.

Of these, improved navigation and content reuse are of most direct relevance to government data. The BBC did not 'create a Web page' for every athlete in the

[115] http://en.wikipedia.org/wiki/Royal_charter

[116] For example http://adamcollyer.wordpress.com/2012/11/13/what-is-the-bbc-for/

[117] For a typical example of public discussion around the topic see http://lists.w3.org/Archives/Public/public-lod/2009Apr/0162.html

[118] http://www.bbc.co.uk/nature/wildlife

[119] http://news.bbc.co.uk/sport1/hi/football/world_cup_2010/

[120] http://www.bbc.co.uk/2012/

[121] http://www.bbc.co.uk/blogs/bbcinternet/2012/04/sports_dynamic_semantic.html

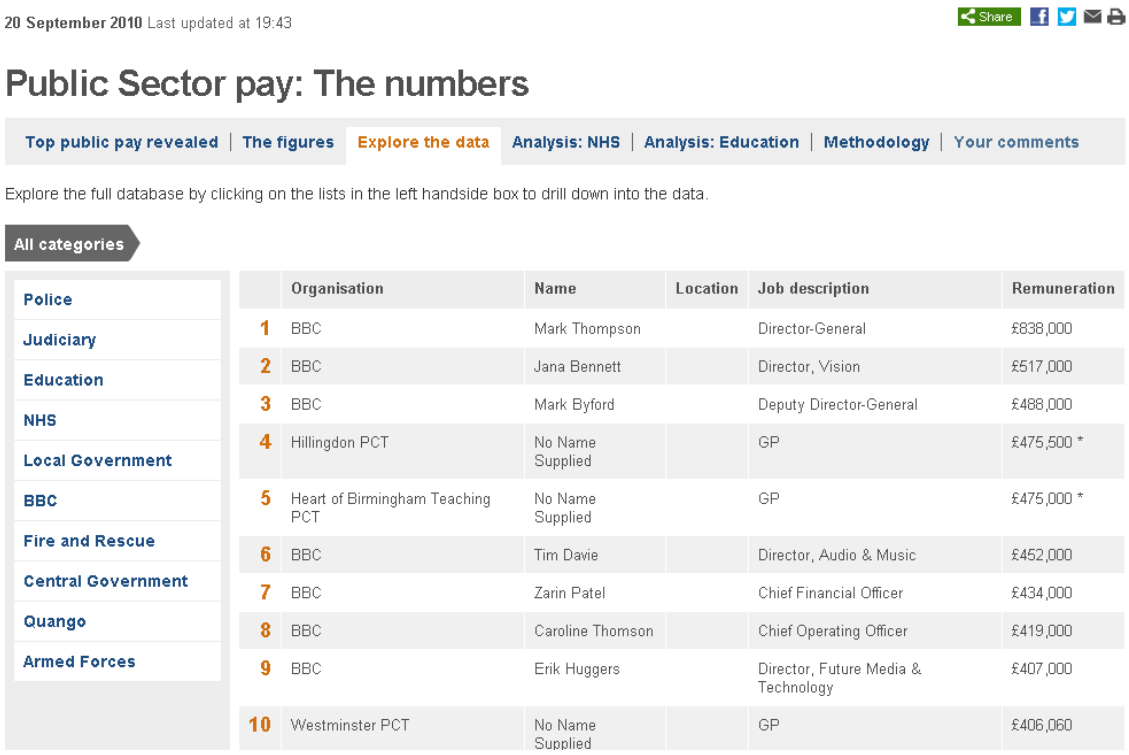
Olympics or player in the World Cup, rather they used data on each individual to generate a Web page that included links to relevant information. The analogy for public sector data might be data about services, institutions and individuals that became much more navigable for both external users and colleagues across departments.

NB: the data behind the World Cup and Olympics products is **not** available as open data.

Another interesting angle on the BBC's use of linked data comes from their work on data visualisation:

"[publishing data] … is an invitation for others to join the search for innovative ways of combining this data with other sources and present it in an engaging manner. Our experience in publishing data views of our knowledge on programmes, music, food or the natural world has shown in the past that there are many ready to answer that invitation. It may then be more useful and accurate to consider using and publishing open data not as discrete and separate activities, but as roles in a spectrum ranging from investigating and gathering the data to the interpretation, publishing and consumption."[122]

In September 2010 they created a visualisation of linked open data concerning public sector pay. The visualisation was just a basic table but it was interactive and allowed users to specialise their search in a way that would be more difficult to replicate using other data formats.



20 September 2010 Last updated at 19:43

Share

Public Sector pay: The numbers

Top public pay revealed | The figures | Explore the data | Analysis: NHS | Analysis: Education | Methodology | Your comments

Explore the full database by clicking on the lists in the left handside box to drill down into the data.

All categories

- Police
- Judiciary
- Education
- NHS
- Local Government
- BBC
- Fire and Rescue
- Central Government
- Quango
- Armed Forces

	Organisation	Name	Location	Job description	Remuneration
1	BBC	Mark Thompson		Director-General	£838,000
2	BBC	Jana Bennett		Director, Vision	£517,000
3	BBC	Mark Byford		Deputy Director-General	£488,000
4	Hillingdon PCT	No Name Supplied		GP	£475,500 *
5	Heart of Birmingham Teaching PCT	No Name Supplied		GP	£475,000 *
6	BBC	Tim Davie		Director, Audio & Music	£452,000
7	BBC	Zarin Patel		Chief Financial Officer	£434,000
8	BBC	Caroline Thomson		Chief Operating Officer	£419,000
9	BBC	Erik Huggers		Director, Future Media & Technology	£407,000
10	Westminster PCT	No Name Supplied		GP	£406,060

[122] How open data is redefining the roles of the journalist, audience and publisher. A Leimdorfer, O. Thereaux, Position paper for Using Open Data for Policy Modelling workshop. http://www.w3.org/2012/06/pmod/pmod2012_submission_9.pdf

Figure 14 Screenshot of the BBC's public sector pay visualisation from September 2010, available at http://www.bbc.co.uk/news/uk-11372185

The successful impact of this relatively simple visualisation is perhaps best shown by the many comments received showing that users had found it easy to draw conclusions from the data.

I am an administrative officer working for the Prison Service. My starting salary doesn't even break the £15,000 mark. I'd like to know what these people do that entitles them to earn more money than I'll see in 10 years! My department is under-staffed and over-worked, and it's because people are being paid these ridiculous salaries. I think it's incredibly unfair, and if something isn't done, and soon, many people are going to strike.

Jen, UK

These numbers are evidently supposed to be shocking, but I was actually quite impressed by how reserved the top salaries are. Any sizeable private company will pay its top earners far more than this, and personally I'd like to think the chief executives of such important national institutions are the pick of the bunch, so salary reduction would be a very bad idea.

Micheal, Newbury

I think it's perfectly acceptable that there are public servants paid more than the prime minister - he has a substantial coterie of aides and assistants to help him with his job whereas the same most likely can't be said of the people on your list. These people aren't making cups of tea for a living - they're professionals working in the public sector who could probably earn a lot more in the private sector. High-level positions require high-level wages to get the best people.

Chris Harker, Durham

Absolutely astounded! University Hospitals of Leicester NHS Trust has 349 of the supposedly most intelligent people in the UK earning more than the prime minister, a man who determines whether or not the country goes to war. That's a lot of people discussing who buys the bandages. Privatise the NHS, kick them to the kerb, reduce debt as they can't get it right.

Figure 15 Some of the comments received from BBC Online users

The BBC's big data-driven sites are created for major events like the World Cup and London Olympics and the infrastructure has been created to support this. In future, content will be added to by journalists directly, manipulated by editors and published in a relatively traditional workflow, albeit using advanced technologies in a custom-built system. A more general programme is now under way to include data embedded within news and sport pages. For example, the Web page at http://www.bbc.co.uk/news/world-europe-23198312 includes the embedded data shown in Figure 16.

```
@prefix og1: <http://opengraphprotocol.org/schema/> .
```

```
@prefix rnews: <http://iptc.org/std/rNews/2011-10-07#> .

<http://www.bbc.co.uk/news/world-europe-23198312> a rnews:NewsItem;
  rnews:creator "http://www.bbc.co.uk#org"@en-gb;
  rnews:datePublished "2013/07/05 13:43:01"@en-gb;
  rnews:description "The Vatican says John Paul II is to become a
    saint, after approving a second miracle attributed to him."@en-gb;
  rnews:headline "Vatican to make John Paul II a saint"@en-gb;
  rnews:thumbnailUrl
"http://news.bbcimg.co.uk/media/images/68569000/jpg/_68569052_68569045.jpg"@en
-gb;
  og1:image
"http://newsimg.bbc.co.uk/media/images/67373000/jpg/_67373987_09f1654a-e583-
4b5f-bfc4-f05850c6d3ce.jpg"@en-gb;
  og1:site_name "BBC News"@en-gb;
  og1:title "Vatican to make John Paul II a saint"@en-gb;
  og1:type "article"@en-gb;
  og1:url "http://www.bbc.co.uk/news/world-europe-23198312"@en-gb .
```

Figure 16 RDF data extracted from BBC News Web page

The effect of embedding data in BBC news Web pages can be seen in search results.

Vatican to make **John Paul** II a **saint**
BBC News - 3 minutes ago
The Vatican says John **Paul** II is to become a **saint**, after approving a second **...** Pope Francis also approved the **sainthood** of **Pope John** XXIII **...**

BBC News - **John Paul** II 'set for **sainthood**' with second miracle
www.**bbc**.co.uk/news/world-europe-23149489
3 days ago - **John Paul** II could be declared a **saint** this year after a Vatican committee approved a second miracle attributed to the Polish **pope's ...**

BBC News - **John Paul** II beatification: Politics of **saint**-making
www.**bbc**.co.uk/news/world-europe-13207940
May 1, 2011 - That fact was on clear display on 1 May, when **Pope John Paul** II was beatified, the final step before **sainthood**, in a ceremony in Rome that **...**

BBC News - Q&A: **John Paul** II's beatification
www.**bbc**.co.uk/news/world-europe-12194694
May 1, 2011 - Here the **BBC's** Rome correspondent, David Willey, answers questions **...** **Pope John Paul** II himself created more new **Saints** and Blesseds that **...**

Figure 17 Google search results showing effect of embedded data in BBC News pages

In Figure 17 we can see that Google has included the image defined in the embedded data, highlighting that particular result compared with the others. The provision of this data, which comes from their Linked Data Platform, helps navigation through the BBC's vast content store but it also reflects a growing

programme at the BBC to improve efficiency by making better use of data. As Oli Bartlett says in one of his blog posts[123]:

"The primary goal of the Linked Data Platform is to make sense of all the BBC's creative works and provide an API to allow the retrieval of any creative work about any 'thing', with the added benefit that we hold a semantic graph of data behind the 'things'.

This means the platform doesn't just know that tomorrow's episode of the Culture Show features Jarvis Cocker. It also knows that Jarvis is from Sheffield, was the lead singer in Pulp, that Pulp were a Britpop band, that they had a single called Common People, and that Common People was played on 6 Music this morning."

It remains to be seen what further use people inside and out of the BBC will make of this data beyond increased search engine exposure.

The terms of use of content on the BBC's Web site do not make specific mention of data as opposed to text, images, video and audio. One must therefore treat data in the same way as those other forms of content. As noted the funding for the BBC is via the licence fee paid by all television owners in the UK (whether they consume BBC programmes or not). With that proviso, all content on the BBC Web site is available free of charge although some premium content is only available in the UK (i.e. where the BBC licence fee is charged).

The licensing terms[124] make a distinction between personal and commercial use. Essentially stating that personal use is unrestricted and commercial use is not permitted.

The corporation has wrestled with how it might make its data available under a more permissive licence for some time but, at the time of writing, has not yet come to a resolution. This fact and the examples we have seen all point to the effort primarily being about using linked data to improve the efficiency of the internal systems rather than to provide data as a service for others to use.

II.8.2 Interview: Dave Rogers and Oli Bartlett

Interview date	9/7/13
Interviewee	Dave Rogers and Oli Bartlett
Interviewer	Phil Archer

Q: How does the supply of Linked Data relate to the public task of your organisation?

A: Exposing programme data as RDF is one strand, but it's separate from solving internal problems. We use LD internally. The stats from the programmes section of the BBC Web site show that the data is not being used a lot.

[123] http://www.bbc.co.uk/blogs/internet/posts/Linked-Data-Connecting-together-the-BBCs-Online-Content

[124] http://www.bbc.co.uk/terms/

The work in making data available through the Programmes pages and Wildlife finder is not driven by a business case, it's more of an experiment: "let's see what happens if we do this." However… BBC 'internal' is huge. BBC is more like a government. You can see different departments of the BBC as separate teams in different places more like government departments than being part of a single organisation.

When we say 'customers' or 'clients' we mean other BBC departments, not people outside the BBC. We're building the linked data platform in an agile, lean way to be able to respond to those customers.

Some use cases: story lines on the news web site. We're also working on a knowledge and learning product. It's all about picking off highest priorities, from online content production through to the audience. We want to be able to sort our content online – bridging gaps between online, iPlayer etc.

The LD platform doesn't go back as far in the chain as the journalists (yet). They will write for the Web and that will link automatically. Multiple journalists can use LD to describe what they're inputting -> this builds the story. Currently it's only using LD to describe finished things but may want to push back up the production chain. We'll only do it where there's an obvious need.

Some problem spaces are easier than others. What are the useful concepts for everyone to share – people, places, organisations, events – some of this quickly starts to get complicated. There are conceptual boundaries around people places and orgs that help it work across different sectors. Lots of programme contributors are marked up with corresponding DBpedia URIs. This helps us to ingest and link up with other people. This is only possible with LD.

Anything the BBC creates is owned by the licence payer – so we must make it available. So the question is not why we make the effort to make our content available, it's why shouldn't we do it.

Linked data makes archive and current content easy to handle in the same way. So news can ask "what have we ever broadcast about person X?" We can find everything people have got – although it's not necessarily available.

Availability is very complex at the BBC. It covers rights, time, geographical restrictions

Q: Does your organisation have a "business case" document that motivates your investment in supplying Linked Data?

A: We have an event: "Connected studio" at end of the month to open our data to other BBC teams. The LD platform means we can understand and aggregate content from across the BBC, there's no need for departmental APIs. The proposed step from closed data to open closed data (the data in the linked data platform is not currently available as open data).

Q: What are your future plans? Do you plan to expand or abolish the supply of Linked Data?

A: The aspiration is to reduce the cost!

Q: Did the supply of Linked Data give rise to new opportunities for flexible data integration within your organisation and with partner organisations? Increase in data quality? New services? Cost reductions?

A: The LDP is hosted in the cloud. We're loading the cloud with lots more data than we have in current LD platform. We can put loads more LD in there and it all joins up to some extent. We can use for production, even if the quality varies. The LDP for the 1st time puts everything in one place.

We have APIs on the platform for conflict resolution and de-duplication. We can access data that has previously been siloed in a week or two (and the data we have in there goes back up to 7-8 years ago). We can make better use of the content than previously as it's been inaccessible until now.

By the end of July 2013 we'll be able to aggregate programme and news articles and Web pages about a given person. The quality of those links is not high enough for public – this is a service for our own staff who can then use it in their own content creation. We want to see what the appetite is for that internally. BBC developers might then create products that then see the public.

We'll be launching this at the Connected Studio event.

Q: What are the main enablers / inhibitors for Linked Data to deliver value for its reusers?

A: Licensing and quality are the inhibitors for us. Also the (lack of) maturity of technology – currently available triple stores are just not good enough. A lot of work on LD has been done in academia where, if a store goes down, it doesn't matter so much. Performance and resilience matters to the BBC. In a production world that's challenging. We're one of the first to be solving the problem. We had to – at the height of the London 2012 Olympics the LDP was handling 2,000 SPARQL queries per second – that's not something other people generally have to cope with.

Data management is probably more important to BBC than other LD people. Internally we can update and remove data. When we move to open data we're only going to expose a stable subset so that URIs persist. We need to be as flexible with internal data as you would be with SQL etc.

Q: Do you supply service level statements for Linked Data?

There is no LD available except programmes which is not official. It's probably coming, but not yet. Even when we do go to LOD, we probably won't so SLAs, people interested in our data are probably not interested in LD for now.

Q: Does your organisation have a URI policy?

We have a URI design policy within team. But at the whim of a bigger org and other depts. we might think otherwise. BBC has a mixed set of policies that may or may not be followed. There is no one canonical URI policy. We do recognise need for canonical URIs in LDP but basically it is bbc.co.uk/things/GUID – a flat policy that avoids ownership. Of course there's a URI for everything in the platform but we don't want to stop product teams exposing their data in their own URI system - i.e. we're very distributed. Trying to impose URI designs across the BBC wouldn't work.

Also don't want to say you can't expose RDF there because we do it all. So you may get 2 BBC URIs for same thing. If exposed they'll have owl:sameAs links.

The BBC itself doesn't benefit from stable organisation structure although our URIs are already stable.

Q: Which skills and competencies did you need in order to supply Linked Data? Do you have them in-house or contracted?

The project wouldn't have got going if we hadn't contracted out LD and software. Competencies for Olympics and World Cup now different. It's now about good Java coders – we no longer need external LD expertise. A good Java developer got up to speed on LD within 2 months. They need to know SPARQL, the mechanics of RDF. They don't need to know about reasoning (Which would present even more problems for performance).

We have 4 data architects across the BBC. One in our team who is the central data architect (Sofia). She's the reference point if we need to go deeper. Sofia has an academic background and does ontology design. We do contract out specific areas, like food – one off projects. We ring fence and price it for outsource. The LDP is about ongoing work. We have 6 developers on the LDP.

