20 eGovernment

Nigel Shadbolt[1] · *Kieron O'Hara*[1] · *Manuel Salvadores*[1] · *Harith Alani*[2]
[1]University of Southampton, Highfield, Southampton, UK
[2]The Open University, Milton Keynes, UK

John Domingue, Dieter Fensel & James A. Hendler (eds.), *Handbook of Semantic Web Technologies*,
DOI 10.1007/978-3-540-92913-0_20,

Abstract: The use of the Semantic Web (SW) in e-government is reviewed. The challenges for the introduction of SW technology in e-government are surveyed from the point of view both of the SW as a new technology that has yet to reach its full potential, and of e-government as a complex digital application with many constraints, competing interests and drivers, and a large and heterogeneous user base of citizens. The spread of SW technology through e-government is reviewed, looking at a number of international initiatives, and it is argued that pragmatic considerations stemming from the institutional context are as important as technical innovation. To illustrate these points, the chapter looks in detail at recent efforts by the UK government to represent and release public-sector information in order to support integration of heterogeneous information sources by both the government and the citizen. Two projects are focused on. AKTive PSI was a proof of concept, in which information was re-represented in RDF and made available against specially created ontologies, adding significant value to previously existing databases. Steps in the management of the project are described, to demonstrate how problems of perception can be overcome with relatively little overhead. Secondly, the data.gov.uk project is discussed, showing the technical means by which it has exploited the growth of the Web of Linked Data to facilitate re-representation and integration of information from diverse and heterogeneous sources. Drawing on experience in data.gov.uk the policy and organizational challenges of deploying SW capabilities at national scales are discussed as well as the prospects for the future.

This chapter will consider the specific issues pertaining to the application of the Semantic Web (SW) to e-government, and look at some of the ways that the Semantic Web community has tried to address them. The first section will examine some of the challenges and opportunities for Semantic Web technologies within e-government, and review progress made. The next section will describe a detailed example of the UK's data.gov.uk program to represent government data on the Linked Data Web. It will also review pilot work that preceded the data.gov.uk work and which was important in providing insights as to the pragmatic deployment of SW technologies in the public sector. A set of resources will be given for further study, before a discussion of likely future directions of research in this area.

20.1 Scientific and Technical Overview

The Semantic Web (SW) is a response to the increasing demand for information in a range of human activities, as storage and processing have become cheaper, while access to digital information resources grows ever-wider. As the benefits are being reaped from the World Wide Web of linked documents, the costs of manual processing are also making themselves felt. One particular challenge that the SW is intended to address is that of the heterogeneity of information sources. Information is created by many processes, is represented in different media in formats of varying levels of formality, and is also of varying quality and completeness. Furthermore, the constructs used in representation – the concepts, predicates, relations, etc. – may have various and different interpretations assigned to them across different resources. However, search, retrieval, and inquiry over the Web, to maximize the utility of the Web as an information resource, should be performable not only within individual resources, but across resources, however heterogeneous in form and quality. From the openness of the Web comes its great value, and hence it is impossible to insist on particular formats, tools, or vocabularies. Information resources must therefore be *interoperable*, allowing sharing, amalgamation, and exchange of information between applications. Search should be accurate across resources. As has been described in the first volume of this handbook, the SW solution is to provide semantics for information, using ontologies, annotations, dedicated representation formalisms, and other aids to interoperability.

One important application area of the SW is that of government, and specifically e-government. In this section, the particular challenges facing the use of the SW in e-government, and significant work in this area, are outlined.

20.1.1 Introduction: The e-Government Opportunity for the Semantic Web

Government and administration generally is thirsty for information; it has been said that the unprecedentedly large information demands of the bureaucracies required to oversee early gargantuan building projects (such as the Egyptian pyramids) led to many important developments in writing and urban living. Although governments are powerful enough to insist to some extent on standardization of information provided to them, and to enforce completeness as far as possible, the complexity of modern administrations means that the centralizing forces are outweighed by the need to delegate the work of government into separate departments, which often become information silos. Intra-departmental efficiency is incentivized by the central government, often at the cost of inter-departmental efficiency.

As the size of government has grown, it has become increasingly reliant upon accurate and timely information about its legislative and policy contexts. Whether that information is gathered by governments, or provided by citizens and businesses, the quality of management of that information is vital. This has led to the promotion of e-government to manage information and deliver services using information technology (IT) where possible. Using IT should create a number of benefits for government, including the standardization of processes, efficiency of information transfer, and storage and effective search, not to mention a decrease in the costs of information management. There should also be visible benefits for the citizen, including the simplification of the interface with government, an increase in the transparency and therefore accountability of government, the ability to manage one's own case, and the lower taxes that should result from the reduction of the government's costs.

Government departments have a requirement to share and exchange information meaningfully. So-called joined-up government is a notoriously hard problem, and the nature of the issue suggests that SW technologies have an important role to play, allowing interoperability and transparency, integrating and reasoning over heterogeneous information sources, and supporting connections between different layers of government as well as across departments. Furthermore, the use of semantic models of interactions should allow for the evolution of systems, reuse of software across contexts and departments, and the creation of customer-focused, multi-viewpoint access to services and information, including the personalization of services.

Heterogeneity is a serious issue for governments. They gather information from so many sources, including their many subdepartments and agencies, and also store so much legacy information (most states have records that predate widespread use of digital technologies) that one must expect the information they use to be extremely heterogeneous. But one must also consider that, unlike most businesses or online services, the users of e-government are particularly varied, as they will include, potentially, every single occupant of a country, including people with minimal computing skills, people who are unable to speak the main languages of a nation, people who, perhaps through disability, find it extremely hard to communicate, and even people who are not citizens of the country in question (as well as citizens of a country not resident there).

Hence the SW, a technology to address heterogeneity, has a great deal of potential for supporting e-government. However, as the SW is yet to reach full fruition, there is much work, research, and implementation to do. Furthermore, as shall be seen in the next section, the e-government domain raises a number of specific problems for the implementation of SW technology.

20.1.2 The Challenges of e-Government

Recent work on e-government has shown that interoperability and reengineering problems can interfere seriously with the effectiveness of government services online. In particular, studies have highlighted the need for standards to support interoperability, security, and privacy requirements that stem from the amalgamation of databases and services, and process reengineering to optimize the benefits of shifting governmental services online. Very few countries are making significant progress to transform their administrative processes – for examples of discussions and policy interventions, see [1–3].

Moreover, despite the opportunities that are available, there is reluctance to follow through, as Peristeras et al. describe:

- Governmental agencies still publish public-sector information (PSI) using a wide variety of nonstandardized and proprietary formats. The sheer volume and wealth of PSI make the potential benefits of reusing, combining, and processing this information quite apparent. However, agencies typically first express reluctance to make their data available, for various cultural, political, and institutional reasons. So, they keep their legacy systems, and the information stored there, fenced and isolated. Even if they decide to move on and free their data, the different data formats, the lack of commonly agreed-upon metadata, and the absence of standardized vocabularies and definitions result in a huge bulk of practically useless data [4].