Some data modelling we did outsource later became incompatible with what we do now. We struggle with typical external ontologies. We start using say 5% of it, then we'll maybe use more. It's more efficient for us to be able to change our ontologies. We can't have the same ontologies internally as on the Web as we need to be flexible. Some ontologies are only used by one product. Lots of modelling for creative works is based on schema.org and we will use it when we make the data open. Internally in the triple store it's all our own ontology so that we're constrained by externalities.

Q: Which organisations are key partners in the supply of Linked Data?

A: Even in the Olympics we took a raw XML from our external suppliers as we felt we were the experts, not them. We partner with GeoNames, MusicBrainz and Wikidata around consuming their data. Our Music department has mechanisms for feeding back to MusicBrainz so they're collaborators but it's not commercial. Individual product teams may have key partners, Jeremy Tarling for example partners with the Guardian. The Sport ontology was designed with the Press Association. So there are some relationships.

Q: Which investments has your organisation made to enable it to supply Linked Data?

A: We don't invest in linked data. We invest in a product that is delivered using LD.

Q: How many FTEs in your organisation are involved in provisioning the supply of Linked Data?

We have a team of 12 for LDP. There are 3 more architects, other developers, developing APIs, bunch of R&D people in this area.

Q: Which costs have you incurred to publish Linked Data, maintain, and promote it? What are the trends?

Connected Studio is an internal BBC hack day. There has been some travel to cover. LD gives the opportunity to move more quickly and understand what we

should be doing. Prioritisation is not easy, especially when you're not building for the public. We want to showcase what we've done and get people to think about it – think LD when building products.

Q: Under which licence is Linked Data made available for reuse?

A: TBD when we open the data. Do we want data to be attributed or not? A key factor is brand protection. It's a challenge we haven't handled yet.

We keep things in the triple store that we know the BBC are comfortable with being public. We don't want opening the data to be a massive tech challenge. Most of our stuff is content metadata, which is not the high value stuff.

Q: Do you provide / make use of any mechanisms on user feedback or evaluation?

It's pretty informal. We have lots of performance monitoring in case things go wrong but there's no real value cycle. A new feature may not be used for some months, then we often won't have a good way to get feedback. Hard to say that LD provided that value.

Q: Can you talk about the number of de-referenced URIs / queries?

A: The no. triples dereferenced is an indicator for success – tells us that people find it useful. We currently have 12.5 Million triples in the LDP. That number can be doubled for 'implicit triples.' Connected Studio might have an effect. Current focus is on news. A typical query rate is 1K/s when football is on.

Is RDFa linked data in the current sense? It's not linked. The benefit is seen in SEO. Thanks to us including RDFa in our pages we're more often above the Guardian and the Daily Mail in search results.

Since the Olympics, our work has been about building up the platform so that any product can use it. That's been a lot of work. See also the curriculum website. BBC/Music will soon be using it.

Interview Summary

The BBC has an unusually large amount of high value audio, video, images and text content spanning nearly 90 years. Managing that content, managing the information within it and making sure that it is available to people throughout the organisation is a huge task. The development of the linked data platform is a response to this demand and is in stark contrast to the Digital Media Initiative[125] that was infamously ditched in May 2013 after more than £80 million had been spent on it.

It is the internal usage of the data that is driving the development of the LDP; public access to that data is very much a secondary concern. The BBC's demands of resilience and performance outstrip many other applications as the rate of two thousand SPARQL queries per second during the Olympics demonstrates. The BBC expects to share its expertise as well as some of its data in due course but, again, the priority is improving internal efficiency.

[125] http://en.wikipedia.org/wiki/Digital_Media_Initiative

II.9 UK – Companies House

II.9.1 Desk research

Companies House[126] is the register for limited companies, limited liability partnerships (LLPs) and other entities in United Kingdom. It makes basic data available as linked data for more than 3 million registered limited companies.

It has two main areas of activity:

- Information registration, including the incorporation and striking off of companies and maintaining a register of the documents delivered under companies, insolvency and related legislation;
- Information provision to the public on companies.

An overall assessment of the Companies House business case was included in the POPSIS study[127] of the European Commission. This study disambiguates the provision of linked data from open data.

Value proposition: Data sold / made available

The list of all information and data types provided is available in the Companies House price list[128]. Linked data for all registered companies is provided free of charge. However, while Companies House provides a lot of information for free to the public (e.g. company appointments, insolvency details), only basic company data is available as linked data.

URIs return the following data: company name, registered office address, company status, incorporation date, country of origin, company type, nature of business (SIC), accounting reference date, date of last accounts/annual return filed, date of next accounts/annual return due, previous names as linked data.

Revenue system: Price structure and licensing usage conditions

Companies House offers access to most of its data for free. This includes basic company details, insolvency information, history of company transactions, etc. There is also a free "monitor" service, which provides alerts when documents registered by specified companies become available on the public register.

Information that is not available for free, such as full mortgage details or personal appointments, is paid for on a usage base (e.g. screen charges or document download). The basic free company information can also be accessed through a free mobile application.

Only basic company details are offered as linked data. This service is free.

Moreover, Companies House offers a subscription service for a monthly fee, Paid subscription is optional to retrieve/purchase information, and subscribers paying a monthly fee have the same price list[128] as non-subscribers.

[126] http://www.companieshouse.gov.uk/
[127] European Commission: Pricing Of Public Sector Information Study, Summary Report, p49.
[128] http://www.companieshouse.gov.uk/toolsToHelp/ourPrices.shtml

Companies House, as a public information provider, makes all information relating to limited companies available for public inspection. It places no restriction on how the information is used after purchase other than the following:

- All customers, including bulk customers, must take their own legal advice regarding possible breach of third party copyright.
- Customers cannot reproduce the Crown insignia or use the Companies House logo.
- If information is used from guidance notes, the website, publications or statistical tables the customer is required to credit Companies House as the source of the information.

Cost structure

Specific cost information about the publication of open linked data is not publicly available.

Companies House considers the following when setting its fees:

- As a Trading Fund, Companies House has a statutory duty to break even over time and to achieve an average annual return (surplus) of 3.5%.
- Managing Public Money requires fees to be set to recover the full cost for each separate service allowing for a cost of capital of 3.5%. This prohibits cross-subsidy between different services.
- The EC First Company Law Directive, requires copies of company records to be made available at the "administrative cost" of producing them.
- The EC Capital Taxes Directive allows company registration costs to be met from fees, but prohibits charges that are effectively taxes. This means that prices cannot be set above costs for the relevant services.
- European Case law provides further guidance on the costs that can, and cannot, be taken into account for fee-setting

Key findings

- **Premium for some information**: most of the data available is available as linked data, however, only basic company information is provided as data in any format, the bulk of the information held by Companies House is as unstructured information.
- **Linked data, free of charge**: linked data is provided free of charge.

II.9.2 Interview 1: Mark Fairhurst, Chris Smith, Stacy Smith

Interview date	26/7/13
Interviewee	Mark Fairhurst, Chris Smith, Stacey Smith
Interviewer	Phil Archer, W3C

The public task[129] of Companies House is to gather data on companies in the UK and make it available to the public. It fulfils its role in disseminating information via a number of Web services and via bulk download. As noted above, some of these services incur a fee.

One of the services made available by Companies House is WebCheck. This allows you to do a search for a company and find the information about it. The linked data service is simply a serialisation of that data as RDF. In other words, Companies House, like OpenCorporates, does not run a triple store. Interestingly, the Companies House staff don't refer to the linked data service, for them it is 'the URI service' – i.e. they provide stable URIs as identifiers for companies that can be de-referenced to obtain basic information but it's not designed to be 5 star linked data.

Many of the questions asked in other interviews were pre-empted and answered in documents made available before the interview. They are copied verbatim in the following text with permission.

Begin quote 1

Companies House URI

Background

The Companies House URI service went live on the 14th October 2011, providing a URI for every Incorporated Company registered at Companies House, which when accessed, provides basic company details in a variety of formats. For details of the formats and the data provided please refer to the user guide available at: http://www.companieshouse.gov.uk/about/pdf/uniformResourceIdentifiersCustomerGuide.pdf

Business Case

The register exists so that people can access information on companies and use it to make decisions. Over time the way we have done this has changed dramatically as technology has enabled new methods of delivery. E.g. microfiche, image, bulk products etc. We are now in a position where over 99% of all searches of company information are made electronically and with the majority of information now received in data format, we are moving rapidly into a world where electronic data is the default. The development of the semantic web (web 2 technologies) is presenting a whole new range of opportunities to link data, mash data and enrich basic information providing greater value to the end user.

The company number is a unique identifier for each company; however on the web the number is not unique but just a large number which if searched would deliver a wide range of results. Currently there is no way of uniquely identifying a company on the web. This is a common problem and the method to uniquely identify anything over the web is by using a Universal Resource Identifier (URI).

[129] http://www.companieshouse.gov.uk/about/ifts.shtml

If a URI was available for companies then this could be used wherever information on a company was published on the web. For example, when government departments published supplier information or enforcement data.

Project

The project to deliver the service took around 4 months from start to finish, with the key tasks taking:

- Project initiation / closure – 1 month
- Design – 1 month
- Development – 1 month
- Testing/Implementation – 1 month

The design phase involved defining the URI, agreeing the formats to be provided and the screen design of the HTML format. Defining the URI was based on the principles as set in the Cabinet Office Guidelines[130] and in collaboration with leading figures within the open data arena. The URI was officially signed off by officials from the Cabinet Office with the structure detailed in the attached document. [This is included below as well]

Development utilised existing infrastructure and application code, used to provide other online search services, which resulted in the short delivery timescales. The main focus of the development phase was to cater for the delivery of the data in the agreed formats. Existing search services provided data in either HTML or XML so work was required for the others.

During project closure lessons learnt were documented which are summarised as:

Insufficient time was planned for the collaboration with the Cabinet Office and prominent industry officials on the URI structure. This resulted in changes to developed code to compensate for the changes made after final sign off.

Live Service Operation

Since the service was put live it has been well received with positive feedback from the general public via Twitter and other sources. We have also received some positive feedback from prominent figures within the linked data industry endorsing the provision of the service, such as Professor Nigel Shadbolt.

Since its introduction statistics show a weekly average of 720,000 hits, with the following a breakdown of the last 2 months:

Format	05/03/2012	12/03/2012	19/03/2012	26/03/2012	02/04/2012	09/04/2012	16/04/2012	23/04/2012
HTML	113,360	90,195	110,146	131,120	133,643	137,003	118,796	133,870
XML	108,965	113,450	96,207	13,976	13,310	449,139	1,334,929	10,727

[130] http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-URI-sets-uk-public-sector.pdf

RDF	16,892	10,351	9,297	11,768	12,545	12,474	13,067	8,120
YAML	11,240	8,402	4,565	5,439	5,147	6,797	6,965	3,300
JSON	428,794	853,666	637,454	624,449	609,282	568,601	547,954	427,964
CSV	196,614	126,218	86,788	85,484	87,200	87,828	107,055	111,042
Total	**875,865**	**1,202,282**	**944,457**	**872,236**	**861,127**	**1,261,842**	**2,128,766**	**695,023**

There is no specific insight into what the data is being used for or by whom but we do know that the OpenCorporates website has embedded the URI links within their site.

Next Steps

- Use for Local Government financial data comparability
- URI name search service
- Link to Ordnance Survey
- Mobile App provision using the URI service.

End quote 1

Mark Fairhurst was able to provide more recent stats are as follows:

Format	07/01/2013	14/01/2013	21/01/2013	28/01/2013	04/02/2013	11/02/2013	18/02/2013	25/02/2013
HTML	217,024	219,390	258,211	277,471	342,371	544,764	587,417	472,100
XML	83,784	66,449	700,452	941,766	390,631	234,271	758,456	142,516
RDF	22,518	25,101	24,371	17,649	38,046	145,217	115,236	85,608
YAML	6,463	8,259	8,452	26,245	23,172	34,142	16,102	10,744
JSON	448,311	820,701	1,146,598	1,037,332	562,091	948,538	1,033,599	908,937
CSV	0	0	0	0	19,049	75,478	79,245	51,652
Total	**778,100**	**1,139,900**	**2,138,084**	**2,300,463**	**1,375,360**	**1,982,410**	**2,590,055**	**1,671,557**

Format	04/03/2013	11/03/2013	18/03/2013	25/03/2013	01/04/2013	08/04/2013	15/04/2013	22/04/2013
HTML	283,926	311,656	335,996	316,616	300,656	328,612	285,798	304,762
XML	767,765	81,110	103,314	71,122	77,604	65,263	75,892	421,936
RDF	32,885	15,879	33,028	27,272	34,277	25,797	22,265	27,299
YAML	267,077	286,227	181,958	109,172	57,011	13,278	15,873	15,414
JSON	821,712	509,678	535,503	474,954	475,934	938,496	808,092	539,339
CSV	15,709	15,879	18,007	13,609	9,147	5,412	6,243	6,340
Total	**2,189,074**	**1,220,429**	**1,207,806**	**1,012,745**	**954,629**	**1,376,858**	**1,214,163**	**1,315,090**

These later statistics are plotted in Figure 18, which shows a gradual increase across the reporting period with a substantial peak in early 2013. It's clear that JSON and HTML are the key formats used and that the peaks and troughs in the overall graph mirror specific technologies. For example, the big peaks in demand seen in early 2013 seem to be driven entirely by requests for the data as XML. The peak in April 2013 mirrors a peak demand for JSON. This suggests that the URI service attracts a small number of large scale users. The Companies House team

noted that excessive use of the data from a single IP address will trigger them imposing a block on that IP address.

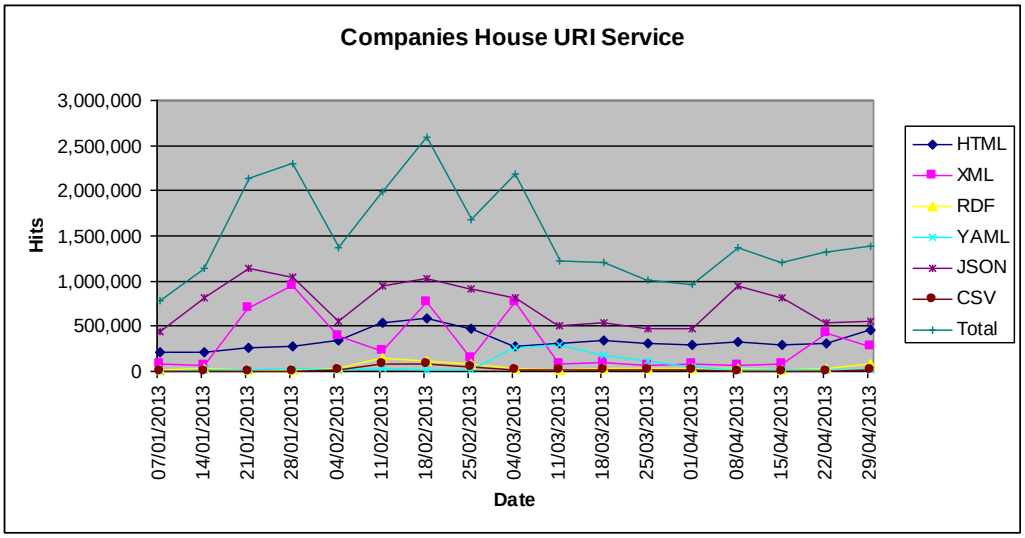


Figure 18 Companies House URI Service hit rate

Q: Does your organisation have a "business case" document that motivates your investment in supplying / re-using LOGD? Can you share it with us?

A: The register exists so that people can access information on companies and use it to make decisions. Over time the way we have done this has changed dramatically as technology has enabled new methods of delivery. E.g. microfiche, image, bulk products etc. We are now in a position where over 99% of all searches of company information are made electronically and with the majority of information now received in data format, we are moving rapidly into a world where electronic data is the default. The development of the semantic web (web 2 technologies) is presenting a whole new range of opportunities to link data, mash data and enrich basic information providing greater value to the end user.

The company number is a unique identifier for each company; however on the web the number is not unique but just a large number which if searched would deliver a wide range of results. Currently there is no way of uniquely identifying a company on the web. This is a common problem and the method to uniquely identify anything over the web is by using a Universal Resource Identifier (URI).

If a URI was available for companies then this could be used wherever information on a company was published on the web. For example, when government departments published supplier information or enforcement data.

Q: How does the supply / reuse of LOGD relate to the public task of your organisation?

A: The public task of Companies House is to gather data on companies in the UK and make it available to the public. It does this via a number of Web services and via bulk download. As noted above, some of these services incur a fee.

Q: What are your future plans? Do you plan to expand or abolish the supply / reuse of LOGD?

A:

- Use for Local Government financial data comparability
- URI name search service
- Link to Ordnance Survey
- Mobile App provision using the URI service.

Q: Can you describe the Linked Data infrastructure of your organisation?

A: One of the services made available by Companies House is Web Check. This allows you to do a search for a company and find the basic information about it. The linked data service is simply a serialisation of that data as RDF. In other words, Companies House, like OpenCorporates, does not run a triple store. Interestingly, the Companies House staff don't refer to the linked data service, for them it is 'the URI service' – i.e. they provide stable URIs as identifiers for companies that can be de-referenced to obtain basic information but it's not designed to be 5 star linked data.

Many of the questions asked in other interviews were pre-empted and answered in two document made available before the interview. They are copied verbatim below with permission.

Q: Which investments has your organisation made to enable it to supply / reuse LOGD?

A: The project to deliver the service took around 4 months from start to finish, with the key tasks taking:

- Project initiation / closure – 1 month
- Design – 1 month
- Development – 1 month
- Testing/Implementation – 1 month

The design phase involved defining the URI, agreeing the formats to be provided and the screen design of the HTML format. Defining the URI was based on the principles as set in the Cabinet Office Guidelines[131] and in collaboration with leading figures within the open data arena. The URI was officially signed off by officials from the Cabinet Office with the structure detailed in the attached document. [This is included below as well]

Development utilised existing infrastructure and application code, used to provide other online search services, which resulted in the short delivery timescales. The main focus of the development phase was to cater for the delivery of the data in the agreed formats. Existing search services provided data in either HTML or XML so work was required for the others.

[131] http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-URI-sets-uk-public-sector.pdf

During project closure lessons learnt were documented which are summarised as:

Insufficient time was planned for the collaboration with the Cabinet Office and prominent industry officials on the URI structure. This resulted in changes to developed code to compensate for the changes made after final sign off.

Q: Do you provide / make use of any mechanisms on user feedback or evaluation?

A: Since the service was put live it has been well received with positive feedback from the general public via Twitter and other sources. We have also received some positive feedback from prominent figures within the linked data industry endorsing the provision of the service, such as Professor Nigel Shadbolt.

Q: Who are the main users of your LOGD services? Is LOGD only used by external customers or also for back-office exchange (e.g. with other agencies)?

A: There is no specific insight into what the data is being used for or by whom but we do know that the OpenCorporates website has embedded the URI links within their site.

Q: Does your organisation have a URI policy? Can we have a copy?

Quote 1

URI Definition

The proposed structure of the URI to retrieve company data from Companies House is as follows. The proposal has been made considering the guidance as published in the "Designing URI Sets for the UK Public Sector"[132] paper.

URI Structure

http://business.data.gov.uk/id/company/{compositecompanynumber}

The URI can be split into the following component parts, with each explained further below:

Domain - http://business.data.gov.uk

Path Structure (Type) - /id

Path Structure (Concept) - /company

Path Structure (Reference) - /{compositecompanynumber}

Domain

The recommendation is to use "business" to represent the business sector, in-line with the examples of education and transport, and within the data.gov.uk collection of UK public sector URIs. NB – business.data.gov.uk exists!)

Path Structure (Type)

[132] http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-URI-sets-uk-public-sector.pdf

Contains the string "id" to show this is an identifier URI. There could be potential in the future to use other types such as "doc", for example to access specific filing images for a company?

Path Structure (Concept)

Contains the string "company" to identify company data. Other concepts could be used in the future to reference business data within other government departments such as tax, for example, which could be used to search for data within HMRC.

Path Structure (Reference)

This references the actual company data and is a composite key consisting of:

{compositecompanynumber} – The 8 digit company number as appears on output services.

Examples

http://business.data.gov.uk/id/company/01777777

http://business.data.gov.uk/id/company/SC216315

http://business.data.gov.uk/id/company/NI074051

http://business.data.gov.uk/id/company/FC020100

http://business.data.gov.uk/id/company/OC362663

http://business.data.gov.uk/id/company/SL002100

Resolving the URI

It is proposed that the http://business.data.gov.uk/id/company is effectively the part of the URI which is used to resolve and route the requests to Companies House and is proposed would be achieved as DNS routing. Once the URI is received at Companies House a 303 redirect will be performed to access the data.

REID

The aim of the REID (Registered Entity IDentifier) initiative is to establish a way in which entities in business registers can be identified by a number that is unique at the world level. The current proposal is for the identifier to be:

CCRRRR.NNNNNNNN-P where

CC – ISO country code

RRRR - Register identifier within country

NNNNNNNNN - Number unique within the register – Maximum 35 characters

PP – Check characters

The above has been considered in designing the URI and it is considered that if the above is adopted then the URI would support as follows:

CC – This would be the UK and would translate to the data.gov.uk domain.

RRRR – Optional and as GB only has one register is not required.

NNNNNNNN – This is the {compositecompanynumber}

PP – Check characters not considered.

Design Principles

Within the Cabinet Office guidelines there are a number of design principles documented. The following table lists each with a narrative on how Companies House is attempting to resolve.

Principle	Must / Recommend	Comment
Use HTTP so that URIs can be resolved	Must	HTTP will be used
Use a consistent path structure to explicitly indicate the type of URI	Recommend	
The publisher will make it clear whether the set is promoted for reuse by other parts of government and/or the public	Must	Expected to be reused and is within the data.gov.uk domain
Public sector URI sets should publish their expected longevity, and potential for reuse	Must	The URI has been designed to be reused and is expected to remain valid for at least ten years. In effect the first part of the URI is to be used for DNS routing and the second part is simply an identifier to the data set. Therefore if Companies House ceases to exist as a department as long as the data set remains it simply means re-routing and ensuring the data can still be accessed via the identifier.
Those public sector URI sets that are promoted for reuse should be designed to last for at least 10 years	Recommend	See above
Where more than one Representation URI is available, provide a Document URI where Content Negotiation can be used to provide the most appropriate representation	Recommend	Content negotiation will be performed to either return data as RDF or HTML.
Avoid exposing the technical implementation of a URI in its structure	Recommend	No technical implementation is exposed.
As a minimum, provide a machine-readable Representation URI	Must	RDF will be available
If appropriate, provide a human-readable Representation URI in HTML	Recommend	HTML will be available

Provide a means of discovering each of the available Representation URIs for a single Document URI	Recommend	
A URI set will publish its authorisation, authentication, and data quality characteristics using a common vocabulary	Must	
A URI structure will not contain anything that could change, such as session IDs	Must	Nothing can change.
A URI path structure will be readable so that a human has a reasonable understanding of its contents	Recommend	Easily understood domain and path structure used.

Q: Can I ask a bit more about the motivation, beyond what is written in the documents?

A: The business driver really was if we can do this then we should. We were able to provide the URI service with very little effort – it's just part of what we're here to do. It wasn't demand-lead.

Q: Can we talk a bit about charging structures…

A: We only charge Recovery Costs – that is, we can only charge what it costs us to run a service. Typically that means individual transactions cost £1. We can't cross-subsidise, i.e. we can't take money from one service and set it against another.

We're not funded by the Treasury at all. Companies House gets its income from registration fees and data consumption where the administrative cost of recovery is low enough to make it worthwhile. The cost of the URI service is *de minimus* meaning it's so low that it's simply not worth charging for.

Q: What are your future plans?

A: The most important service for us – the one most of our customers use – is the XML Gateway. That is free to use although you do need to register and so it's not fully 'open'. We're now working on providing a JSON API for it. That will be free too although it includes a developer API Key – which is pretty standard and not a barrier. We expect the new JSON API to significantly reduce the demand on the URI service.

We have been considering improving the URI service by referring to the Ordnance Survey Post Code data and maybe working with the Land Registry and Royal Mail to include UPRNs – the unique identifiers for every postal address. Improvements such as this are things we'd like to do but the XML Gateway and new JSON API are more important to us. Developers who use our data want JSON. In fact, is it really up to us to do the amalgamation? If we provide our data and make sure it's reliable and authoritative, is that not where our responsibility ends?

Q: Do you know who your data users are?

A: No. Yes, we know about OpenCorporates of course (Chris Taggart helped us) and we see evidence that a lot of 'users' are actually bots crawling to collect the

data en masse. In our view that's misuse as it suggests that they're then storing the data and using the local copy - we really don't want people to use that because of course it quickly goes out of date. We want them to use the live data – more reason to focus on the new JSON API.

Q: Can I push a little on the development cost, please. The report you've given us says the project took 4 months – how many people were involved?

A: There will have been an inception document and a closure document but we've been unable to find them, sorry. There was an architect and a developer plus testers. Overall 4 people were involved. You could boil it down to about 2 months FTE.

Q: I notice the documents you sent us include mention of the REID standard developed by the European Business Register – did you consider using it?

A: We considered it but the identifiers aren't resolvable so we don't consider that they add any value to what we provide already. The main thing about URIs is that you can look them up and REID doesn't offer that.

Interview summary

The 'URI Service' at Companies House was an easy add on to their existing services. It is well used but the expectation is that use will decrease, not increase, with the launch of the new (JSON) API to the Gateway service. The provision of stable, de-referenceable URIs that return basic information about companies is clearly a useful building block on Britain's information infrastructure but, beyond incremental improvements (such as using the Registered Organisation Vocabulary), it is unlikely that any significant further development will take place. The service is easy to run and costs are *de minimus*, therefore, so long as there is any demand one can look forward to it continuing.

II.10 UK – Department of Environment, Food and Rural Affairs (DEFRA)

II.10.1 Desk research

The Department of Environment, Food and Rural Affairs, DEFRA, has a good track record of publishing data. For example, many geospatial data sets created by DEFRA's agencies are available through http://www.geostore.com/environment-agency/ and DEFRA has published hundreds of data sets through data.gov.uk (at the time of writing a simple search for 'DEFRA' on data.gov.uk returns 364 results). Many of these datasets are of high value and/or interest: flood maps, otter surveys, sites of special scientific interest, climate records and more. The data is generally published under the UK Open Government Licence[133] but the climate data published by the Meteorological Office is subject to more restrictive licensing terms and users must register before downloading anything. The provision of open data by DEFRA and its agencies is the result of the adoption of an Open Data Strategy[134].

There are two linked data sets published by the Environment Agency (an agency of DEFRA): Bathing Water Quality and Integrated Catchment Management. The development of these data sets is closely tied to DEFRA's implementation of the INSPIRE Directive and the definition of URI designs for spatial objects[135] currently undergoing revision with an updated version due for publication imminently.

Work on the Integrated Catchment Management Data has followed that done on the Bathing Water Quality and, at the time of writing, is still marked as being 'experimental' even though the data model has been defined and the reference data published. For example, http://location.data.gov.uk/so/am/RiverBasinDistrict provides reference data about river basin districts and the water bodies and management catchments that they contain. The documentation provided for this work[136] is extensive and reflects the sophistication of the data modelling used. An application to view the data is being developed[137] but is not yet publicly available.

The Bathing Water Quality data[138] was first published as linked data in 2011. It exposes the weekly sampling data taken at over 500 locations around Britain's coast and uses a *relatively* simple data model given the need to identify locations, pollution hazards and a time series of observations.

The data is accessible through a SPARQL endpoint and using the Linked Data API[139]. This is the method used by UK government linked data publishers to provide ready-made data views as Web pages, for example, details of the sampling point at Aberafan can be seen at http://location.data.gov.uk/doc/ef/SamplingPoint/bwsp.eaew/36800. The primary

[133] http://nationalarchives.gov.uk/doc/open-government-licence/version/2/

[134] http://data.gov.uk/sites/default/files/Defra%20Open%20Data%20Strategy.pdf

[135] http://www.data.gov.uk/sites/default/files/Designing_URI_Sets_for_Location-V1.0.pdf

[136] http://environment.data.gov.uk/icm/icm-data-model.html

[137] http://environment.data.gov.uk/icm/icm-application.html

[138] http://environment.data.gov.uk/bwq/

[139] http://code.google.com/p/linked-data-api/

means of access to the data for end users, however, is the Bathing Water Quality Explorer[140].

Figure 19 Screenshot of the home page of the UK Environment Agency's Bathing Water Quality Explorer

This application allows users to find out about bathing waters of interest. Navigation is possible either by clicking on the map or searching for a location by name. For each location the application offers not just water quality measurements but an image and short text description as well as other information about the water catchment area, local sewerage facilities etc. to give a detailed view. Importantly, the data is updated regularly. The screenshot in Figure 20 shows data recorded less than 2 weeks previously.

[140] http://environment.data.gov.uk/bwq/explorer/

English

Bathing Water Data Explorer

Environment Agency

Bathing water at Felixstowe South

home · bathing water information · sampling data · suspensions · help

Suffolk, England **Local authority**: Suffolk Coastal District
Year of designation: 1988

Tweet +1

Felixstowe is a resort town but is also home to the country's largest container port on the Orwell Estuary. Felixstowe South is a sand and shingle beach gently sloping to the sea in a sheltered bay. There is sand at low tide. The beach is backed by a promenade and gardens.

Water sampling point location
lat/long: 51.9543304067933, 1.34146436108153; easting/northing: 629700, 233700
View in: Google Maps, Bing Maps, OpenStreetMap

Nearby bathing waters:
Felixstowe North, Dovercourt, Walton, Frinton

Bathing water quality results

Recent results from water quality assessments under the Bathing Water Directive appear below, or view more details of assessment results

Annual Compliance Results					Latest weekly in-season
2008	2009	2010	2011	2012	Tue, 25 Jun 2013 10:15:00
Higher	Higher	Higher	Higher	Higher	

Catchment map

EA bathing water sampling location
Surface water catchment boundary

Bathing water map

EA bathing water sampling location:
higher standard
minimum standard
fail
closed
not classified

Surface water outfall
Emergency or storm overflow
River or stream
Treated sewage works outfall
Disused outfall

Catchment description

The River Deben is 5km to the north and this drains a large mixed catchment. The Orwell and Stour estuary is 3km to the south and drains a large catchment containing some industry and the Port of Felixstowe

Figure 20 Partial screenshot of a page from the Bathing Water Quality Explorer showing details of a specific location (http://environment.data.gov.uk/bwq/explorer/info.html?site=ukh1406-11000)

As well as the Bathing Water Quality Explorer, the Environment Agency also provides a tool to generate widgets[141] that can be embedded on any Web site and so include live data about a given location. The example shown matches the 'full' version available in the Explorer.

The linked data published by DEFRA is available free of charge under the UK Government Open Data Licence.

The provision of both the data and the tools appear to meet a real demand. Britain's Marine Conservation Society (MCS) produces an annual Good Beach Guide[142] and national newspapers often feature stories about bathing water quality[143], especially when it doesn't meet EU guidelines[144]. However, sites like Visit Britain[145] and more local sites like the Felixstowe Star[146] don't appear to make use of the service. The most impact on the topic seems to come from the annual publication of the MCS Good Beach Guide. That guide is produced using the samples collected by the Environment Agency and others but the data is not re-exposed or referred to directly. Curiously, the MCS highlights that the data it gathers from the Environment Agency is supplied under different (more restrictive) licence conditions than other data they use[147]. This is at odds with the Open Data Licence under which the data is published through http://environment.data.gov.uk/bwq/.

The published linked data does not appear to be used either by the European Environment Agency. Their State of Bathing Waters[148] visualisation does a similar job to the UK Bathing Water Quality Explorer but at a larger and therefore less detailed scale. The data made available from the EEA tool is provided as a set of Excel files.

[141] http://environment.data.gov.uk/bwq/widget/design

[142] http://www.goodbeachguide.co.uk/

[143] http://www.guardian.co.uk/environment/2013/jul/07/england-polluted-beaches-tide-of-filth

[144] http://www.independent.co.uk/news/uk/home-news/dont-come-on-in-the-waters-terrible-major-reduction-in-number-of-recommended-beaches-8550798.html

[145] http://www.visitbritain.com/en/Top-10-British-beaches/

[146] http://www.felixstowestar.co.uk/summer/beaches

[147] http://www.goodbeachguide.co.uk/goodbeachguide-data

[148] http://www.eea.europa.eu/themes/water/interactive/bathing/state-of-bathing-waters

II.10.2 Interview: LOGD supplier

Interview date	9/7/13
Interviewee	Alex Coley, DEFRA/Environment Agency, Chair UK Government Linked Data WG
Interviewer	Phil Archer

Q: How does the supply of Linked Data relate to the public task of your organisation?

A: The Bathing Water Quality data (BWQ) isn't just one dataset but rather a number of related data sets. The old way we would have published would have been 5+ separate data sets where any context would be lost. Lines are blurry between 'sets' and as we move to data as a service it can sometimes be hard to define what is a distinct dataset. We have a number of data sets as linked data but for example data.gov.uk (DGU) has no entry for BWQ because it doesn't fit the DCAT/CKAN model. We could bodge it but there's no current convention to work out how to do it in a meaningful way.

There's a set on environmental info regulations that underpin the work. Bits of legislation that describe the public task. These cover public register type activities – info we must provide plus public service type. We're trying to build data management activities that fit multiple uses, but we just have to manage once. LD is the technology that fits that.

The Environment Agency (EA) gets 20 – 40K Freedom of Information (FOI) / Environmental Information Requests (EIR) requests per year – more than any central gov department. Central gov FOI is generally about salaries, money spent on X etc. EA – the request are more in the form of what's the data that the supports this policy outcome. FOI is a burden in that sense. Need to provide data in a way that is usable and that underpins activities. People can build on top of it – so we don't get asked as much. Data is up to date and relevant.

If we change the data (fixing an error etc.), we don't delete the old – we update it and keep history about what we changed and why. Transparent history. We do it because in the past people have miss understood changes or NGOs have been critical about us changing data so we show why. E.g. lab results show that calibration of equipment was wrong.

Our widgets[149] are used by at least 10 local authorities. Some local authorities are very proud of their beaches. Previously they linked to the EA site. We created widgets so that local authorities can use it. The Marine Conversation Society (MCS) has a copy of the widget in their site but they also have an extra manual data supply. They have contractors. MCS don't get open data but they do have the widget. They take raw data behind the widget – there's a skills issue that prevents

[149] http://environment.data.gov.uk/bwq/widget/design

them taking the LOD. That's not ideal for us as there is a risk in understanding any data changes.

ARAP use the data in their Beach Selecta app. It uses our LD through the API, not SPARQL.

Q: Does your organisation have a "business case" document that motivates your investment in supplying Linked Data? Can you share it with us?

The important thing is to maximise use of our info and LD is part of that. We're about environmental outcomes that we want to achieve – decisions based on evidence – we want other people to base their decisions on evidence too so we make the data available for that. In the past we've spent money on building websites. At that point the business need has changed and we have to spend more on updating it. Now we have our data and we can build site and applications on top of that. Principle relies on us using open data standards. Means using things that allow us to link together. We link to Office of National Statistics and ordnance Survey data.