The representation of this mountain of data is clearly a vital first step toward semantically enabled government services, and would clearly be an important gain in its own right. In the example system discussed later in this chapter, data.gov.uk, real evidence of progress toward this goal is shown.

Implementation of e-government services of all kinds is usually seen as a stage by stage process of increasing political and technological sophistication [5]. Layne and Lee, in common with other commentators, set out a four-stage process, of which the first is cataloguing, creating an online presence, putting government information on the Web, creating downloadable forms, etc., giving a one-way communication (broadcasting) facility. In the second stage, the internal government information systems are connected with the online interfaces, and citizens can transact with government, make requests, provide information, fill in forms, etc., making communication two-way. The third stage is one of vertical integration, where local systems are linked to higher-level systems with related functionality. The result for the citizen is a "one-stop shop" that appears seamless. Resources of greater value than information can be exchanged. The final stage is horizontal integration, where systems are integrated across functions, services become ubiquitous, and the individual departments of government become increasingly irrelevant from the point of view of the citizen, who just picks up the services he needs from a single portal. In particular, it is the third and fourth stages that are genuinely transformative of government information infrastructure.

However, it is fair to say (a) that very few e-government systems have been genuinely transformative [1], and (b) that the application of SW technology in this space has been more in the realm of prototypes or proofs of concept than fully fledged delivered systems or procedures. A different approach, embraced by the US's data.gov (http://www.data.gov/) and the UK's data.gov.uk (http://data.gov.uk) initiatives, is to make publicly available much of the nonpersonal data upon which public services and government depend. One then encourages a developer community and third party organizations to use that data and build their own applications using it, thereby guaranteeing user-relevance and often delivering capabilities not envisaged by the government and its departments. This unanticipated reuse was what drove the success of the Web of documents and it is argued that this is what will happen in a Web of Open Government Data [6, 7].

In this section, some of the challenges to SW technology and research in this domain will be surveyed. However, some problems follow from the nature of the administrative and political process, and clearly cannot be addressed from within the SW community.

20.1.2.1 Specifically Political Problems

Because businesses routinely have to perform reengineering of legacy systems, and because they face similar difficulties, it is tempting to treat government as a large business in the analysis of the problem. However, government has many drivers and difficulties of context that businesses do not face: in particular, whereas businesses have the (relatively) straightforward goal of creating value for shareholders within the law, governments need to meet a wide range of targets and have a range of public tasks and duties. Furthermore, different governments need (or want) to meet different targets.

Governments are extremely large operations compared even to multinational firms, and they are rarely as "lean" and efficient as private enterprises. Government bureaucracies are large employers, and so the politics of employment are brought into play. Changes in working conditions (including redundancies) are sometimes difficult for governments to negotiate. They may suffer at the ballot box if they restructure in too radical a way. Even if they do make workers redundant, they still have to support their former employees, via social security or unemployment benefits.

Hence government has special problems with information, reengineering, and change management generally. Although the SW is clearly an important tool in the future for e-government, these problems will of course apply to any move to adopt the SW, imposing costs and requiring new capabilities. A business can provide a business case for reengineering its information systems, and show how value is created and profits maximized. Although value for taxpayers' money is an important factor in government decision-making, it is not the only one. Indeed, in democracies, where a government will typically face its voters within 4 or 5 years, high perceived short-terms costs will often outweigh long-term benefits, which may accrue to a government of a different party.

These problems of reengineering information, which often requires wholesale reengineering of government structures and ways of doing business, and the preservation of trust and privacy, mean that the e-government area is an especially complex and demanding application for the SW. It is true that governments do have advantages in that they can enforce rules and standards, and their demands for information are usually met, but the most prominent forces acting upon them are political ones that tend to promote inertia rather than radicalism. The generalized benefits of efficient information flow accrue to everyone to a small degree (including to outsiders who do not vote for the government), while the smaller number of losers, who either have to consume more resources to support reengineering, or who may lose their information-processing jobs altogether, suffer relatively greater losses and are incentivized to become a vocal minority opposed to change.

The problems that attach to any kind of technological or administrative change in a democratic polity cannot of course be solved by SW research, although in e-government they are always part of the background. The aim of this chapter is to look specifically at how SW technology and SW research can make a difference. Hence the next subsection will consider the specific challenges where technological research can be expected to make a difference.

20.1.2.2 Challenges that can be Addressed by SW Research

It is clear that, despite the opportunities for knowledge management and integration that the SW affords, there are serious challenges to its implementation in the e-government domain, and that therefore need to be ameliorated or overcome. However, progress is being made on all fronts, and so despite the length of the list of the challenges, there is still room for optimism. The major challenges can be enumerated as follows.

1. *Representing information.* Governments need to capture information about themselves, and describe themselves to their citizens and interested stakeholders (such as companies competing for government business), yet doing this effectively remains an unmet challenge [4]. SW technologies, particularly ontologies, can support the creation of portals and knowledge maps that allow stakeholders to discover how e-government works in a particular context, and to provide an instrument of analysis for improvement of services [8]. On the other hand, models of citizens' requirements may be harder, although Ilgar et al. [9] have argued that the use of emergent semantics, based on the evolution of "folksonomies" as a result of thousands or even millions of interactions between citizen and government, might "circumvent the problem of ontological drift by dynamically tracking the changing ways in which people conceptualize their domain."
2. *Integration of information.* The particular advantage for the SW occurs when interoperability issues have a semantic dimension [10] – in other words, where the lack of interpretation of data causes administrative obstacles. The information resources in e-government contexts are a wide set, including data, documents (including multimedia), files for download, transactions, links, services, and user-provided or user-related items (such as credit card details). These will need semantic markup if SW technology is to allow machine understanding of e-government interactions. Also, governments are complex entities, and each entity has its own information and descriptive terminologies. Integration will need to operate with these different terminologies to minimize cost and disruption, and top-down approaches for enforcing integration are not suitable for governments [11].
3. *Publishing information.* Governments need to release information to their citizens under freedom of information initiatives that are spreading across the democratic world, and support and facilitate democratic debate. Traditionally, information law has focused either on protecting information of commercial value (copyrights, trade secrets, patents), or on protecting confidentiality of certain relationships within which private information needs to be readily disclosed (e.g., doctor–patient, lawyer–client). However, many writers and activists, and some politicians, have argued that the asymmetry between government and governed in access to information exaggerates asymmetries of power, and if governments are to be democratically held to account, more information needs to be passed back to citizens. Furthermore, in an argument basically borrowed from John Stuart Mill, others have argued that the quality of policy-making and administration would be improved if more people were involved in them, for which they need information.
4. *Search and discovery of information.* Data in most government organizations are stored in many databases in different departments and locations, yet to realize the potential gains from this data it is necessary to bring it together from across government and elsewhere on the Web to achieve synergy and maximize value. However, the sort of federated search and retrieval that is required is known to be a hard problem. Furthermore, without an idea of what information is available, it is difficult for someone in a particular government department to realize what inferences he or she

could potentially make. The development of policy depends on government officials understanding what information was available to the policy-making process.