Open data Strategy is being rewritten. Likely to see much more linked data-based (not public yet).

Q: What are your future plans? Do you plan to expand or abolish the supply of Linked Data?

It's going to expand. The Evidence Director of the EA is very supportive. We'll work case by case. We're working on publishing and connecting code lists which we'll do using the new registry software[150] being developed by Epimorphics under the UK Gov LD WG. DEFRA is a large collection of organisations and this will help us to provide a managed service.

Q: Did the supply of Linked Data give rise to new opportunities for flexible data integration within your organisation and with partner organisations? Increase in data quality? New services? Cost reductions?

A: It allowed us to more explicitly expose the provenance of the data. If someone uses the data then we have some assurance that their data is up to date. URI structure gives some traceability. We're working with Rivers Trust[151] on catchment data – they're going to be publishing dome linked data as well. Catchment management should come from different data sources, using consistent ontologies etc. That's about meeting the public task and reducing costs. DEFRA has been funding the Rivers Trust to do this as well as training them/paying for them to be trained.

We did the BWQ work because it was the right thing to do. During that exercise, the history came up as a problem. We had a lot of water quality problems in 2012 due to excessive rainfall. The way we'd modelled the BWQ data meant that the history was there. People could refer to it, talk about our data – this was very helpful for our policy people. There was a shift in policy mid-season due to this, and that directly allowed us to change in the middle of an abnormal situation. We were

[150] https://github.com/der/ukl-registry-poc

[151] http://www.theriverstrust.org/

able to extend the data model which lead to an easy update of the website. That would have been a year's work under the old system. This time it took 3 weeks and cost under £10K – that's an unheard of level of responsiveness. No relational database to be updated.

Q: What are the main enablers / inhibitors for Linked Data to deliver value for its reusers?

A: Enabler – the directors – senior management bought into it. They don't get the detail but they do get the principles. Enables best use of data. Nigel (Shadbolt) came in to tell them how well they were doing. Senior management feel that they own it.

Skills and background of the IT department is the big inhibitor. IT departments traditionally focus on systems provision, traditional technology stacks and have vested interests in that persisting.

Q: Do you supply service level statements for Linked Data?

Not yet. We have a service where we are paying for it to be available. We have an SLA with our supplier, but that's not public yet. All services are hosted on Amazon Web Services with back up in case of DDoS attack. It's been remarkably resilient. The water catchment data we're developing is not performing well enough to make public yet. SPARQL is the problem. BWQ queries are less complex, but the explorer does more client side stuff, hits client much more.

Q: Does your organisation have a URI policy? Can we have a copy?

A: We have a framework policy. Not an environment.data.gov subdomain policy – but is in development. Things like what to do in our namespace. We're working on it – conventions coming from what we've done. (Editor's note: the UK advice on URIs for location[152] was done, and is being updated by, Stuart Williams of Epimorphics who was also the lead contractor for the Bathing Water Quality data modelling and Explorer application. The close relationship between DEFRA and Epimorphics is significant in this regard).

Q: Can you describe the Linked Data infrastructure of your organisation?

Jena/Apache hosted on AWS. It uses ELDA[153], the Epimorphics implementation of the Linked Data API, and will use registry service for code list repositories within next 4 months. DEFRA also uses TSO[154] , SWIRRL[155] and others through G-Cloud

Q: Which skills and competencies did you need in order to supply Linked Data? Do you have them in-house or contracted?

A: 99% skills are bought in. Also buying in training to upskill staff. It's been a hard slog. Initially it was down to an individual (Alex Coley) who provided vision. There was one person from the IT dept who was supportive from a strategic point of view. Alex's background: Atlantis project, a collaboration between OS and Met Office,

[152] http://www.data.gov.uk/library/designing-uri-sets-for-location

[153] http://www.epimorphics.com/web/tools/elda.html

[154] http://www.tso.co.uk/

[155] http://www.swirrl.com/

Centre for Hydrology and Ecology, etc. Worked on Digital National Framework – DNF which designed the use of TOIDS – persistent identifiers (URNs) used by OS. Wanted to reuse that rather than replicate everything. This became a reference for INSPIRE, became reference model for INSPIRE when it switched to use HTTP URIs. Detailed River Network for EA -> underpinning river data set, unique persistent identifiers (URNs) to be translated. So in that context it was natural to see what LD can do at the EA. BWQ was a case study. The EA found some money, John Sheridan offered some time which lead to the pilot which proved the usefulness for meeting a whole range of use cases. Then we had to find real money to do it bigger. We wanted to share data. We should use it ourselves as well as sharing it. The IT dept often had a focus on systems rather than a data centred focus of delivery.

The continued provision of the BWQ data has survived Alex leaving (although temporarily) the Environment Agency. He's still doing a bit of steering but not leaving. EA has a commitment to provide BWQ. The old system has now been switched off so the new one has to be maintained.

The ethos here is: Demonstrate benefit and then kill off other systems. There's a minimum 5 year commitment.

Q: Which activities do you carry out to supply Linked Data?

There are 3 tasks involved: designing URI space, vocabulary modelling, and publishing. Then we can build apps over the top of it. Pointing R2DQ[156] at a relational database isn't enough. It's not the whole thing. Sadly the traditional data architects don't yet all talk the same language.

Q: How many FTEs in your organisation are involved in provisioning the supply of Linked Data?

Several people. We will one in DEFRA dedicated to doing the registry work.

Q: Which costs have you incurred to publish Linked Data, maintain, and promote it? What are the trends?

A: We have spent money, but less than equivalent traditional processes. Costing a lot less than it would and cost is going down each time we do it – experience counts.

Q: Who are the main users of your Linked Data services? Is Linked Data only used by external customers or also for back-office exchange (e.g. with other agencies)?

BWQ is used by ourselves to produce our apps. Users don't care about LD. Response to a query can be a tweet with a URL that tells them everything they want to know. In this regard ELDA is an important and very powerful tool.

Q: How often is Linked Data used? What are the trends?

A: We have seen growth over a year.

[156] http://d2rq.org/

Q: How is the provisioning of Linked Data funded?

A: Tax payer

Q: Under which licence is Linked Data made available for reuse?

A: The Open Gov Licence – we're working on switching it to OGL2.

We expect to do some non-open LD, used for data integration within DEFRA. Reasons for not opening the data include national security, commercial and personal confidentiality.

Q: Which channels are predominantly used to consult Linked Data: Web API? Web site? (mobile) App? Data market?

BWQ Explorer is our main for now. This financial year should publish the full water quality archive with at least monthly updates. That covers 7K sites monitored multiple times per year back to 1950s. That will be published as LD with an explorer.

One aim is that FOI requests will self-serve.

The European Environment Agency uses the SPARQL endpoint. There's a reporting requirement that the EEA can take, pull not push. Dan Smith made a nice app[157] (a variation on the BWQ Explorer).

Main external user of the linked data is Arup's Beach Selecta app (see next interview).

We have explicitly not asked people to register. The only way we know who has used the data is look at the logs. From that we can see that some people suck in all the data. We're seeing increasing use of the widget. We can tell a widget or BWQ Explorer request, a Beach Selecta App etc. We did BWQ because we wanted the data, not worried about licences. Water Quality archive has licence issues.

Sites that use the BWQ service include beachalive.co.uk (see Figure 9). The seaside town of Bude has gone one better and, through A T Engineering[158], has installed a 24 inch touch screen showing the BWQ data.

[157] http://danpaulsmith.com/#projects

[158] http://www.atengineeringtadley.co.uk/blog/water-quality-monitoring-and-interactive-screen-enclosure-bu/

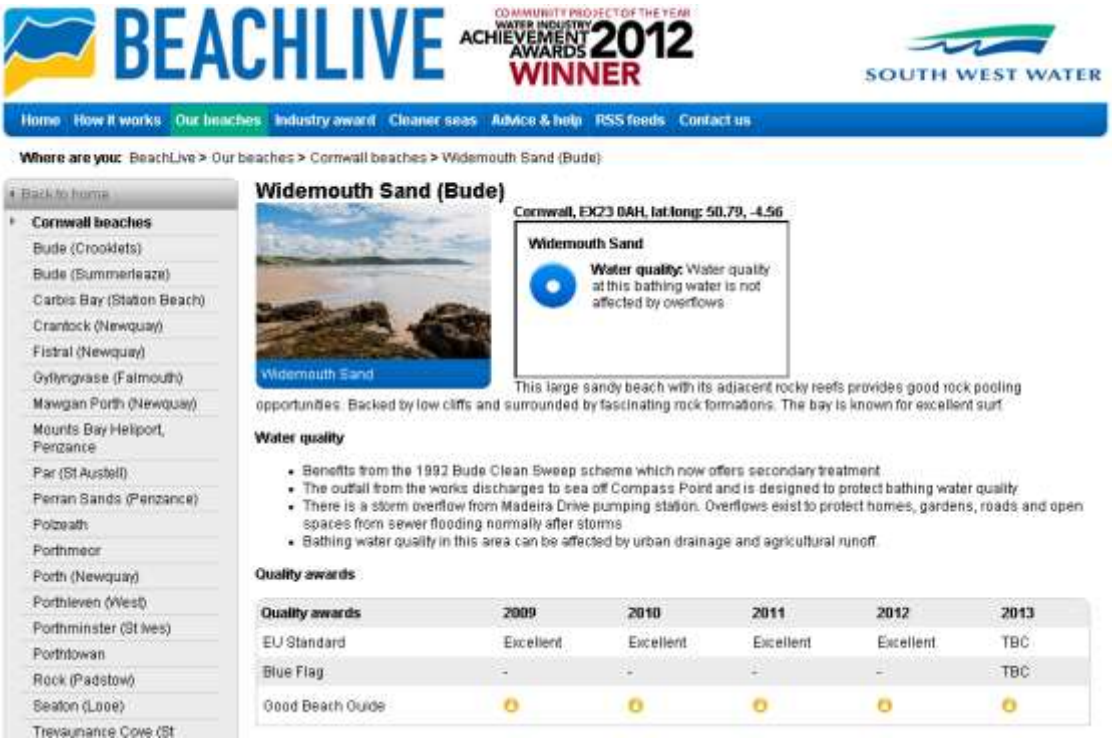


Figure 21 Partial screenshot from beachalive.co.uk that uses the EA's BWQ services (http://www.beachlive.co.uk/index.cfm?articleid=9511 taken 2013-07-29)

Q: Do you have a branding strategy for Linked Data?

The partnership licensing team paid for the production of the 'Powered by Linked Data Environment Agency' logo which is available in diff colours. ARAP wanted to show the provenance of the data – makes them look good and avoids endorsement perception.

Q: Do you invest in advertisement for Linked Data?

No. Government restricts marketing spend.

Q: Do you provide / make use of any mechanisms on user feedback or evaluation?

A little. Went through ARUP etc. and sought feedback from the developer community. We have a Web site with a feedback button. We intend to build more. Every page on the explorer has an e-mail address. It gets some usage, often querying detail on the data.

Interview Summary

DEFRA and the Environment Agency need a great deal of detailed information about locations, water and air quality, vegetation, land use and more. Since this

information is typically created and managed by disparate agencies, linked data is the obvious technology to improve efficiency. Alex's anecdote about being able to add a new feature in a few weeks rather than the more traditional large scale re-design and deployment of an entire system is worth bearing in mind.

As ever, the technology is only a means to an end. The provision of the widgets and the BWQ Explorer are the real story here, topped nicely by the interactive screen on Bude beach.

II.10.3 Interview: LOGD User

Interview date	22/7/13
Interviewee	Ewan Peters, GIS Technology Architect for Arup
Interviewer	Phil Archer

Arup[159] is *a global firm of consulting engineers, designers, planners and project managers* that delivers major infrastructure projects, sports facilities, municipal buildings and, of direct relevance here, water management installations. Arup does a lot of work with the Environment Agency and has been a strategic partner to Yorkshire Water since 1996[160]. In that context, the bathing water quality data is of great importance and an authoritative measure of high quality of bathing water can be seen as a sign of success for Arup's work.

Arup created the Beach Selecta application[161] and made it available on both the iOS and Android platforms. It offers information on the nearest beaches including:

- bathing water quality;
- the description of the site (also from the Environment Agency);
- the location of car parks displayed on a Google map;
- the local weather forecast;
- tide times;
- photographs of the area (from Flickr).

Q: How did the Beach Selecta application come about?

A: We started working with a developer and looked at the bathing Water Quality Explorer. It looked very interesting and we wanted to gain some hands-on experience of using the data and the technology

rkshire_Water.aspx

ourselves and seeing how easy it would be to integrate related legacy data. We also worked directly with Epimorphics (the company used by the Environment Agency to model and publish the BWQ data and build the BWQ Explorer).

Since the data and the approach looked interesting and seemed to have real potential in our wider work, we wanted to have a demo, an application that we could use to explore the potential of linked data for ourselves and our clients.

A question for us was: given the BWQ data, how easy is it to add in more? We found data that we wanted to include easily enough but we had to convert it to linked data ourselves. That extra data - things like the car park data - is not as up to date as the BWQ data. That's the good thing about the BWQ data itself - by using the SPARQL endpoint provided by the Environment Agency we always get the most up to date authoritative information.

The developer had an Oracle background and found it quite an easy journey to learn how to use linked data although I have to say - the documentation could be a lot better.

The app cost £13 - £17K (€15 - €20K) so not huge in the app development world. We've had around 1K downloads.

The app is free so there's no revenue for us here. What we're interested in really is the potential for using linked data in Building Information Modelling (BIM). At the moment, information about a building created at the design stage is not carried across to the construction phase and still less into the operational phase. We'd like to see if we can save money through more efficient data management and a linked data approach seems to have that potential. We get a positive response from our clients when we talk about this but there's a lot of resistance. The idea that you should make your data available to other people when you're really not 100% of what they'll do with it doesn't sit well.

We'd be interested to pursue the idea of vocabularies around building information modelling.

Interview Summary

The purpose of Beach Selecta is not to provide information about local bathing water quality - although it does that too. It's to provide experience of working with the technology and building an application that uses linked and other data sources. Can you build a good looking, easy to use application based on Linked Data? Beach Selecta proves that you can. The real test will be when the same techniques are used in Building Information Models.

II.11 UK – National Archives

II.11.1 Desk research

The initiative legislation.gov.uk is perhaps the benchmark against which any other linked open government data effort can be judged. Few will be match it in terms of the quality of data provision and engineering excellence.

Every piece of legislation passed and amended by any parliament in the UK, Scotland, Wales and Northern Ireland since 1267 is available as both a Web page and as XML, and is described in RDF. For example:

`http://www.legislation.gov.uk/id/ukpga/1998/29`

is the identifier for the Data Protection Act 1998. To get the act as XML append 'data.xml'. To see the RDF description of the act, append 'data.rdf'. To read the act, just put the URI in your browser. All data published on legislation.gov.uk follows the MetaLex standard[162] which is also used to provide a similar service for Dutch legislation[163]. The latter service is provided by the University of Amsterdam rather than the Dutch government which, for now at least, maintains its traditional data portal[164].

Legislation in any country is complicated. There are differences between an act being *passed* and being *in force*. An act can be revised or be directly affected and overturned by other legislation. There are differences between primary and secondary legislation, statutory instruments and so on. The URI schemes developed for legislation.gov.uk are consequently highly structured and detailed[165]. A feature of legislation.gov.uk is the way in which user agents are redirected when dereferencing a URI, handling British peculiarities like Regnal Years (the year of the reign of a particular monarch) which were used to date legislation prior to 1963.

The technology that underpins legislation.gov.uk is, of course, a means to an end, the end being to make legislation readily accessible to users, whether they be interested citizens or legal professionals. A crucial feature of the service therefore is the search function which can help to find specific pieces of legislation as well as carry out more advanced searches for legislation affecting a particular place, or that was in force at a particular time.

Head of Legislation Services at the National Archives, John Sheridan, wrote a blog post in March 2012[166] that is very pertinent to the current study.

[162] http://metalex.eu/

[163] http://doc.metalex.eu/

[164] http://wetten.overheid.nl/

[165] http://www.legislation.gov.uk/developer/uris

[166] http://digital.cabinetoffice.gov.uk/2012/03/30/putting-apis-first-legislation-gov-uk/

"We developed the API and then built the legislation.gov.uk website on top of it. The API isn't a bolt-on or additional feature, it is the beating heart of the service. […] Since launching the API we've seen several third party applications be developed, including two different iPhone / iPad apps, as well as innovative new products, such as a service for law lecturers to create and self-publish relevant extracts of legislation for their courses."

"The API is enabling us to develop an entirely new approach to updating legislation, inspired by the open source software movement – our Expert Participation Programme. We are inviting people from outside The National Archives to work with us, to apply changes to the legislation in the database. Quality is maintained through our editorial practice and a stringent process of review. It is an exciting time as private sector companies are now investing, employing people to work with us to bring the database up to date and to maintain it. The API gives them ready and easy access to the data, which they can include in their own products and services."

"The legislation.gov.uk API has changed everything for us. It powers our website. It has enabled us to move to an open data business model, securing the editorial effort we need from the private sector for this important source of public data. It allows us to deliver information and services across channels and platforms through third party applications. We are developing other tools that use the API, using Linked Data – from recording the provenance of new legislation as it is converted from one format to another, to a suite of web based editorial tools for legislation, including a natural language processing capability that automatically identifies the legislative effects. Everything we do is underpinned by the API and Linked Data. With the foundations in place, the possibilities of what can be done with legislation data are now almost limitless."

All data available through legislation.gov.uk is available free of charge under the **UK Government Licence for public sector information**. The licence document lists some types of data that are explicitly not covered and requires users of the data to give due attribution but in essence this is a very liberal licence that allows users to:

- copy, publish, distribute and transmit the Information;
- adapt the Information;
- exploit the Information commercially for example, by combining it with other information, or by including it in your own product or application.