5. *Web Services.* Government services need to be configured and composed for delivery to citizens with heterogeneous requirements. For the provision of services, SW services promise greater flexibility and controllability than standard Web Services. Although Web Services can be composed quickly and effectively, their syntactic definitions cannot describe a service's functionality precisely, while the description also needs a human to interpret it in order to determine those contexts of application where it will work, together with the required inputs and outputs. There is also a cost in remodeling every time a new service is deployed. Because of the semantics attached to SW services, context and capabilities can be modeled effectively, and service invocation, discovery, composition, and mediation can be automated, while methods for creating SW services exist that allow new methods to be deployed, or old methods repurposed, without rewriting the entire business process [12].
6. *Privacy and access control.* Although governments are now expected to release information, they have at the same time a requirement for preserving the privacy of their citizens and attending to issues of national security, which is naturally a recurring tension. Data are extremely useful both in the commercial and academic worlds as well as government, yet governments must be circumspect about what they release. Overlapping the need for privacy is the requirement for data protection, which is intended to strike a balance between privacy and fair use, focusing on such matters as good information management, security, the quality of democracy and society (requiring the free flow of some information), freedom of information, freedom of expression, regulation of data stores, and the ability of data subjects to inspect and if necessary amend faulty data [13].
7. *Reengineering and change management.* Governments need to manage the effort of reengineering (both of processes and of legacy data), especially in the context of e-government where it is inefficient simply to bolt e-government technologies, such as one-stop portals, onto preexisting information silos. A seamless front end usually requires well-designed back-end processes. The use of SW technology to develop e-government services and to manage information requires the creation of a number of knowledge resources, including ontologies and process models. Such models are potentially very beneficial for the reengineering effort, as they themselves can be used to visualize the system for its designers, and provide a common understanding of processes for the governmental system as a whole. However, their development can undo, or cut across, many years of information-handling practice, requiring the reorganization of information management. There needs to be a willingness to look at reengineering the fundamental workflow of processes. Furthermore, as boundaries between information sources are broken down, the necessity of particular departments taking specific responsibility for gathering certain types of information becomes less. The logic of merging and amalgamating data is to lower the barriers between people and units traditionally kept institutionally separate.

8. *Perceptions of the costs of implementing SW technology.* Cost-benefit analyses are clearly important in “selling” the SW to an organization, but the problem seems to be that the benefits are hard to quantify, whereas the costs – including conversion costs, maintenance costs, organizational restructuring costs, and transaction costs – seem overwhelming [14]. The original *Scientific American* article popularizing the SW in 2001 [15] put forward a futuristic vision of software agents acting on useful information to perform complex tasks for their users. The standards that are a precondition for such large-scale agent-mediated information processing are progressing, and some are now in widespread use, but the early vision will not emerge until SW standards are well-established and used on a global scale [16]. Likewise for e-government, the benefit of adopting SW approaches will only become evident once semantics inside as well as outside governments are more widespread. Perceptions abound about the high cost of developing SW technologies, such as building ontologies and converting data into RDF. As Peristeras et al. [4] argue, the natural inertia of standard hierarchical governmental structures can also help push perceptions of costs in an unfavorable direction, and make the benefits seem less tangible.

20.1.3 Meeting the Challenges

In this section, approaches to the eight challenges outlined above will be reviewed, looking at the potential for improvement and actual progress made in SW research projects.

20.1.3.1 Challenge 1: Knowledge Representation

The increasing digitization of government services has led to a plethora of government-sponsored schemes and architectures to represent governmental knowledge. The Federal Enterprise Architecture (FEA) has been developed by the US Office of Management and Budget to provide a common set of terms and practices for the procurement of IT services, enabling a uniform approach to describing and evaluating IT investment and collaboration. FEARMO (http://web-services.gov/fea-rmo.html) is an associated reference model ontology consisting of executable specifications in OWL-DL of five reference models focusing on performance, business, service, technical issues, and data. The specific aim of FEARMO is to serve as a catalyst for innovation in SW technologies to provide interoperability.

The FEA contains a data reference model that provides principles for the description, categorization, and sharing of data. In terms of description, the aim is to provide means for an agency to agree data structure and semantics, using logical or conceptual data models to provide metadata. In terms of categorization, taxonomies (which can be in the form of XML topic maps or OWL hierarchies) are used to give additional meaning to data by relating it to the context of its creation and use.

The UK government has also established a metadata standard, eGIF (http://www.govtalk.gov.uk/schemasstandards/egif.asp), with an associated metadata standard eGMS (http://www.govtalk.gov.uk/schemasstandards/metadata.asp), which defines terms for encoding schemes, thematic categories and relations using a thesaurus approach, and draws on common and well-used standards, in particular Dublin Core. The Integrated Public Sector Vocabulary (IPSV), a structured thesaurus of administrative activities that was set up for use within eGMS, provides most of the semantics.

As part of the data.gov.uk project the UK government has initiated the development of several core ontologies, including the SDMX/SCOVO ontology for statistical data and the organizational ontology to represent government structures. Also ontologies are available for particular sectors – for example, administrative geography, education, and transport. In a parallel effort the W3C eGov Interest Group has started the development of the DCAT (http://www.w3.org/egov/wiki/Data_Catalog_Vocabulary) ontology to represent data assets.

One foundational ontological structure that has been found valuable for e-government is the *life event*, a meaningful entity for citizens (e.g., a wedding, the purchase of a house) that links together administrative services, procedures, and requirements. A standard vocabulary of these, general enough to express the particularities of the varied legislation across polities, while specific enough to give a modeler enough expressive power to build meaningful models, has emerged [17]. Life events have been argued as having several advantages with regard to search, service automation, and usability for e-government portals [18]. So, for example, the Access-eGov project (http://www.accessegov.org/acegov/web/uk/) [19–21] adopted life events to model government processes from the point of view of users, information consumers, citizens, and businesses. This enabled requirement-driven development of semantic structures based on consumers' needs, and allowed users to browse the e-government site structured by life events.