II.11.2 Collected metrics

A news item published by the National Archives in June 2012[167] includes some indications of the scale of the operation.

[167] http://www.nationalarchives.gov.uk/news/732.htm

"The website receives more than 2 million monthly visitors and is many people's first port-of-call for looking up the law."

"The National Archives' in-house editorial team currently applies up to 10,000 legislative changes, called 'effects', to the database every year. This is the effect that a new law can have on existing legislation. However in the same period, the UK's Parliaments and Assemblies may make 15,000 new effects. As a result around half the laws on legislation.gov.uk are not yet up-to-date. Where outstanding effects are yet to be applied, this is clearly indicated on the website. By providing experts from outside government with the tools and training required to revise legislation, we will be able to ensure all legislation on the website is current, while at the same time retaining ultimate responsibility for the accuracy of the content."

II.11.3 Interview 1: LOGD supplier

Interview date	9/7/13
Interviewee	John Sheridan, Head of Legislation Services
Interviewer	Phil Archer

N.B. In this interview, John Sheridan made many references to a new service that will make the Official Gazettes (of London, Edinburgh and Belfast) more readily available. They have been available online for some time but not in a data-driven way that will be the case with the new system, due online in September 2013. Under the new system, notices will be published in HTML5 + RDFa. The RDF data will then be collated and made available via a SPARQL endpoint (and as a bulk download). This service was not online at the time of this study. A key feature of the new service is that by using a common structure for official notices, marked up with RDFa, insights can be gleaned from what had hitherto been unstructured text. In particular, the new system will create a *de facto* insolvency register.

Q: How does the supply of Linked Data relate to the public task of your organisation?

A: It allows us to fulfil our statutory duty to provide access to legislation and the UK's official gazettes. It allows us to ensure provenance of the info that we're making available so that it can be used as part of legal purposes or processes. No other tech gets you close. You can make nice structured data sets but nothing comes close to LD for structure.

Q: Does your organisation have a "business case" document that motivates your investment in supplying Linked Data?

A: For legislation, don't have a business case doc, but the essence of approach to managing legislation as data draws deeply on LD principles so there wasn't a business case to write. A business case for using linked data would like making a business case for using electricity. The business case is for the use of open standards. For gazettes, we have framed a proposition which is clearly enabled by

the use of LD. And there was a business case around the proposition. It can only be fulfilled by the use of LD. LD is not an end in itself, it's a means. You don't write business cases about your means, only your ends.

The work on legislation.gov.uk made the gazettes work easier. I wouldn't fancy writing a business case for supplying LD, It's fantastically useful but…

Q: What are your future plans? Do you plan to expand or abolish the supply of Linked Data?

A: Expand. We have a major service coming online later with the gazettes. Online from September, completion by Dec 2013 which will include UK company and solvency register but that will be the *de facto* corporate insolvency register. The UK has 250 different types of notice and the data we provide about them will be very rich. Transport, planning, the moving of car parking places etc. Common characteristics are facts that are contained in notices – formalised in (semi) structured text – that lend themselves to being expressed as structured data. You get value from the collection as well as each individual notice. Notices are admissible in court.

The aim is to remove reliance on immutability of paper as the way of demonstrating that something was published on a given date. Future plans are to expand what we provide for both legislation and gazette. No question of abolition.

We want to work more closely with partners, esp. European Publications Office.

The idea of URIs as European Legislation Identifiers is taking off – maybe not RDF – but there's activity in FR, IE, LU, DK & NL - not bad for something only agreed last autumn.

Q: Did the supply of Linked Data give rise to new opportunities for flexible data integration within your organisation and with partner organisations? Increase in data quality? New services? Cost reductions?

A: Yes. We have new services around provenance and authenticity. There are audit trails around process, assurance that the process has been followed so that the data you're using is the same as what's being laid before parliament, what you're signing. For the gazettes it's about due diligence that the person making the notice is who they claim to be. Those checks are recorded as part of the audit trial. Therefore, someone using the data can trust it more as they can check what processes it's been subjected to. I can't necessarily point to cost reductions within the National Archives any more than I can answer: How did your use of electricity save you money?

LD is inherently open standards based. Using open standards brings contestability and portability - and that saves you money. And yes you can quantify that. We've seen 30% improvements in services that we've procured by using open standards. Evidence for that is included in our submission to the UK government consultation on the use of open standards[168].

[168] https://www.gov.uk/government/consultations/open-standards-open-opportunities-flexibility-and-efficiency-in-government-it

Q: What are the main enablers / inhibitors for Linked Data to deliver value for its reusers?

A: Main inhibitors would be that people's knowledge is low, tools aren't amazing. LD enables us to make info available in many other formats that people do find valuable. It's a good core tech even if users aren't familiar with it. We need to be realistic about level of proficiency. Enabler - LD allows you to offer reusers choice as a core tech in a way other techs don't.

Q: Do you supply service level statements for Linked Data?

A: Yes. In that we publish the contracts we have with our suppliers and those contain SLAs. Aiming for services to be equivalent to ODI Expert certificate[169]. Guarantees and SLAs are v important.

Q: Does your organisation have a URI policy? Can we have a copy?

A: Yes and yes. We specify UK Gov URIs[170] in each contract in the areas that they cover. For gazette we also require there's a URI template. IETF work at the heart of ELI and this[171]. We're trying to eliminate the reliance on paper and persistence is at the heart of what the National Archives does of course so you can be confident that the URIs will persist as long as they're useful.

Q: Can you describe the Linked Data infrastructure of your organisation?

A: Legislation uses XML for docs, RDF for data and process. There are many people in the chain recording a 'fact'. The model for the gazettes is different. We're using HTML5 for the notices with data embedded using RDFa. We then extract RDF from the notice for ease of querying. For the Gazettes, data is a slave. Master is the HTML5/RDFa notice.

Q: Which skills and competencies did you need in order to supply Linked Data? Do you have them in-house or contracted?

A: Largely contracted. We needed to have sufficient knowledge to be able to secure the services which includes some hands on skill. We have employees who can write SPARQL.

Q: Which organisations are key partners in the supply of Linked Data?

A: Our contractors are TSO[172] (AKA Williams Leigh)

Q: Which investments has your organisation made to enable it to supply Linked Data?

A: Investment in people – knowledge and understanding. Government policies allow this to happen, e.g. ELI. We're spending public money to supply/use LD. Investments have not been made with a view to supply LD – they have been made to improve access to legislation/notices, and create a *de facto* corporate insolvency

[169] https://certificates.theodi.org/about

[170] http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-URI-sets-uk-public-sector.pdf

[171] http://tools.ietf.org/html/rfc6570

[172] http://www.tso.co.uk/

register. I am sure that the equivalent of millions of € have been spent on those 2 services (not tens of M). ROI has been doing things that wouldn't otherwise have been possible.

Q: How many FTEs in your organisation are involved in provisioning the supply of Linked Data?

A: For the 2 services: 1.5 FTE in house. Contractors ca 10 – 12.

Q: Which costs have you incurred to publish Linked Data, maintain, and promote it? What are the trends?

A: The Gazette contract is concessionary contract, so we make money, not spend it. For legislation.gov.uk we spend money. (The value of the contract is not made public, even under an FOI request[173])

Q: Who are the main users of your Linked Data services? Is Linked Data only used by external customers or also for back-office exchange (e.g. with other agencies)?

A: We use it all the time. It's about solving problems for us. A legacy service for gazettes has been used by some, e.g. Garlik around personal insolvency.

Q: How often is Linked Data used? What are the trends?

A: We're making more and more use of it.

Q: How is the provisioning of Linked Data funded?

A: legislation.gov.uk is part of TNA's services, funded by tax payer. The gazettes service is funded by contractor since posting notices in gazettes is a legal requirement in several situations and such postings are chargeable.

Q: Which pricing mechanism or other source of income exists?

A: All data is free in both cases.

Q: Under which licence is Linked Data made available for reuse? Can we have a copy?

A: The Open Government Licence[174]. We are in the process of moving to version 2 of that licence[175].

Q: Which channels are predominantly used to consult Linked Data: Web API? Web site? (mobile) App? Data market?

A: We use the API - the whole public Web site is based on it.

Q: Do you have a branding strategy for Linked Data?

A: No.

Q: Do you invest in advertisement for Linked Data?

A: No but there probably will be around the gazettes service when it comes online.

[173] http://www.nationalarchives.gov.uk/foi/costs-running-official-documents-per-annum.htm

[174] http://www.nationalarchives.gov.uk/doc/open-government-licence/version/1/open-government-licence.htm

[175] http://nationalarchives.gov.uk/doc/open-government-licence/version/2/

Q: Do you provide / make use of any mechanisms on user feedback or evaluation?

A: Yes – for both sets of services we obtain user feedback. We have an expert participation model using LD tools – experts contribute back. E.g. by identifying when one piece of legislation affects another, where it affects it and so on.

Q: What is the number of de-referenced URIs / queries?

A: In the region of 5M/week (including content negotiation)

Q: What were the costs?

A: There was the original development cost: the cost of all activities that were required to identify, model, transform, harmonise and publish Linked Data.

Q: What is he annual maintenance cost?

A: The cost of publishing updates of the data, maintenance costs of relevant infrastructure is around 0.5M/year

Q: What are the promotion costs per year: the cost of promoting the availability of the data as Linked Data?

A: Nil

Q: What is the number of linked datasets (outgoing and incoming links)?

A: Not many. Hard to find other sources of info that you trust. The gazettes link to legislation.gov.uk and Companies House URIs. They also link to other data sets. legislation.gov.uk us starting to link to Publications Office using ELI. For transposition and implementation. Gazette will use Ordnance Survey data (OS data not needed for legislation.gov.uk). It's more likely that links will go from OS to legislation.gov.uk for defining boundaries.

Q: What is the number of derived applications?

A: This is more likely for gazettes. Applications use the doc oriented approach such as iLegal[176], MobileLegislate[177] and Longman Law Bespoke. These are all different types of commercial product or service but they don't make use of the linked data.

Q: Can you estimate the number (and increase) of corrections requested; and cost reductions on information integration?

A: Linked data allows you to do things you couldn't so otherwise. See benefits in terms of cost.

Linked data is a means, not an end.

Interview summary

The main point that John Sheridan emphasises repeatedly is that the use of linked data is entirely practical. No other technology comes close to providing what the National Archives need: being able to provide data about legislation at the level of the Act itself down to individual paragraphs. legislation.gov.uk was built as a linked

[176] http://ilegal.org.uk/

[177] http://mobilelegislate.com/

data installation and the Web site you see is just a user of the API. Whether anyone else uses the API is completely unimportant - the task is to provide trustworthy, accurate information about legislation. There are users of the information provided on legislation.gov.uk but no known users of the data as linked data.

For the new gazettes service, the infrastructure is different, being based on text written in HTML5 and enhanced with RDFa. By harvesting the RDFa it is possible, for the first time, build a *de facto* company insolvency register and other services.

II.12 UK – OpenCorporates

II.12.1 Desk research

At the time of writing, OpenCorporates offers data concerning over 55 million registered companies with more being added all the time. The data comes from multiple sources. Some is taken directly from published company registers, such as in the UK and Alaska, other data is screen scraped (occasionally with the help of ScraperWiki). In some cases the data is collected and re-published with the express permission of the relevant company register, in other cases OpenCorporates has just gone ahead and done it, so far without negative consequences for anyone.

OpenCorporates does not use linked data in its own infrastructure - the data is actually held in a relational database. Its API returns JSON by default but can also return XML - but not RDF. RDF data *can* be obtained directly from OpenCorporates through content negotiation when dereferencing individual company URIs or by simply appending a company's URI with '.rdf'. In this sense should therefore OpenCorporates not (currently) be seen as a publisher of linked data. The RDF data that is available, however, does make use of the Registered Organisation Vocabulary (see Figure 22).

A notable feature of OpenCorporates is that it crowd sources extra data too. On a simple level, users are able to add details such as a company's Web site, postal address etc. There is a link on each company page through which problems with the data can be reported. On a more complex level, the community is encouraged to indicate that a company is a member of a particular group.

Almost all major companies are actually networks of companies, some with complex ownership rules and hierarchies. What Wikipedia calls 'BP'[178] comprises various subsidiaries in different jurisdictions around the world. The precise relationship between those units is not in the public domain (and one suspects in a company the size of BP may not even be known internally with any degree of certainty) however the fact that Burmah Castrol in some way has membership of BP is useful information and it's this undefined 'grouping' that OpenCorporates offers. The list of companies in any one group may be incomplete and, like all crowd-sourced data, is subject to errors, but the crowd itself is usually good at reporting those errors so that, to continue the example, the list of 8 companies shown as being members of the BP group is very likely to be accurate and may or may not be complete. Group data is available in the same multiple formats as individual company data.

[178] http://en.wikipedia.org/wiki/BP

```
<rdf:Description rdf:about="http://opencorporates.
  <rdfs:label>
    APPLE BINDING LTD
  </rdfs:label>
  <rdf:type rdf:resource="http://s.opencalais.com/1
  <rdf:type rdf:resource="http://www.w3.org/ns/rego
  <opencorporates:legalName>
    APPLE BINDING LTD
  </opencorporates:legalName>
  <rov:legalName>
    APPLE BINDING LTD
  </rov:legalName>
  <opencorporates:companyType>
    Private Limited Company
  </opencorporates:companyType>
  <rov:orgType>
    Private Limited Company
  </rov:orgType>
  <opencorporates:companyStatus>
    Active
  </opencorporates:companyStatus>
  <rov:orgStatus>
    Active
  </rov:orgStatus>
  <rov:registration rdf:resource="http://opencorpor
```

Figure 22 Partial view of the RDF data made available by OpenCorporates. Highlighted terms are from the Registered Organisation Vocabulary

Groupings are important, especially when trying to track down where corporate money goes. Politicians around the world are tackling the problem of large companies using tax havens to hide their profits and it's instructive to see company relationships where possible. Making connections requires detailed investigation and access to the kind of data that, currently, only OpenCorporates provides coupled with filing data that may need to be paid for and read by humans. In a talk given to the Personal Democracy Forum 2013[179, 180], Chris Taggart explained the effort required to take the simple registration of Facebook Inc. and discover its network of companies.

[179] slides: http://www.slideshare.net/countculture/understanding-corporate-networks

[180] Video http://personaldemocracy.com/media/mapping-corporate-graph

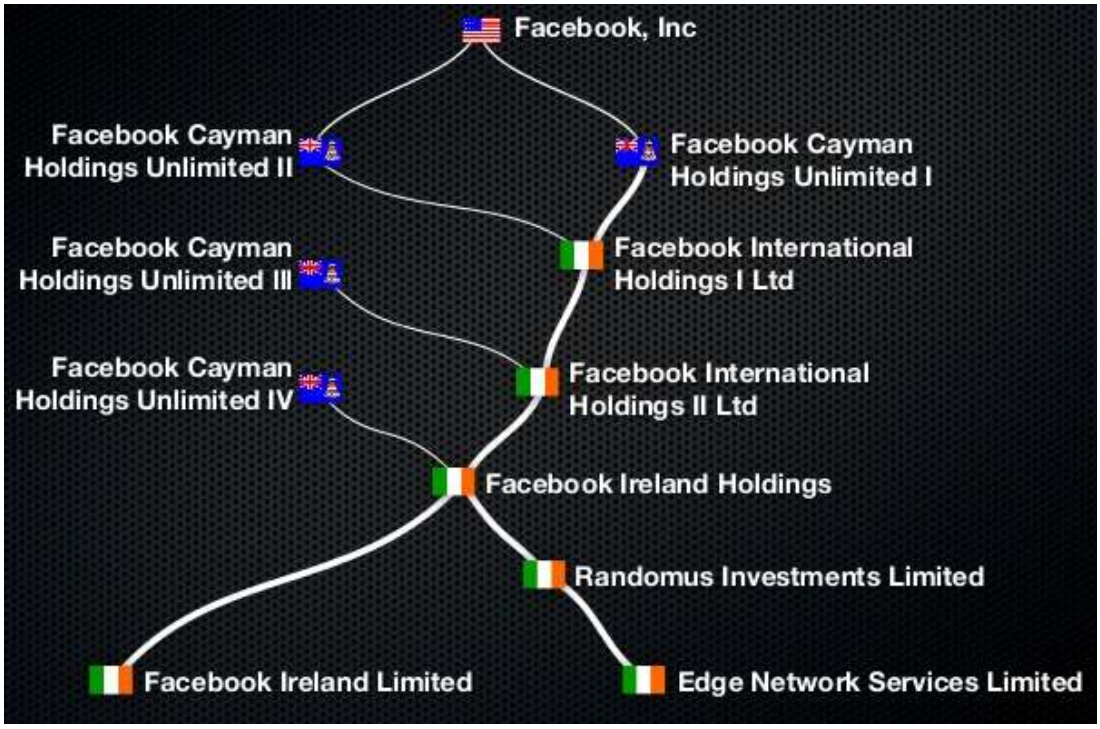


Figure 23 OpenCorporates' visualisation of the structure of Facebook

It should be noted that OpenCorporates' company grouping feature is in the process of being replaced with Corporate Networks Data[181]. TO find a company's network, append /network to that company's page. For example http://opencorporates.com/companies/us_de/4767700/ is the page for the State Street AIS Europe LLC - which is registered in the US state of Delaware. Add '/network' to its URI to see the diagram shown in Figure 24. The diagram is interactive so that if you click on the parent company, you see its full tree… which won't fit on this page in any meaningful way.

[181] http://blog.opencorporates.com/2013/07/11/open-corporate-network-data-not-just-good-but-better/

Figure 24 Partial screenshot of the network diagram for (Delaware-registered) State Street AIS Europe LLC

The work on networks has been done largely by hand and funded by a grant from the Alfred P Sloan Foundation. The original work on groupings may be kept, at least in part. It's essentially a mapping between companies and the Wikipedia article about them, but expect OpenCorporates expects the primary grouping of the entities that make up a large corporation to be using the corporate network feature, which is far more granular, has detailed provenance, confidence, and time dimensions.

It is possible to generate insights purely by programmatic means as Tony Hirst showed when he linked the OpenCorporates data to trademark data published by the World Intellectual Property Organisation[182] and when researching Tesco[183].

In addition to the main RESTful API offered by OpenCorporates, it also runs a reconciliation API designed principally for use in Google Refine but available for use

[182] http://blog.opencorporates.com/2012/12/17/guest-post-data-sketching-with-the-opencorporates-api/

[183] http://blog.ouseful.info/2012/04/12/mapping-the-tesco-corporate-organisational-sprawl-an-initial-sketch/

directly. The Open Knowledge Foundation uses the reconciliation API in its openspending.org project to match transactions to individual companies[184].

All data published by OpenCorporates is available for use free of charge under a share alike licence. Companies wishing to augment the data and not share their work are charged a fee. Use of the APIs is also free but it rate limited, particularly for anonymous users.

II.12.2 Interview 1

Interview date	14/6/13 plus e-mail follow up
Interviewee	Chris Taggart, co-founder
Interviewer	Phil Archer

It's worth repeating that OpenCorporates makes no use of linked data internally. Neither is it a public sector body and therefore many of the questions asked of other interviewees are not relevant.

Q: Why does OpenCorporates publish linked data?

A: Why wouldn't we? It's easy to include it as one of the format options and we want to reach as many people as possible. We don't make a special case for it any more than we do for the JSON or XML versions of the data, however.

Q: Are you aware of anyone using the linked data that you publish?

A: Not as such, no, no. The primary access route for our database is via the API (which returns JSON by default and XML if asked). The Google Refine Reconciliation Service is used more than the RESTful API.

Q: You recently updated the RDF data to reflect the changes in the Registered Organisation Vocabulary. Did anyone get in touch to ask about/object to that change?

A: Surprisingly, no. I made the change when the vocabulary was updated (I'd helped create the original Core Business Vocabulary) and worried that it might have an adverse reaction. The fact that it didn't may or may not be a god thing.

My attitude to linked data is that I know it's there and if the demand for it is there, or if it makes my core job easier, then of course I'll use it. For our current work there's no advantage in using it so we don't. What's important is making the data available in the first place and then using that to draw insights - which we do - see the work we're doing now on Open Data Corporate Networks for example (see Figure 25).

[184] http://api.opencorporates.com/documentation/Google-Refine-Reconciliation-API

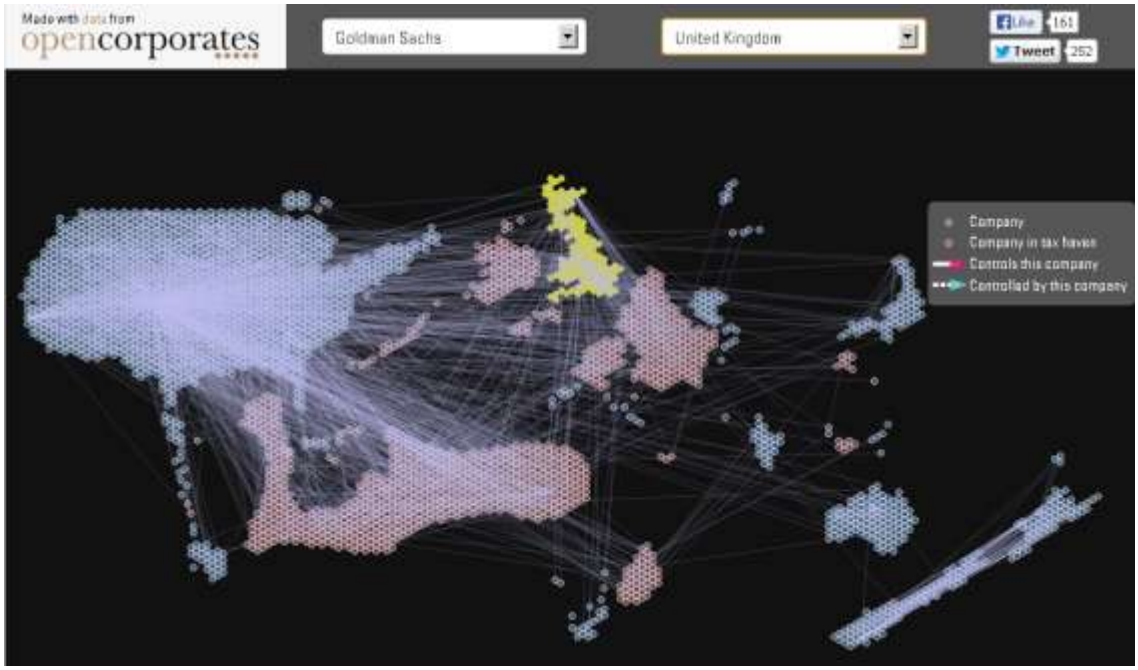


Figure 25 Screenshot of OpenCorporates visualisation of the structure of Goldman Sachs

There are some really good visualisation tools and code libraries available to create things like that and they use either JSON or CSV, not linked data (see Figure 26). Working out the structure of a corporation like Goldman Sachs, or Starbucks or Facebook is a labour-intensive activity that requires the skills of a detective, not a developer.

Q: There's currently one example of a company register that is itself published as linked data: do you make use of the Companies House linked data?

A: We make use of the data and we link to it, including owl:sameAs triples, but no, we don't make use of the linked data that they publish.

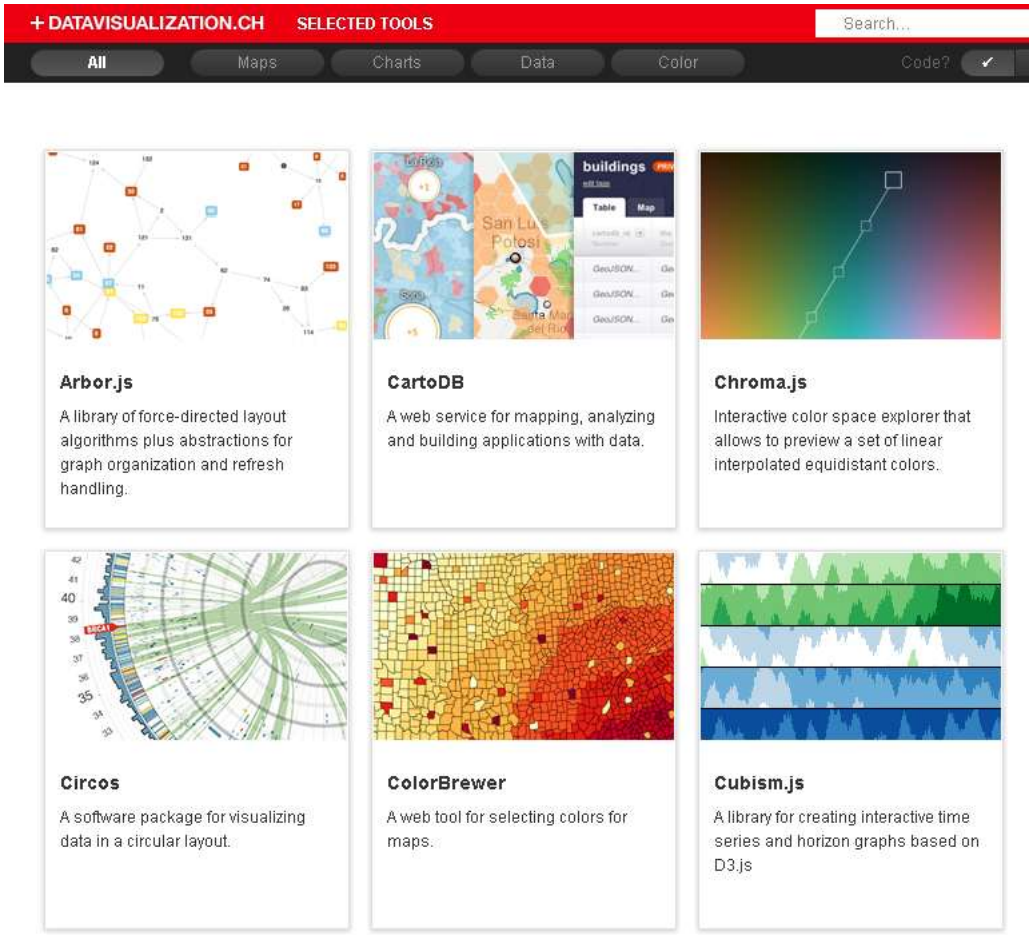


Figure 26 Partial screenshot of the catalogue of data visualisation tools highlighted at http://selection.datavisualization.ch/

Interview Summary

OpenCorporates is an exceptionally important venture with the potential to make a real impact on the way businesses around the world are regulated. The value comes from the fact that the data is open and free, and that its work is entirely transparent. The visualisation work makes the power of the information obvious to all. Whether its provision of its data as RDF proves important in the long run remains to be seen.

II.13 UK – Ordnance Survey

II.13.1 Desk research

The **Ordnance Survey**[185] is Great Britain's national mapping agency. The agency is responsible to collect, maintain and distribute geographic information (GI) of England, Scotland and Wales to government, business and individuals.

The Agency generates revenue through licensing the intellectual property rights. This is done directly with customers and via more than 500 private-sector partners.

Via OS OpenData, Ordnance Survey provides free of charge **linked open data**[186] offerings from Boundary-Line, Code-Point Open and the 1:50 000 Scale Gazetteer, which provide a Uniform Resource Identifier (URI) for every postcode, administrative region and civil voting area of Great Britain.[187]

Value proposition and licensing usage conditions of Linked Open Data products

The following products are licensed under the OS 'OpenData Licence' **(Free of Charge)**

- **OS OpenData:** is the opening up of Ordnance Survey data that makes a number of datasets available as open linked data, free of charge under the terms of an OS OpenData Licence. These datasets – which include raster and vector mapping, height, boundary and gazetteer products.
- **OS OpenSpace:** is a service free of charge that enables web applications to be built using datasets, such as the OS Open Data, that are hosted by Ordnance Survey.

In general, the OS OpenData Licence Terms and Conditions[188] allow users to:

- copy, distribute and transmit the data;
- adapt the data; and
- exploit the data commercially, whether by sub-licensing it, combining it with other data, or including it in your own product or application.

These are conditioned to:

- **Attribution**: Acknowledge the copyright and the source of the data by including the following attribution statement: 'Contains Ordnance Survey data Crown copyright and database right 2011'.
- **Share-a-like**: Include the same acknowledgement requirement in any sub-licences of the data that you grant, and a requirement that any further sub-licences do the same.
- **Endorsement**: Ensure that you do not use the data in a way that suggests Ordnance Survey endorses you or your use of the data.
- **Misrepresentation**: Ensure that you do not misrepresent the data or its source.

[185] http://www.ordnancesurvey.co.uk/oswebsite/about-us/index.html

[186] http://www.ordnancesurvey.co.uk/oswebsite/opendata/linkeddata.html

[187] http://www.ordnancesurvey.co.uk/oswebsite/docs/annual-reports/ordnance-survey-annual-report-and-accounts-2011-12.pdf

[188] http://www.ordnancesurvey.co.uk/oswebsite/docs/licences/os-opendata-licence.pdf

Other data products and licensing usage conditions

There are variants of standard licences offered by Ordnance Survey[189] that regulates and prices the use and **access to premium paid-for data**. Data and services can either be licensed directly from Ordnance Survey or through a Licensed Partner who can supply both basic products and value-added solutions utilising Ordnance Survey data. Additional rights apply to Infrastructure Body (Utility); or a Public Sector Body (licensed under collective purchasing agreements). Premium paid-for products are **not yet available as linked data**.

Key findings

- Basic linked data free of charge.
- Premium for non-linked paid-for data.
- API for non-commercial purposes gives access to limited linked data.

II.13.2 Interview 1

Interview date	15/7/13
Interviewee	John Goodwin, Principal Scientist, Ordnance Survey Research;
Interviewer	Phil Archer

Q: How does the supply of Linked Data relate to the public task of your organisation?

A: In 2010 we were required by the Cabinet Office to open at least some of our data (and subsequently contracted to do so). The decision to use linked data was heavily influence by Tim Berners-Lee and Nigel Shadbolt. We opened three data sets as LD:

- 1:50 000 Scale Gazetteer;
- Code-Point Open (UK post codes);
- administrative geography gazetteer for Great Britain.

Linked Data per se doesn't really serve as part of our public task. However it is our public task to map administrative boundaries.

Q: Does your organisation have a "business case" document that motivates your investment in supplying Linked Data? Can you share it with us?

A: Not relevant in our case - we were simply paid to do it.

However... now it's in place we're looking at the proposition that linked data offers the rest of OS and our services. It's not a big overhead to run the linked data service, it's one of several data formats we use to publish.

[189] http://www.ordnancesurvey.co.uk/oswebsite/licensing/agreements.html

Q: What are your future plans? Do you plan to expand or abolish the supply of Linked Data?

A: We plan to expand our use within digital products and research. We're starting to pull in government data to enrich our geo data.

One issue we're tackling at the moment is tracking changes over time. There's no defined way to do this so we haven't been able yet to embark upon the notion of life cycles for the data.

One thing that makes it harder for us is that you don't need to register to use the LOD, which you do to access our other free data services. We can see server stats but they only tell you so much. We keep in touch with our customers/users via blogs[190] and forums.

Q: Did the supply of Linked Data give rise to new opportunities for flexible data integration within your organisation and with partner organisations? Increase in data quality? New services? Cost reductions?

A: I certainly find it useful and would find it hard to go back to relational databases. It's also easier - the Ordnance Survey on Demand Service, Web Map service etc. come from data that has to be loaded into specific silos for the job. We're starting to see internal interest in using the post code linked data, i.e. outside the silos. In that sense we are starting to see benefits internally.

Q: What are the main enablers / inhibitors for Linked Data to deliver value for its reusers?

A: A lot of developers don't like it. If they can't start using it within an hour they'll ditch it - and it takes longer than that to learn how.

It has the appearance of being harder than it really is though with more of an academic community cf. the JSON-centric developer community.

There aren't the resources to take people through the learning curve during a hack day. Our community is GIS oriented. The resources that are available seem to all expect too high a level of existing knowledge.

We're also lacking guidance on what vocabularies to use. How does anyone know that FOAF and DC Terms are the ones to use unless you already know? We could do with some help to get the community going.

As for enablers… it's much easier to manage. Ordnance Survey and Transport for London can both manage their data separately - no need for anyone to download and integrate different data sets manually. It would be a lot easier to work out the location of bus stops by post code if the bus stop data were available as LD as well, for example - you could then just throw a query at it, no need to do any more.

It's hard for us to point to other real tangible benefits though.

The overall community is growing, not diminishing so it makes sense for us to make our data available as LD.

[190] http://blog.ordnancesurvey.co.uk/author/dee/

Our linked data was originally hosted by Talis so when their linked data platform folded we hired three of their ex-staff to help us build our own infrastructure. It's now all hosted on Amazon Web Services using open source software. We're now increasing our internal skills but it remains true that very few people can take a dataset, develop a linked data model, set up and manage the hosting for it.

Before the original 2010 contract we had one person at OS with knowledge of linked data. We'll be adding another 6 soon and will be up to 20 by the end of 2013.

Q: Do you supply service level statements for Linked Data?

A: Not since Talis folded. We only have staff to monitor the servers in office hours. Providing an SLA adds an overhead. We might set up SLAs with some of our customers, like the Environment Agency and DCLG (Department of Communities and Local Government) but we haven't done so yet.

Q: Does your organisation have a URI policy?

We have an identifier policy of which the policy on URIs forms a fragment.

Q: Which activities do you carry out to supply Linked Data?

A: Going back in time, the research department did linked data for a while. We had released boundary data before the UK Government's open data policy came in. We had to do some modelling of course and found that we had to mint several new predicates to do that – it was not very open. Now it's a lot easier and more automated. Having the Code Point data (postcodes) in place definitely helps. We're now trying to make sure that we have a repeatable process which will help lower the costs.

It should be a lot easier to develop new products. Adding a new column on a database can take up to 5 years for customers to adjust – now we can offer product enhancements very easily. We're using our LD infrastructure as a test bed for new products of our own.

Q: Which investments has your organisation made to enable it to supply Linked Data?

A: It's all part of the contract with the Cabinet Office – impossible to break down and identify how much the linked data aspect cost.

Q: How often is Linked Data used? What are the trends?

A: We don't see a lot of repeat users. People come in, have a look and go away, but the trend is upward. We notice a lot of hits from the NHS and the trend there is upwards.

Q: If the Cabinet Office stopped paying you, would you continue using and publishing linked data?

A: Probably

Q: Which pricing mechanism or other source of income exists?

A: We do have chargeable products, like those based on OGC standards, but the linked data is all available for free. Our biggest request actually is for a complete dump of our data.

Q: Do you invest in advertisement for Linked Data?

We do spend money to promote the data, yes. GeoVation[191] encourages the use of linked data (see Figure 27). We give financial rewards of around £100K twice a year to the top submitted projects (as judged by peer review). The challenges are around different themes like transport, environmental data etc. We've run hack days too, for example with the Environment Agency.

We generate interest through Twitter too (John Goodwin 1600+ followers, Leigh Dodds 1800+). John often asked to push linked data by developer community.