Peristeras et al. [4] argue that one important challenge for e-government, and the related concept of e-participation of citizens in government processes, is to be able to represent Web 2.0 discussions in order to understand the shifts in public opinion and mood (it may be that work to track opinion and bias, as in the Living Knowledge project [22], would be helpful in such a challenge). The authors advocate the use of linked open data and other Web vocabularies, combined with other technologies such as natural language processing and argument representation, to represent and "easily 'sense' what a community wants." This suggestion, however, ignores some of the dangers of such techniques. For instance, the opinion "worm" that is generated in real time by changes in opinion during leadership debates has become not only an indicator of the progress of the debates, but one of the outcomes. The worm becomes a quantified proxy for debating success, and success in turning the worm upward becomes the aim of the debate for its participants. One could expect a similar result if real-time changes of opinion in the blogosphere or Twitter were tracked. To widen the point, representation of information, particularly about opinion, is not always a neutral move, but can also supply politicians with new incentives and new weapons.

20.1.3.2 Challenge 2: Integration

Integration, or interoperability, is an issue that raises itself in the e-government domain in a number of areas. Klischewski [23] gives the following types of interoperability, which together add up to a highly complex set of interrelated problems.

- Citizens' understanding of a situation needs to be mapped onto that underlying the Website or service they are accessing.
- Citizens are also likely to be accessing other Websites and services in relation to the issue at hand (either for advice, or because the issue requires both governmental and nongovernmental support – for instance, moving house needs contact with estate agency and banking services as well as engaging governmental change-of-address, regional tax, capital gains tax, and local government services).
- Support for the citizen may require integration of private and public resources.
- It may also require process management across government organizations and IT services.
- Citizens' requirements in one area may overlap with requirements in other areas. Someone moving house will need to interface with tax authorities, local government, and utilities. Furthermore, this event may cause other entities to seek services, so a company may need to seek a work permit, utilities may find they need to establish some services to the new address, while the local government may also be affected (e.g., in terms of care for children or elderly people) as a result.

Semantic interoperability – ensuring that the meaning of shared information is interpreted in the same way by sender and receiver – sits alongside other types of interoperability therefore, including organizational interoperability (business processes and collaborations understood in the same way across organizations) and technical interoperability. Of course these are linked – Klischewski [24] argues that within a single organization process integration requires more planning and a higher intensity of cooperation and financial investment, but has a higher potential payoff for success; however, when the integration required is across global, open partnerships, process integration is not usually successful. With a large number of more or less independent units (administrative units within a single government, as well as its network of stakeholders), the successful exchange of information across their heterogeneous IT systems and procedures is far more likely than cross-organization process management, which anyway will begin with information integration as a first stage. Devoted interoperability frameworks usually contain technical standards catalogues, which serve to provide basic guidance to departments, as well as help standardize work procured from outside IT suppliers [25]; and with the spread of SW concepts and services, the SW provides new technical options for achieving information integration [24].

It is also important to distinguish between front-office and back-office integration. In the absence of any kind of integration, the user has to access services in the order required to do a job, and so is in effect required to manage the process him- or herself. During back-office integration, services are integrated at the system level, so a user accesses a service that then invokes other services as it needs them. In front-office integration, the user accesses a personal assistant, which then coordinates services behind the scenes, thereby integrating services at the user level (e.g., generating a plan for the user's current life event) [26].

The key contribution the SW makes to integration and interoperability is to allow a common ground to be reached without forcing a single perspective on all the actors involved [23]. Without a shared basis, for example, in terms of the actual creation or reuse of ontologies, markup languages, and markup methods, integration will not be achieved because of differences in semantic assumptions made by practitioners in, say, law, policy-making, service delivery, ideology, administrative processes, and IT procurement and management, not to mention the fact that in many nations such services have to be marketed and explained in a number of different languages. Weinstein [27] argues that the requirement to unify, link, or otherwise align models and ontologies enables the identification of commonalities and diversity, thereby facilitating policy discussion in such a way as to foster creative negotiation.

The SW's contribution is based on its philosophy and the formalisms that make up the layered diagram. RDF provides machine-readable descriptions of data, while ontologies expressed in OWL give interpretations of the descriptions with a common vocabulary. Rules and ultimately other types of logic will enable systems to reason about the data. When such a cascade of formalisms is in place, it should enable operationalized methods for resolving differences in interpretation over digital data and services despite structural differences and variable underlying assumptions. If two systems use the same underlying ontology (which is quite possible in well-understood domains), then interoperability will be obtained via a straightforward mapping, but even if they do not and interoperability issues reemerge one level up, their different ontologies can be published, and then mapped or merged. Ontologies are also important methods for guiding citizens through difficult areas with well-established vocabulary and principles that have reached a stable consensus (e.g., law), preventing mistakes in query construction, or misinterpreting answers.

In Europe, the European Interoperability Framework (EIF) has been set up to support interoperability in data exchange, and as part of its principles the Semantic Interoperability Centre Europe (SEMIC.EU – http://www.semic.eu/semic/view/) has been created by the European Commission. This is a service for the seamless exchange of data, and focuses on the semantic aspects of interoperability, promoting the reuse of assets both syntactic, such as XML schemas, and semantic, such as ontologies, in the e-government domain. It has set up an open repository of what it called interoperability assets, and also maintains a quality assurance framework. The EU has also sponsored a number of projects in the e-government field.

For instance, the SEEMP project [28] is an EIF-compliant architecture to support interoperability across public and private employment service agencies. Although each service has its own local ontology for describing at the semantic level the services it exposes and the messages it exchanges, because they are all operating in a reasonably standardized domain, these ontologies are pretty similar. SEEMP has developed a single consistent ontology out of the local ones, which it hopes will become a reference standard for the employment services who should provide mappings between the local and global ontologies. In a similar approach, the BRITE project [29, 30], which aims to build interoperability between business registers across the EU to facilitate cross-border e-government services for businesses, links (and is generated from) divergent national ontologies with a high-level domain ontology (HLDO) that acts as the intermediary between the local domain ontologies.

SmartGov is another EU project designed to specify, develop, deploy, and evaluate a knowledge-based platform to generate online transaction services, which can be easily maintained and integrated with legacy systems. The approach has been to store knowledge units (which could be anything including help, best practice guidelines, examples, troubleshooting advice, etc.), which are often short unstructured pieces of text, in a knowledge base whose structure reflects that of a transactional service. Domain-specific transactions use a domain map to link the transaction structure (and hence the knowledge units) with the domain concepts; the map is based on an e-government service ontology [31].

Barnickel et al. [32] point out that a global ontology is impractical for at least some e-government scenarios because of the international dimension (systems from different states often have to achieve interoperability), and they argue that service composition in a semantic interoperability infrastructure is the way forward. The combination of domain-specific ontologies and upper ontologies for Web Services, such as OWL-S and WSMO, allows SW services to be wrapped around already existing Web Services (services are discussed in more detail below in ➋ 20.1.3.5). Semantic bridges [33] describe the relations between distinct concepts defined in different ontologies that nevertheless are intuitively close in meaning; Barnickel et al. [32] use a rule language to do this, ensuring that the transformations can be implemented with an inference engine. Creating such bridges requires cooperation and sharing between domain experts familiar with the (local, domain-specific) ontologies being linked. Semiautomatic tools support service composition, by reasoning over semantically described relationships (such as inheritance between concepts) and recommending suitable assignments between output and input parameters of different services.