Figure 27 Partial screenshot of http://www.geovation.org.uk/ (24/7/13)

I built an app that finds pubs in low crime areas – something that, yes, you could do with GIS, but not as easily.

A final soft benefit is that our work on linked data shows us as thought leaders.

Interview Summary

Ordnance Survey was the first major UK public sector body to publish linked data. They did it because they were paid to and the advice from the Cabinet Office was strongly influenced by sirs Tim Berners-Lee and Nigel Shadbolt. It was tough at first. They needed the support of an external specialist provider. When that outside business folded, OS hired some of the relevant people from that company as consultants to help them move to an open source, cloud-based solution. Now that the LD infrastructure is in place, they're beginning to find new uses for the data themselves. They recognise significant potential internally. Users of the data outside OS are the Environment Agency (see DEFRA case study) and the Department of Communities and Local Government.

[191] http://www.geovation.org.uk/

II.13.3 Interview 2: Ordnance Survey Data User

Interview date	26/7/13
Interviewee	Steve Peters, Strategic Statistics Division, Department for Communities and Local Government
Interviewer	Phil Archer

Steve Peters is a developer in his spare time and makes a living at the Department of Local Government and Communities, DCLG. Along with Paul Davidson of Sedgemoor District Council, he's a key person in the open data movement within UK local government and is responsible for the Open Data Communities Web site[192]. The site makes extensive use of linked data, whether DCLG's own or other people's, such as Ordnance Survey's, as well as non-linked data APIs. An important example of the latter is the Nomis API[193] that provides access to extensive data about the Labour Market, and outputs from the 2011 Census. Nomis is run by the University of Durham on behalf of the Office of National Statistics.

DCLG sees itself as the 'department for place'. The Open Data Communities site is therefore offered as a service to find out what DCLG knows about a given local area. DCLG is also actively promoting Open Data Communities across central and local government, persuading and encouraging others to employ Linked Data standards and technologies in their systems and outputs. The objective here is to further extend and enable effective use of the "web of data" about local areas – e.g. to combine DCLG and other related data on the local labour market, housing and skills (i.e. some of the data within the Census), or street-level crime.

As part of its communications toolkit, DCLG has developed various demonstration applications to show and tell the benefits and opportunities arising from inter-linking multiple sources using open standards. At the time of our discussion, Steve was working on understanding the Police API. Two days later… the application was available (see Figure 28).

The best source of information about Steve's work, often supported by Swirrl[194], is his blog. The short blog entry about this application is worth repeating here more or less verbatim:

[192] http://opendatacommunities.org/

[193] http://www.nomisweb.co.uk/api/v01/help

[194] http://www.swirrl.com/

Figure 28 Data being loaded from the Police API into the Local Crimes Application

Quoted blog post[195]

Exploring local crime data via data.police.uk

Posted on July 28, 2013 by openviz

I noticed last week that the good folk at data.police.uk have recently introduced some new features in their API, in particular the ability to get street-level crime data for custom geographic areas. So, I decided to give that a try, focussing on the possibilities for querying for crime data in Wards. The result is this application.

The app opens with drop down lists of:

- Dates (year and month, as "YYYY-MM") for which street-level data is available via the Police API
- All Local Authorities in England – which I retrieve by querying DCLG's OpenDataCommunities
- Wards falling within the selected local Authority.

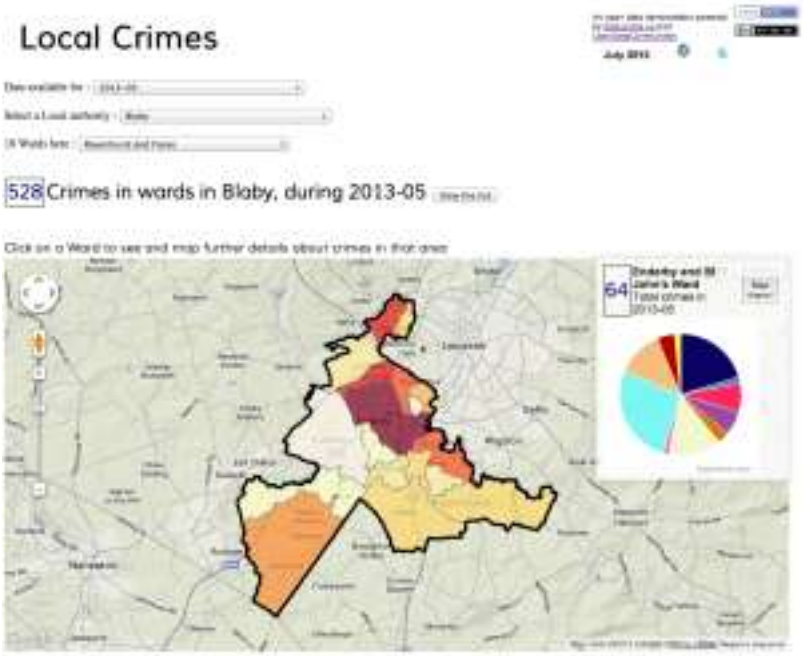
The app defaults to Adur district council, the first entry in the alphabetical list, but you can select any council from the drop down. On selecting a local authority, the app automatically discovers associated Wards by querying boundary files that I've

[195] http://openviz.wordpress.com/2013/07/28/exploring-local-crime-data-via-data-police-uk/

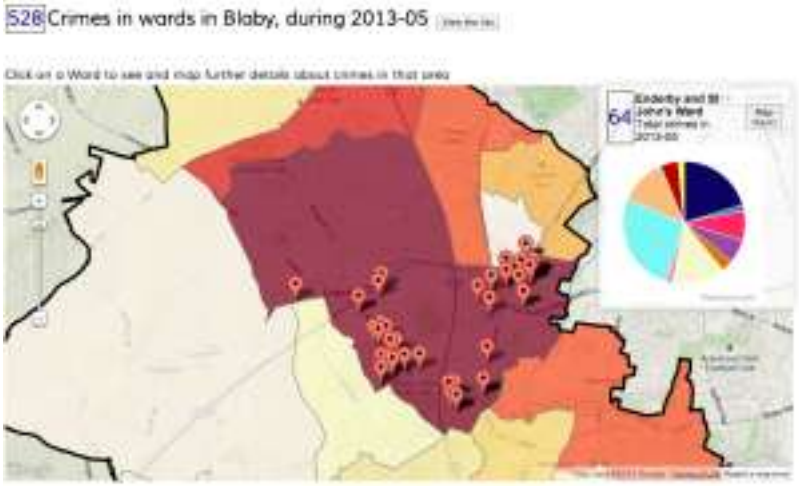
acquired and processed from ONS's Open Geography Portal : specifically, the December 2012 set, accessible via the 'Download products' tab.

The app then cycles through the list of wards, querying data.police.uk to retrieve the list of street-level crimes in each ward, for the selected month. The result is a visualisation with three main elements.

First, we have a map of wards in the selected district, shaded according to the total number of crimes listed in the given month. You can click on any Ward, and see more detailed information (in a pie chart) about the number and type of crimes in that area.

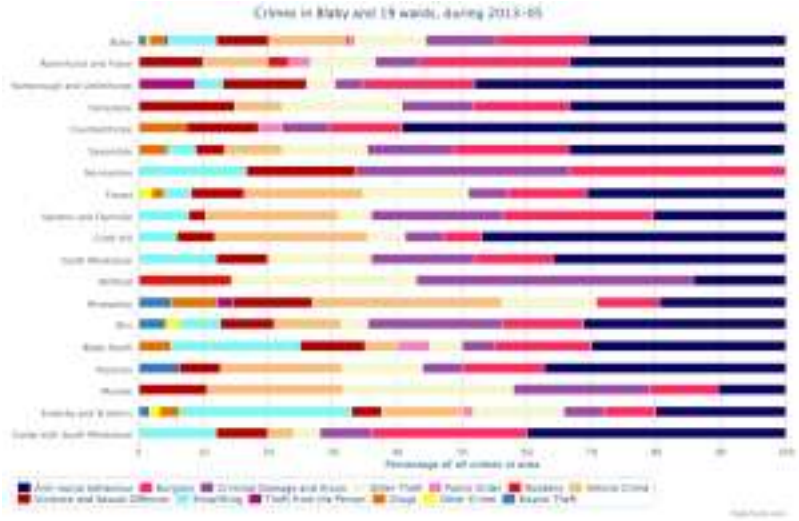


Next to the pie chart, you'll find a button to map individual crimes within the selected ward.



Hovering over a map points will display details about that crime's category and location.

The second main element is a bar chart showing the percentage distribution of different crime categories in the select month, across the selected authority and its constituent Wards.



The third element lets you explore the full list of crimes in the selected authority, by clicking on the "view the full list" button – located in the title just above the main map. This opens a separate window, containing a table that you can sort by clicking on column headings, and filter by typing key words in the search box.

Caveats and conclusions

As the Police API does not allow custom areas larger than 20 square kilometres, the app is unable to retrieve data for a number of wards – particularly in the more rural areas. In these cases, wards will be shaded grey, with no data on street-level crimes.

It may also fail to load data for a few local authorities, due to some residual bugs with the underlying map data for those areas. I'm working on sorting that soon.

Oh, and it doesn't work in Internet Explorer version 8 or earlier. I'll spare you the horrible technical explanation.

Apart from that, it seems to works fairly well, and has given me a shopping list of new features for the Police API.

I also have ideas on combining crime data with related public sources that are available now via third-party APIs. Watch this space, and in the meantime please do try it out and let me know what you think.

End quoted blog entry

The timescale between "I'm just trying to get my head around the new Police API" on Friday lunch time and this app appearing two days later - even allowing for Steve's passion for developing over a hot summer's weekend - is remarkable. The key is in the fact that a lot of groundwork was laid a month earlier with the development of Parish Areas application (Figure 29).

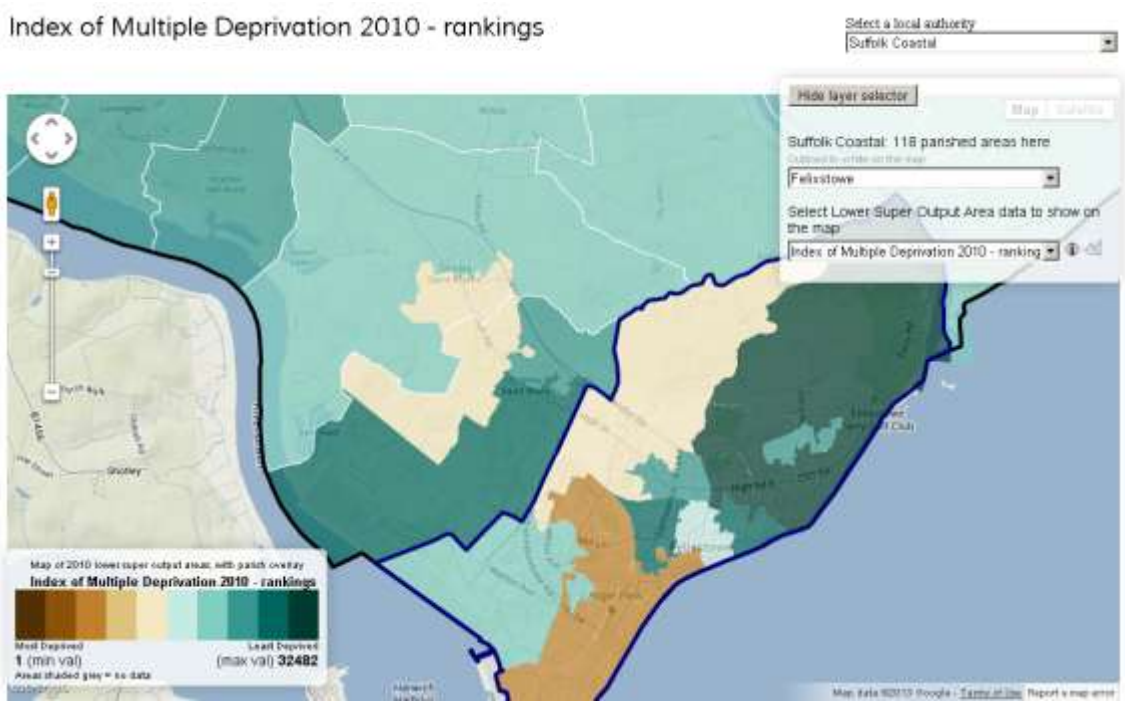


Figure 29 Screenshot of Parish and Local Demography App (http://dclgexamples.mywebcommunity.org/npf/parishdata-v3.html)

This application uses Ordnance Survey data, ONS data etc. and provides a useful view of the local demography, deprivation index etc. In order to add in police data, all that was necessary was to understand how to query the police data API, which as Figure 28 shows is done live every time you use it to ensure that the latest data is used. The amount of time spent developing the new local crime map was almost *all* about understanding the police API and very little about developing a new app.

Steve Peters' job would have been a lot easier if the police data were made available as linked data in the first place. Linked data allows you to break out of API

silos - the police API is not the same as the Nomis API and so on. "Some APIs are almost spitefully different," said Steve at one point. If an organisation is going to go to the bother of modelling its data and making it available through an API, the extra effort to make it available as linked data is trivial. Linked data comes with its own standard functions, especially when using the Linked Data API (which a lot of UK government sites do). Steve said he'd like to see greater implementation of federated SPARQL[196] - i.e. the ability to send one query to lots of different endpoints at once.

"At DCLG we really want to engage with partners to solve problems like providing care for the elderly, relieving homelessness and so on. We can't do that without the data being available. We don't know what the Department of Health knows - but our local services would be improved if we could know."

One site that *has* combined health data with some of our local data is Lambeth in Numbers (Figure 30). Lambeth Council gets its local deprivation data from the DCLG triple store

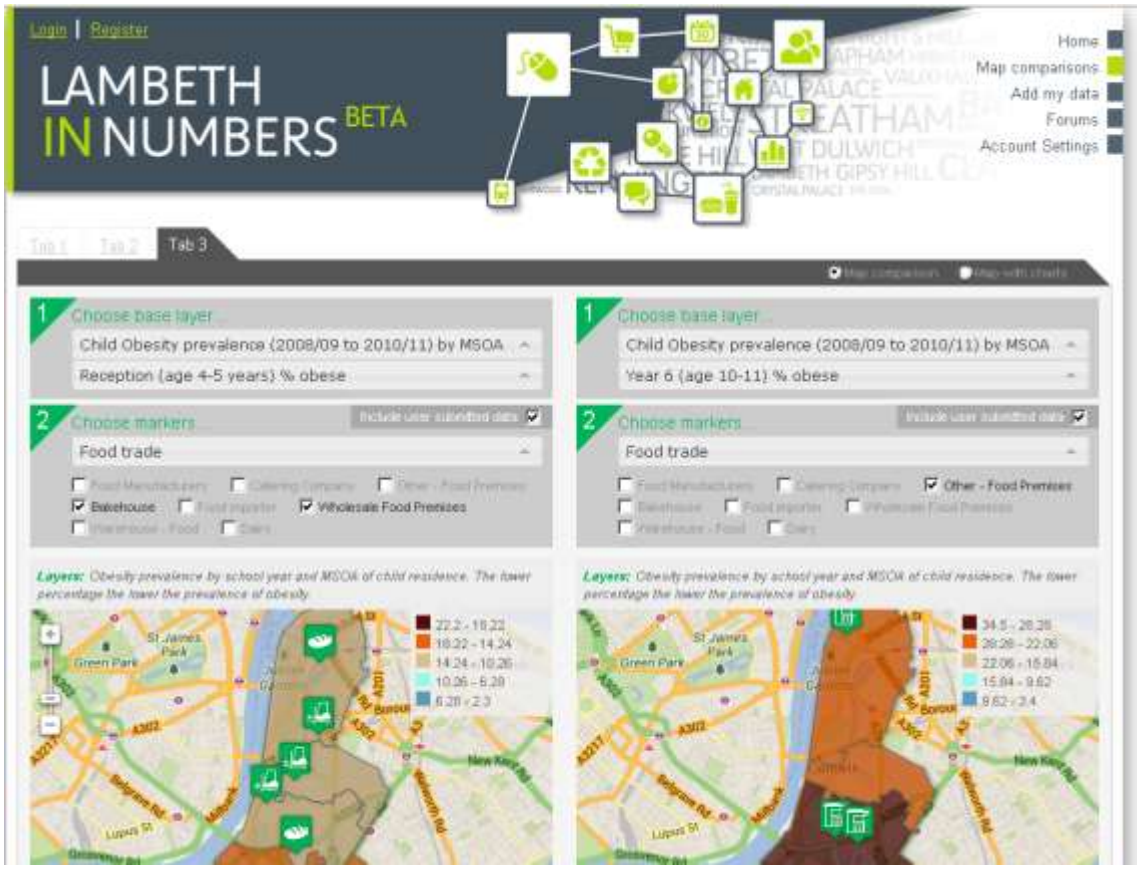


Figure 30 Partial screenshot of Lambeth in Numbers (http://www.lambeth-in-numbers.co.uk/)

Users can select which aspect they want to examine and how that affects different age groups. By making two copies of the same app available side by side, it's possible to make comparisons quickly and easily. Different local facilities can be

[196] http://www.w3.org/TR/sparql11-federated-query/

shown or hidden which may correlate with the statistics chosen in some way (for example: childhood obesity and the number of takeaways).

Interview summary

We spoke to Steve Peters as a consumer of Ordnance Survey's linked data - which Open Data Communities certainly is. However, it's also a provider of linked data. The Department of Communities and Local Government has a public task of making data available about individual localities and so in that regard, accessing data from multiple sources is important.