As an alternative approach [34], bridges disjoint SW applications by using automatic ontology alignment followed by automatic translation of metadata from one application to the other, allowing services to communicate. Given two SW applications to bridge together, aligning their ontologies produces an alignment file that maps terms from one ontology into the other, which can then be used to translate the output and effect specifications of one application into the ontology of other, which then can be measured against the preconditions of the second application.

In the linked data world several services are alleviating the issue of decentralized data integration. Co-reference systems, such as sameAs.org, help to interconnect linked data resources that represent the same concept or object. A simple RESTFul API can be used to discover more information about the same thing in different linked databases. Other services like the EnAKTing PSI Backlinking Service (http://backlinks.psi.enakting.org/) help the integration of linked data by resolving the problem of foreign URIs [35]. Linked data approaches argue for lightweight ontologies and do not insist on large-scale overarching ontologies [16], which can be expensive to build and difficult to maintain.

20.1.3.3 Challenge 3: Publishing

Following the severe financial crisis that began in 2008, the administration of President Barack Obama, which took control in January 2009, put in train a package of economic stimulus of unprecedented size. One key aspect of the package was the need to build and retain voter trust given that the stimulus was close to $1 trillion, that lawmakers had a poor reputation after a series of scandals, and that banking executives were already perceived by voters and consumers as having manipulated previous systems in order to award themselves large bonuses at the cost of profits for the shareholders, and the gross inflation of systemic risk. Hence transparency about the conduct of the stimulus was seen as central. To this end, a Website, recovery.gov (http://www.recovery.gov/), was created to:

- . . .feature information on how the Act is working, tools to help you hold the government accountable, and up-to-date data on the expenditure of funds. The site will include information about Federal grant awards and contracts as well as formula grant allocations. Federal agencies will provide data on how they are using the money, and eventually, prime recipients of Federal funding will provide information on how they are using their Federal funds. Interactive graphics have been used to illustrate where the money is going, as well as estimates of how many jobs are being created, and where they are located. And there will be search capability to make it easier for you to track the funds.

The interest of this site is twofold. First of all, it is underpinned with SW technology, representing and linking data using RDF, supporting queries with SPARQL, and so on. It is also using cutting-edge ideas, such as Semantically-Interlinked Online Communities (SIOC), developed at the Digital Enterprise Research Institute at Galway, Ireland, which is intended to support the development of online communities and linking debates by providing an ontology for representing social Web information in conjunction with FOAF [36].

In the United Kingdom, the Office of Public Sector Information (OPSI – http://www.opsi.gov.uk/) led the early drive for the release of public-sector information (PSI), with the aim of "understanding the potential of freeing up access, and removing barriers to reuse, [which] lie at the heart of our push to raise awareness of the potential for transforming how the citizen and state interact" [37]. OPSI operates from within the National Archives, and is at the center of information policy in the UK, setting standards, delivering access, and encouraging the reuse of PSI. It has responsibility for the management of much of the UK government's intellectual property; it is also the regulator of the information-trading activities of public-sector information holders. As such, it has been in a central position to support the release of PSI using SW technology and formalisms to facilitate discovery and sharing. David Pullinger, Digital Policy Director of the Central Office of Information has supported the use of SW formalisms to link data in a national information infrastructure, allowing reusable information, represented with SW conventions, in decentralized form but including identity management systems to allow personalization, while an important aim of OPSI is to raise awareness of the SW through government [38]. An extended example of the use of SW technology involving OPSI, called AKTive PSI, is described below in the release of UK PSI (cf. also [39]).

Other participants in AKTive PSI, such as Ordnance Survey, have also made use of Semantic Web technology. For example, the OS GeoSemantics team has released an ontology and dataset of 11,000 instances for Administrative Geography, which describes the administrative divisions in the UK – a complex dataset ideally suited to semantic representation [40]. The success of AKTive PSI was one of the factors that led to the development of the UK transparency and open data initiative data.gov.uk, which will be discussed in detail as an example of the use of the linked data paradigm in government information.

In May 2009 the US government's federal Chief Information Officer Vivek Kundra launched the data.gov Website to increase public access to valuable machine-readable datasets generated by the federal government. The philosophy of opening data for citizens' use has been influential in this development. At the time of writing, in the 18 months since its launch it has released thousands of datasets containing billions of RDF triples. In June 2009 Tim Berners-Lee and one of the authors of this paper (Shadbolt) was asked to establish a single point of access – data.gov.uk – for UK public data. It too now hosts thousands of datasets and is promoting the use of linked data standards in the UK Government.

This was not the first time SW technologies had been used by the US or UK governments. RDFa was already in use, to a small extent, behind the scenes on the White House Website (http://www.whitehouse.gov/) – but the set of technologies in use now is comprehensive enough, as one blogger put it "to enable the citizen masher to do their wizardry." In 2004 experiments began in the UK to evaluate SW technology. Both data.gov and data.gov.uk are recent sites but their performance and the examples they set will be of great importance as early adopters of large-scale use of linked data in the public realm. Both are likely to be bellwethers of the success of SW formalisms and technology for handling public information. They are governed by similar philosophies of releasing open data in such a way as to allow it to be integrated with and linked to other information sources, using lower-level SW technologies appropriate for the Web of Linked Data. Both initiatives have benefited from top-level political support from Presidents and Prime Ministers.

20.1.3.4 Challenge 4: Search and Discovery

Improved discovery of information is of course an imperative for an e-government system. Comte and Leclère [41] argue that using semantic reasoning will be particularly helpful to address issues of lack of interoperability, poor document management, and the absence of intelligent mechanisms. Interoperability has already been discussed above. With respect to document management, e-government information systems are often based on database management systems, but the resources with which they have to deal are often unstructured, and may not even be digital. The failure to index information well enough can also cause it to be effectively lost, as irretrievable. The absence of intelligence and inference is important, say Comte and Leclère, because of the low level of familiarity and expertise of the client, the citizen. They give the example of someone with a request in the legal domain who needs to have not only an answer, but also a reassurance that the answer is complete and correct, that no other relevant information has been missed out, and that information held implicitly in the information source has been rendered explicit. In their system, they rejected the classical database closed world assumption, and built a portal in which each government information resource is described by metadata using vocabulary from a heavyweight ontology, which can be displayed in networks. A set of protocols and processes allows the importation of data specified in RDF and OWL, and the reasoning uses an AI formalism based on Sowa's conceptual graphs that has an expressivity equivalent to RDFS.

Sidoroff and Hyvönen [42] argue that the application of SW techniques to the problems of content discovery and aggregation in e-government portals is highly beneficial, allowing semantic search and dynamic linking between pages. This in turn allows the consolidation of heterogeneous content related to the same information need. They have used SW techniques on the Suomi.fi e-government portal based around the idea of life events [17]. This allows the development, for instance, of *compound pages*, which tie together several information sources into a list in a single resource, aggregating information from different sources in a clear way. Explicit logic rules in SWI-Prolog allow the generation of links dynamically, making it possible to link any information resources in the portal that satisfy a linking rule, for example, exploiting similar properties of the resources expressed as metadata, or providing navigation hints based on recommender systems. Search allows content to be classified and viewed along several orthogonal views simultaneously. Ontologies (defining views along dimensions such as life events, topics, location, target audience, etc.) are used to describe Suomi.fi's information items, which are then used to represent the information in taxonomies for the user interface. A top-level view shows the whole subject topic taxonomy, which the other views allow alternative ways of classifying the content; the approach insists that new kinds of view for search can be created easily and the content projected onto them as the user requires.

Peristeras et al. [43] describe a use case and a platform for discovery of e-government services based on an OWL representation of the Governance Enterprise Architecture (GEA), which enables the semantic matching of citizens' profiles with formal specifications of available and relevant public services. A similar ontology-driven approach has been described in [44], in the development of a semantically enhanced search system to retrieve statistics, a large and growing space of data. This paper puts forward an intelligent search engine based on modeling electronic catalogues in the EU Combined Nomenclature with OWL. The search technique combines standard keyword-based search of the actual database with a higher-level RDQL (RDF Query Language) query of the ontology. The SAKE project [45, 46] provides a framework and tool workbench for an agile e-government change management, knowledge sharing, and knowledge and service creation, detecting change in information sources and work contexts and proactively delivering resources to users.

The hierarchies can be very useful navigational aids for the user as well. Sacco [47] argues that the use of *dynamic taxonomies*, which can adapt dynamically to the user's focus, allowing items to be classified under an arbitrary number of concepts at any level of abstraction, solves many of the problems of search and discovery for citizens who are perhaps only tenuously aware of what they are looking for – the search does not depend on the starting point, and it produces all potentially relevant information (e.g., all services offered to senior citizens) before drilling down to find the particular requirement.

A different principle is in operation in the exploitation of linked data. As part of the data.gov and data.gov.uk, an increasing number of linked data mashups have been noticed. As already noted the overarching principle here is that government makes data available and then a large developer community outside government builds the applications. Examples of these will be given in the example application section below.

20.1.3.5 Challenge 5: Web Services

The Web Services field is seen by many as the silver bullet for the SW within e-government, especially as the provision of services is a key function of government. Moving services to the Web would provide high availability and facilitate reusability [48].

The composition of government services (both on- and off-line) has generally been ad hoc and time-consuming, and this issue has also afflicted the field of Web Services as well. Traditional languages for describing operational features of Web Services to enable service composition, such as the Web Services Description Language (WSDL), the Simple Object Access Protocol (SOAP), or Universal Description, Discovery and Integration (UDDI) have little or no support for the semantic description of services that would allow automatic discovery, selection, and composition, although they could be supplemented by ontologies [48–52]. Chun et al. [50] argue that automation requires rules describing nonfunctional capabilities and properties of services as well as syntactic and semantic descriptions, and organizes its approach around policy rules specifying how the actual administrative unit handles contracts, negotiations, and other pragmatic, contextual aspects. The Web Service Modeling Ontology (WSMO – http://www.wsmo.org/) is a conceptual model that can be used to create domain ontologies, as well as for specifying nonfunctional properties of services (such as their cost and quality), describing the goals of the functional capabilities, describing service-related information (such as the functional capabilities, its communication protocols, or its decomposition in terms of other services), and specifying mediators by identifying and removing obstacles to compatibility at the level of data format or underlying processes.

The EU project DIP (Data, Information and Process Integration with Semantic Web Services – http://dip.semanticweb.org/) developed an e-government use case in close collaboration with Essex County Council, a large local authority in South East England (UK) containing a population of 1.3 million, to deploy real-world SWS-based applications in such a way as to support reuse and composition [12]. In order to provide semantics and step toward the creation of added value services, DIP adopted WSMO and IRS-III, a tested implementation of this standard [53]. Since government legacy systems are often isolated

– they are not interconnected and/or use distinct technological solutions – the innovations of the DIP approach firstly enabled the data and functionalities provided by existing legacy systems from the involved governmental partners to be exposed as Web services, which are then semantically annotated and published using SW service infrastructure.

Using SW services with formal descriptions of their semantics, machine interpretation enables discovery by matching formal task descriptions against the descriptions of the services, mediation, and composition [54]. The advantages of this for e-government, as set out by Gugliotta et al. are:

- Providing added value joined-up services by allowing software agents to create interoperating services transparently and automating integration.
- Enabling formalization of government business processes, allowing a common understanding and visualization across heterogeneous administrative units, possibly promoting reengineering.
- Reducing risk and cost, moving from hard coded services to reusable ones.
- Allowing one-stop customer-focused and multiple viewpoint access to services.

Gugliotta et al. [55] based one of their scenarios around the "life event" concept, envisaging an active life event portal. WSMO augmented with ontologies based around the "life event" concept is also the basis for projects such as Access-eGov [20], while Wang et al. [56] describe the extension of WSMO to encompass public administration concepts, linking the generic public service object model of the GEA with WSMO to produce WSMO-PA. Further experiences of using the GEA are presented in [43, 57, 58].

20.1.3.6 Challenge 6: Privacy and Access Control

Privacy and trust are essential factors for e-government, and the SW would seem to be a valuable tool to promote them, with ontologies providing vocabularies to express privacy and security policies machine-readable allowing reasoning over them (the variability of people's attitudes to privacy, and therefore the policies they will endorse, is one major reason why SW technology has great potential here). Where a government possesses large quantities of information, the guarantor of privacy is often what might be termed *practical obscurity*: the phenomenon that information, often paper-based and held in discrete repositories, though theoretically in the hands of governments, is actually not useful because it cannot be found effectively in a timely way [59]. This is particularly true of information that does not exist explicitly in government archives, but could be deduced from information held in two or more other sources. Hence in some polities lack of trust in government, however well-founded, can lead to skepticism regarding the benefits of efficiency, because efficient use of information can lead the citizen to feel that their personal affairs and actions can more easily be scrutinized by government. Hence, when it comes to the citizen trusting the government's use of IT innovations, privacy issues loom large.