As well as those built by Steve Peters, DCLG has engaged commercial Semantic Web company Swirrl to create applications. These include the Local Communities Indicators dashboard (Figure 31). This dashboard uses linked data even though DCLG publishes the underlying data as Excel spreadsheets (it doesn't even bother saving them as CSV). The motivation here therefore is not to make linked open data available to others (although it does that too), but to reduce development time in making important data accessible to end users, whether they be citizens or policy makers.

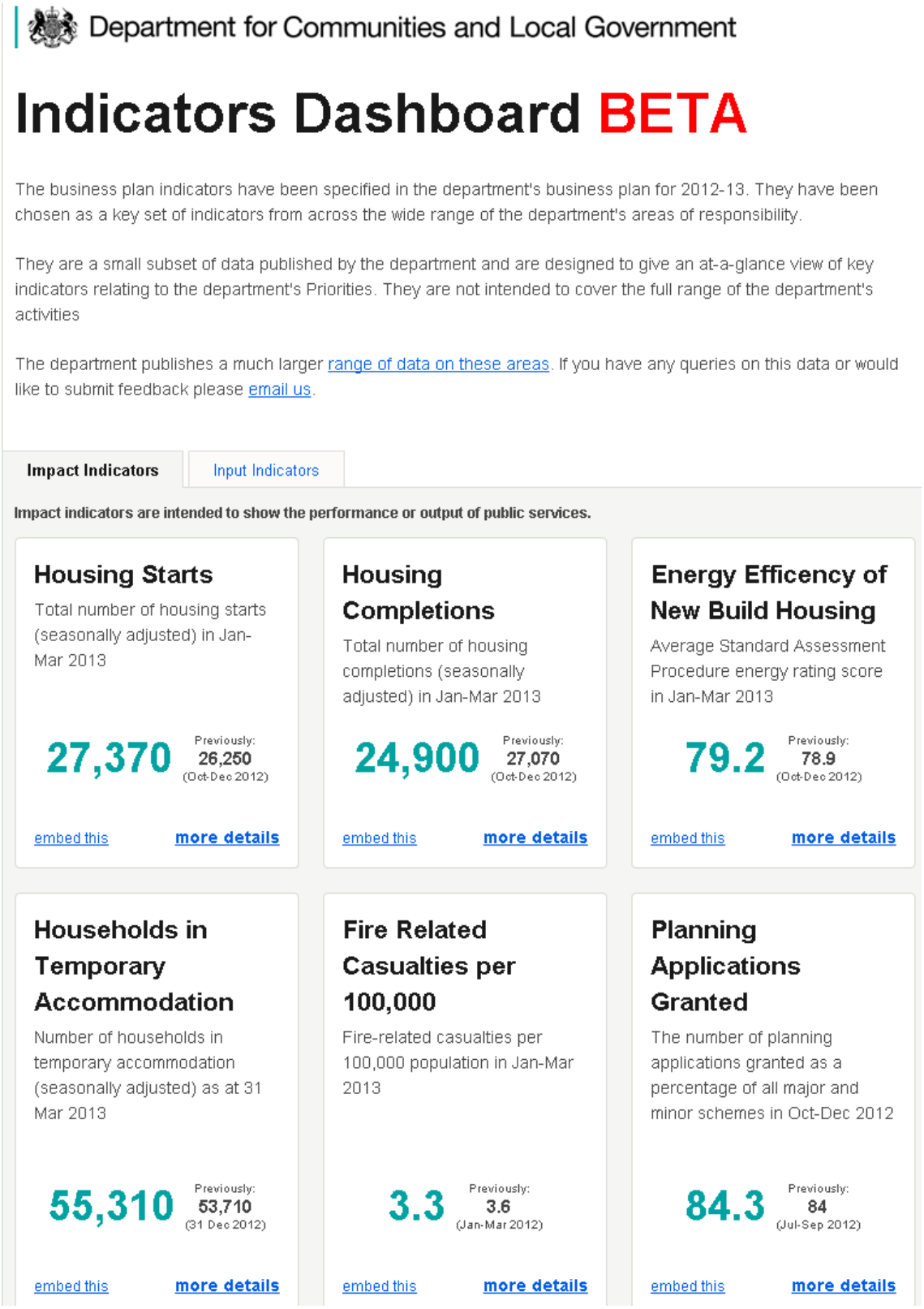


Figure 31 Partial screenshot of DCLG's Indicators Dashboard[197]

[197] http://dclgapps.communities.gov.uk/indicators/

II.14 UN – Food and Agricultural Organisation of the United Nations

II.14.1 Desk research

Source: http://agris.fao.org/openagris-linked-data and additional information from FAO

FAO, the Food and Agriculture Organization of the United Nations, is an agency that leads international efforts to defeat hunger. FAO is also a source of knowledge and information, and helps developing countries to improve agriculture, forestry and fisheries practices, trying to achieve food security for everyone.

AGRIS

In this context, in 1974 FAO set up an initiative called AGRIS to make information on agriculture research globally available. AGRIS is a collection of more than 4 to 5 million bibliographic records and it is one of the most important world-wide information systems in the agricultural domain. AGRIS serves a million pages a month, with more than two hundred fifty thousand users accessing the system every month: thus it has a strong and very important audience.

The Issue

AGRIS bibliographic records are created by cataloguers. Only 3% of the entire collection has a full text link, but using the structured information from the bibliographical record the entire text can be easily found through using the Google APS. A bibliographical record contains structured information (title, author, conference, journal, keywords) that virtually links to many other resources on the web. To allow users to access the full text of a publication and other related information in the agricultural domain, we developed OpenAGRIS, following Linked Open Data principles.

Linked Open Data

Linked Data is the way to publish structured data and to interlink with other existing datasets, in a machine readable way: to overcome the lack of information of AGRIS records, it was necessary to convert the whole repository in RDF and become part of the Linked Open Data cloud. Thus, more than 100 million triples were generated and used by OpenAGRIS, a web application that aggregates information from different Web sources to expand the AGRIS knowledge providing much data as possible about a topic or a bibliographical resource.

AGROVOC

AGROVOC is a 30 years old multilingual vocabulary containing more than forty thousand concepts in 22 languages and published as Linked Open Data. AGRIS records have been indexed with AGROVOC keywords.

Source: http://aims.fao.org/standards/agrovoc/linked-open-data

AGROVOC is now available as a linked data set published, aligned (linked) with several vocabularies:

- FAO Glossary of Biotechnology for Food and Agriculture, http://www.fao.org/biotech/biotech-glossary/en/

- Eurovoc, the EU's multilingual thesaurus maintained by the Publications Office of the European Union, http://eurovoc.europa.eu/
- GEMET, the GEneral Multilingual Environmental Thesaurus of the European Environment Information and Observation Network of the European Environment Agency, http://www.eionet.europa.eu/gemet
- Library of Congress Subject Headings (LCSH), http://www.loc.gov/aba/cataloging/subject/
- The National Agricultural Library's (NAL) Agricultural Thesaurus, http://agclass.nal.usda.gov/
- RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) maintained by the Bibliothèque nationale de France, http://rameau.bnf.fr/index.htm
- STW Thesaurus for Economics, a thesaurus providing vocabulary on any economic subject maintained by ZBW - Leibniz Information Centre for Economics, http://zbw.eu/stw/
- TheSoz, the thesaurus for the Social Sciences, maintained by GESIS – Leibniz-Institute for the Social Sciences, http://www.gesis.org/en/services/research/thesauri-und-klassifikationen/social-science-thesaurus/
- FAO geopolitical ontology, http://www.fao.org/countryprofiles/geoinfo/en/
- Dewey Decimal Classification (DDC), http://dewey.info/
- DBpedia, http://dbpedia.org
- SWD (Schlagwortnormdatei) of the Deutsche Nationalbibliothek, now part of the Gemeinsame Normdatei (GND), http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html
- GeoNames, http://www.geonames.org/

The Linked Data version of AGROVOC is in RDF/SKOS-XL, and is stored using Allegrograph as a triple store. Data is accessible to machines through a SPARQL endpoint, and to humans by means of a HTML pages generated with Pubby. Data is hosted by MIMOS Berhad (Malaysia), an FAO partner.

OpenAGRIS

OpenAGRIS is a Web application that aggregates information from different Web sources to expand the AGRIS knowledge providing much data as possible about a topic or a bibliographical resource. Using Agrovoc as backbone, OpenAGRIS can interlink with a lot of existing datasets (currently DBpedia, World Bank, Geopolitical Ontology, FAO fisheries dataset, AGRIS serials dataset, and so over), showing as much information as possible about a specific topic, as statistics about fish species or geographical distribution of plants. In this way, OpenAGRIS pages are landing pages that aggregate information the Web knows about a specific publication, topic, research area (in the agricultural sector).

OpenAGRIS is based on four internal FAO RDF datasets:

- The AGRIS records dataset, the direct translation of AGRIS XML records to RDF. Considering that AGRIS contains more than 4 million XML records, this new dataset consists of 100 million triples.
- The Agrovoc RDF dataset: AGROVOC is the world's most comprehensive multilingual agricultural vocabulary that contains close to 40,000 concepts in 22 languages covering subject fields in agriculture, forestry and fisheries together with cross-cutting themes such as land use, rural livelihoods and food security.
- The AGRIS journals dataset: since 79.54% of AGRIS records are journal's articles, FAO created a dataset of more than 22,000 agricultural journals with complete information about each journal (ISSN, start date, frequency, publisher…).
- The AGRIS centres dataset, which contains information about data providers, thus the AGRIS source of information.

The external datasets which OpenAGRIS is currently linking to are:

- DBpedia
- World Bank
- Google (Google Custom Search API)
- nature.com
- FAO Country Profiles
- FAO fisheries dataset
- GBIF (Global Biodiversity Information Facility)
- IFPRI
- Europeana

II.14.2 Collected metrics

Usage

Usage of the linked data collection is monitored through evaluation of Web statistics for individual users. Downloads of data dumps are registered.

Revenue and other sources of income

All activities are funded from public sources. Data is made available for free. There are no other sources of revenue.

Cost

Development cost is difficult to estimate. The team has spent project money in the range of a million Euros between 2002 and today to develop semantic web competence, methodology and technology. The direct cost of the now functioning LOD infrastructure has been not more than 100,000 Euro. Maintenance and promotion cost are part of on-going funding.

Benefits

OpenAGRIS contains over 5 million records and over 134 million triples.

AGROVOC contains over 32,000 concepts organized in a hierarchy, each concept may have labels in up to 22 languages: Arabic, Chinese, Czech, English, French, German, Hindi, Hungarian, Italian, Japanese, Korean, Lao, Persian, Polish, Portuguese, Russian, Slovak, Spanish, Thai, and Turkish. Four more language versions are under development (Malaysian, Moldavian, Telugu, and Ukrainian).

Benefits are mainly in efficiency gains and in improvement of services rather than in cost reductions.

II.14.3 Interview: Johannes Keizer

Interview date	7 August 2013
Interviewee	Johannes Keizer (Information Systems Officer, Team Leader Knowledge Standards and Services
Interviewer	Makx Dekkers

Q: How does the supply of Linked Data relate to the public task of your organisation?

A: FAO has a mandate to make all of its data available and provide access to data elsewhere for a worldwide community. The provision of Linked Data is one of the channels that are being offered. Contributed data is not always available as Linked Data so data may be imported through specific APIs, e.g. from the World Bank.

Q: Does your organisation have a "business case" document that motivates your investment in supplying Linked Data? Can you share it with us?

A: There is no formal "business case" document. We issued in 2008 a concept note "AGRIS 2010", which outlined the program of using bibliographical records to link to data on the web. Linked Data activities are part of the normal improvements of the technical tools that are being used. Linked Data is seen as the state-of-the-art for the kinds of information provision activities of FAO.

Q: What are your future plans? Do you plan to expand or abolish the supply of Linked Data?

A: The agricultural community pushes in the direction of open and linked data. The team at FAO is ahead of the curve in comparison to the community and takes care that the approach is sustainable and scalable.

Q: Did the supply of Linked Data give rise to new opportunities for flexible data integration within your organisation and with partner organisations? Increase in data quality? New services? Cost reductions?

A: The linking between AGROVOC and GEMET allows people to see connections that were not visible before. Internally, the linked data is used for disambiguation and other types of quality enhancements. The use of linked data leads to better services and efficiency gains, not to cost reductions.

Q: What are the main enablers / inhibitors for Linked Data to deliver value for its reusers?

A: The growth of the information space available to users is driven by data owners who make datasets available as Linked Open Data. Inhibitors are institutional

policies towards opening up data, and the institutional capacities to move upwards in the 5-star model; many data owners are still in the process to get to 1-star data level.

Q: Do you supply service level statements for Linked Data?

A: Not specifically for Linked Data. FAO aims for 24/7 availability, and guarantees accuracy and availability of its own data. No guarantees are given for external resources.

Q: Does your organisation have a URI policy? Can we have a copy?

A: Main principle of the URI policy is that URIs should be stable. Examples of URIs are:

- http://agris.fao.org/aos/records/IR2010000398 for an AGRIS record
- http://agris.fao.org/aos/data/IR2010000398 for metadata about the record
- http://aims.fao.org/aos/agrovoc/c_29966 for AGROVOC terms

Q: Can you describe the Linked Data infrastructure of your organisation?

A: The main AGRIS database is based on XML data. OpenAGRIS is a parallel RDF triple store that is made available through a SPARQL endpoint.

The Linked Data version of AGROVOC is in RDF/SKOS-XL, and is stored in Allegrograph triple store. Data is accessible to machines through a SPARQL endpoint, and to humans by means of a HTML pages generated with Pubby. Data is hosted by MIMOS Berhad (Malaysia), a FAO partner. AGROVOC is implemented as two separate triple stores: one that is used for updating and one that is available for external access.

Q: Which skills and competencies did you need in order to supply Linked Data? Do you have them in-house or contracted?

A: The necessary competencies are acquired by the team through conferences and workshops. FAO contracts with external experts who then transfer knowledge to the team. Maintenance of the main maintenance tool VocBench[198] is outsourced. There is a community that works on further development of the tool.

Q: Which organisations are key partners in the supply of Linked Data?

A: On the data side, there is a group of Agricultural Research Institutes that provide data for the linked data collections.

On the technical side, FOA works with MIMOS in Malaysia who host the linked data files and with several European projects.

Q: Which activities do you carry out to supply Linked Data?

A: Using the mappings between vocabularies, the linked data is produced automatically as much as possible. Checking of vocabulary mappings is done by experts.

[198] VocBench. http://aims.fao.org/tools/vocbench-2

Q: Which investments has your organisation made to enable it to supply Linked Data?

A: Most of the investment has been funded through European projects. The total amount is in the order of a million Euros over the last ten years, but this refers not in a strict sense to the now existing infrastructure, which did not cost more than 100,000 Euros.

Q: How many FTEs in your organisation are involved in provisioning the supply of Linked Data?

A: The team consist of 15 people but they also have other responsibilities.

Q: Which costs have you incurred to publish Linked Data, maintain, and promote it? What are the trends?

A: Unknown.

Q: Who are the main users of your Linked Data services? Is Linked Data only used by external customers or also for back-office exchange (e.g. with other agencies)?

A: Publications Office of the European Union, Luxembourg; European Environment Agency (EEA), Copenhagen, Denmark; Deutsche Zentralbibliothek für Wirtschaftswissenschaften - Leibniz Informationszentrum Wirtschaft (ZBW), Kiel, Germany; GESIS – Leibniz-Institute for the Social Sciences, Mannheim, Germany; German National Library (Deutsche Nationalbibliothek), Leipzig, Germany; MIMOS Berhad, Malaysia; Indian Institute of Technology (IIT), Kanpur, India; University of Rome at Tor Vergata; School of Information and Library Science, University of North Carolina, USA; Department of Computer Science, University of Waikato, New Zealand; Chinese Academy of Agricultural Sciences (CAAS), Beijing, China; Empresa Brasileira de Pesquisa Agropecuária (Embrapa), Brazil; Hasselt University Library, Belgium; Knowledge Synergy Inc., Japan.

Internal usage from FAO Technical Departments.

Q: How often is Linked Data used? What are the trends?

A: FAO has observed a 50% increase in the use of data because of the enrichments that were possible using Linked Data.

Q: How is the provisioning of Linked Data funded?

A: Technical development is funded through projects. Data maintenance is part of internal funding.

Q: Which pricing mechanism or other source of income exists?

A: Data is provided free of charge.

Q: Under which licence is Linked Data made available for reuse? Can we have a copy?

A: AGRIS content is made available and licensed under a Creative Commons Attribution 3.0 Unported License.

Copyright for the AGROVOC thesaurus content in English, French, Russian and Spanish is licensed under a Creative Commons Attribution-NonCommercial-

ShareAlike 3.0 Unported Licence. For any other language, the copyright rests with the institution responsible for its production.

Download of AGROVOC for off-line use can be requested through a web form.

Q: Which channels are predominantly used to consult Linked Data: Web API? Web site? (mobile) App? Data market?

A: Access is provided through Web services and SPARQL endpoint.

Q: Do you have a branding strategy for Linked Data?

A: Yes, this is done through the Linked Data section[199] of the AIMS website.

Q: Do you invest in advertisement for Linked Data?

A: Yes, through talks at workshops and conferences.

Q: Do you provide / make use of any mechanisms on user feedback or evaluation?

A: A feedback page is available for comments and suggestions on OpenAGRIS at http://agris.fao.org/openagris-comments-and-suggestions.

II.14.4 FAO as a reuser

FAO reuses data from many providers. One of those sources is Google. The issue that occurs is that there is a limit on the number of access to Google Open APIs. FAO usually runs over the limit by early afternoon which means that the Google data cannot be presented as additional information with FAO data.

[199] FAO. AIMS – Agricultural Information Management Standards. Linked Data. http://aims.fao.org/advice/linked-data