However, there is relatively little work in the field as yet discussing privacy and trust. One can only speculate why there is such a lack. It may be that, in such a complex field where relatively little progress has been made in the more transformative areas, most effort has been focused on proofs of concept, defining architectures and services that deliver in real-world contexts, with privacy considered as a bolt-on to be addressed in future prototypes that are closer to genuine implementation. For example, Klischewski and Jeenicke [60] explicitly remark that their prototype system for the Hamburg area focused on service functionality, and privacy was a secondary issue not addressed, despite the fact that they focus strongly on requirements analysis. Peristeras et al. [4] argue that, although privacy will inevitably be a problem with linked open government data, "there are many 'safe' candidate start-up areas for which information reuse looks quite harmless: for example, data related to geography, statistics, traffic, and public works." Indeed the data.gov.uk work explicitly focused on nonpersonal public data to avoid issues around personal and private data. Nevertheless, citizens, developers, and governments may discern benefits in combining personal and nonpersonal data held by the state, in which case privacy concerns will need to be addressed.

Medjahed et al. [52] describe the privacy solutions used in WebDG system, an architecture for providing customized SW services in the USA in the domain of benefit collection (WebDG is a generic system, based on WSDL). The benefits area has obvious privacy issues (citizens must disclose their social security number and salary, for instance), and the information is even more sensitive when combined illicitly with other information (e.g., from the tax office, about earned income). They point out that security does not necessarily produce privacy, because one may want one's information kept private from people operating a secure system (e.g., government employees). WebDG has a three-layer privacy model. User privacy preferences are specified through editable profiles (which can be overridden by government regulations), and WebDG assigns a user a credential on the basis of these. The credential determines access and read/write rights to data objects. Each e-government service has its own privacy policy specifying a set of rules applicable to all users, stating the purposes for which the service can use information collected, the length of time, and the conditions of information storage, and specifying how and to whom the information can be released. And data also has associated with it a privacy profile that determines the access views it allows to those who access it. WebDG also contains a series of modules intended for enforcement of policies.

Weitzner et al. [61] argue that attempting to control access to data in a world of digital information, where copying and transmission are virtually costless and the large size of communicating networks means that dissemination can be extremely wide and fast, is fundamentally mistaken and destined to be behind the curve of progress. The authors argue that data use is a more important thing to focus on, and that data access control should be supplemented by legal and technical mechanisms for transparency and accountability, allowing harms to be traced and rectified or compensated. The Policy Aware Web (http://www.policyawareweb.org/) is a conception of the SW intended to allow this idea to happen.

Weitzner et al. [61] describe a transparency and accountability architecture TAMI designed to address data misuse in the scenario of government data mining of transport information in order to identify potential terrorists. The architecture needs an inference engine to support analysis of data and assess compliance with rules of access and use, a truth maintenance system containing proof antecedents from the inference engine, data provenance and justifications of the antecedents, and a proof generator. This architecture is intended to identify factually incorrect antecedents (e.g., misidentifications of passengers), assess compliance with information sharing rules prior to data transfer, and to check that actions are consistent with information usage rules.

This is an interesting example of the way that technical developments can affect process reengineering. Weitzner et al. [61] first argue that a privacy focus on data use is far more realistic in the current technological climate than a focus on access, and then uses the proposed architecture as an intervention in the public policy agenda. It claims that "policy aware systems bring added focus to policy questions regarding data mining privacy," and that to realize the promise of transparency and accountability rules, a series of legal issues will have to be resolved – for example, a question such as "under what circumstances, if ever, can inferences generated in one type of profiling system (e.g., anti-terrorism passenger screening) be used to further criminal investigations?" In this way, the SW can be transformative in e-government, by posing and demanding answers to hard questions. This brings the discussion onto the next challenge, about reengineering and change management in general.

20.1.3.7 Challenge 7: Reengineering and Change Management

The issues underlying reengineering should not be underestimated. It is very hard to turn staff-intensive and paper-based systems into automatic digital systems, especially when the reengineering might be entrusted to the very staff whose jobs are under threat from the transformation, and whose incentives are at best mixed. It is also very hard to integrate systems across platforms to provide seamless service for the citizen. Furthermore, the chief driver of change is not pressure from without, but rather consciousness within government of the opportunity costs of not upgrading systems – a notoriously weak driver. As a result twenty-first century e-government systems are often grafted onto nineteenth century bureaucracies. This locks in the high costs of integration, and tends to create islands of e-government rather than allowing an integrated approach across government.

Stojanovic et al. [62] argue that the use of semantic technologies to describe e-government services can improve the management of change in the resolution of either static modification or dynamic modification.

- Taking into account an enormous number of public services and dependencies between them, as well as the complexity of interpreting and implementing changes in government regulations, the process of reconfiguring the existing legacy systems (the so-called static modification) seems to be quite complex. Indeed, an efficient management system must provide primitives to allow the progressive refinement without rewriting it from scratch, and must guarantee that the new version of the service is syntactically and semantically correct. However, an efficient management system for resolving *static* changes in an e-Government domain does not exist.

The SW, they argue, can provide tools for addressing this problem. This is agreed by Klischewski and Ukena [63], for instance, although their focus on requirements analysis biases their approach toward the creation of SW services in a static regulatory context, which is certainly important, yet ignores the perhaps more frequent situation where the regulatory context is dynamic, and hence methods to deal with dynamic modification are also needed. Peristeras et al. [4] argue that annotations are important in the creation of public knowledge, as "the key to knowledge creation lies in the mobilization and conversion of tacit knowledge," and suggest that documents be enriched with knowledge drawn from other documents and communities.

Accordingly, Stojanovic et al. [62] add ontologies for understanding the evolution of the system's ontologies as procedures and regulations change the usage of the system by end users (the ontology is used to structure the usage log) and the lifecycle of the system as a whole (which describes information flow and decision-making in public administration). Such a system is intended to locate out-of-date services (i.e., ones that have ceased to be relevant because of changes of regulation, possibly in quite a remote domain), and manage the change and change propagation process – thereby creating a bridge between decision- and policy-making and technical realization. To do this, it is not enough to provide semantics for the relevant services, but all the dependencies between stakeholders that come together to create collaboratively the administrative processes need to be modeled. The management of change, including representing change at the right granularity, and propagating change through the system, is described in [64].

The large-scale model, including models of all the services included, is then used, together with a set of constraints that must be satisfied for all services (included in an ontology to describe the services), to ensure consistency with each other, and with the constraints that together define an ideal for e-government services. A typical set of models using such a methodology can be clustered according to whether they are meta-ontologies that define the modeling language for e-government services, domain-specific ontologies, or administration ontologies. The meta-ontologies include legal ontologies, organizational ontologies, and the lifecycle ontology, as well as the common device of a life event ontology [65]. Hinkelmann et al. [66] found such a methodology very useful in dealing with the loose federal structure of government in Switzerland.

The constraints include "each service has to have reference to one business rule" (i.e., each service has to be invoked by some process, otherwise it should be deleted), "each service input has to be either the output of another service or specified by the end-user" (this enables changes in one service to propagate to other services that use its outputs), and "if the input of a service subsumes the input of the next service, then its preconditions have to subsume the preconditions of the next one" (preventing nonoptimal configurations in the event of a change, so that for instance if one service has as a precondition that the user is over 18, there is no point having services that use its output having a precondition that the user is over 16). Most of the constraints are domain-independent in this way, although as Stojanovic et al. [65] point out, domain experts are extremely important in this methodology as they possess very good knowledge about the effects of e-government in their area, and they understand legislation, the legislative history and context, and the public view of it. The constraints are required, because changes in administrative processes would not necessarily introduce inconsistencies with the ontologies, or with the services themselves – the constraints are needed for change management. Once there has been a change in administration or regulation, a procedure analyzes the current model to see where there are weak points, if any, signaling this to managers enabling them to produce technical fixes.

Stojanovic et al. [62] argue that such a system would ultimately be able to suggest changes that improve services even where administrative change has not taken place, by being used to monitor e-government service execution in the context of citizen comment and complaint about the system. Stojanovic et al. [67] explored this possibility in the FIT project (Fostering self-adaptive e-government service Improvement using semantic Technologies), in which the methodology and models they outlined were used to provide a personalized and inclusive (i.e., taking user needs and preferences into account) front office, and a flexible back-office that supports knowledge sharing, best practice, multi-context views, and context-aware service delivery. So-called front-office integration takes precedence over costly back-office integration. With front-office integration, service integration does not require intervention in their implementation, and it is neutral between newly built services and legacy services. Integration need only go as far as is driven by demand, and indeed special purpose customized suites of services can be created for individuals [26]. The system can learn preference rules from citizen interactions, and modifies the service portal according to changes in user requirements and feedback. The FIT ontology, an e-government upper ontology, is the basis for navigation and inference across the system.

A similar idea, of using a model distributed about ontologies, including ontologies of change, informs the OntoGov project (http://www.hsw.fhso.ch/ontogov/) [65, 68, 69], an EU project that ran between 2003 and 2006 to develop, test, and validate a semantically enriched and ontology-enabled platform to facilitate the management, including composition and reconfiguration, of e-government services. This defined a high-level generic ontology for the service lifecycle covering all the phases from definition and design through to implementation and reconfiguration to provide the basis for designing lower-level domain ontologies specific to the service offerings of the participating public authorities. It also developed a platform to enable public administrations to model the semantics and processes of e-government service offerings at different levels of abstraction, and to enrich the provision of e-government services to citizens and businesses with useful metadata. The aim of OntoGov was to bridge the gap between policy-making and the technical realization of e-government services, making knowledge explicit and supporting the management of services over their lifecycle.

As well as evolution of ontologies, versioning of data is extremely important, and is one of the main issues addressed by data.gov and data.gov.uk. For instance, most of the PSI datasets released by the UK government contain past versions and are liable to change in the future. Regional boundaries and legislation are two simple examples of domains that change regularly, and which are crucial to any e-government application. The technical issues behind versioning in linked data and the solution that data.gov.uk is implementing by using RDF graphs will be summarized in the example application section below.

Ultimately, reengineering depends not only on the use of semantic technologies, but also on the *perceptions* of technologists and administrators underlying the management of change. The SW has something of a bad reputation when it comes to implementation issues, and it can be hard to get buy-in from administrators who anticipate a large initial investment cost following a disruptive information management phase. These worries of perception are overblown, and will be discussed in more detail in the next section.

20.1.3.8 Challenge 8: The Perceived Costs of Implementing Semantic Web Technology

As Fishenden et al. [70] argued about the problem of interoperability in e-government, "Interoperability is concerned with more than just low-level technical issues. To be successful, interoperability programs need to address a range of issues that span technical, semantic, cultural and organizational interoperability as well as security, confidentiality, data protection, privacy and freedom of information obligations." Hence there are a number of diverse obstacles to using SW technologies and formalisms to address e-government problems. The investment in reengineering may look daunting in terms of initial cost, and possibly even in terms of expected benefit, at least until enough concrete and irrefutable evidence has been amassed that the SW does deliver. Well-placed champions will be important.

This leads to a serious pragmatic issue about how such champions should be deployed, and where they should be within the organization. Is it better to use a "top-down" approach to conversion, engaging a powerful person or administrative body high up in the hierarchy, which prescribes methods for interoperability, determines how many resources will be devoted to the reengineering, and is prepared to provide incentives for change? Or alternatively should a "bottom-up" process allow interoperability to emerge via smaller units at the leaf nodes of the hierarchy engaging in information sharing, perhaps in a series of bilateral arrangements, culminating in the emergence of a de facto method, which then can be formalized?

In a complex and fragmented domain such as e-government, it is clear that some bottom-up processes will be required [11], because so many different cultures (e.g., formal vs. informal models) and practices will be used. Nevertheless, some kind of top-down pressure will also be required (a) to ensure that those departments reluctant to change still undergo the process, (b) to ensure consistency between approaches and to avoid reinventing the wheel (e.g., by sharing of ontologies), (c) to steer reengineering strategically, and (d) to provide rewards and incentives for good practice, and manuals of best practice. Klischewski [23] argues that the way to reconcile the top-down and bottom-up approaches for producing semantic integration and common ontology acceptance requires three essential steps.

- First, metadata standards such as eGIF or other initiatives drawing on standards such as Dublin Core should define terms for administrative purposes.
- Second, integration should be framed through upper and lower domain ontologies, finding common ground on generic concepts and also on elementary concepts, and providing mappings between the often divergent concepts in between.

- Third, intermediate between conceptual schemas acknowledging the importance of each departmental culture, while creating reference ontologies to support mappings between schemes.

Wagner et al. [11] suggest that the complexity of existing e-government systems is likely to be an important challenge to the SW in this area, and insist that any approach must incorporate the bottom-up methodology in at least some respects. They propose the design of a two-layer wiki with a semantic layer describing semantic relationships, maintained by a community of users, although as they admit such a methodology raises questions about trust (of the information created) and confidentiality. However, as they maintain, there is a broad trade-off between accuracy and trust on the one hand, and broad and easy participation on the other. The overlaid semantic wiki would identify semantic relationships at the content layer and express them in a separate structure in machine-readable wiki pages. Pages expressing the logic of the e-government SW would then allow, for instance, a matching of related pages; question pages, which state questions and link to multiple options, could then be associated with explanation pages that provide explanatory content. This sort of structure could be laid over legacy pages as well as newly generated content, allowing machine interpretation and automatic provision of links.

Nevertheless, as Fishenden et al. argue [70], one needs to look beyond the technical as the spread of SW technology in e-government is assessed, whichever problem one is concerned to address. It is probably fair to say that many organizations still view the Semantic Web with some skepticism, and this culture needs to be addressed. It thrives no doubt partly because of a suspicion that administrators are expected to pioneer an approach in which there are few "quick wins." Moreover, there may be worries about the cost and privacy issues that arise whenever increasing amounts of information are linked into the Web.

Some have produced a reverse pragmatic argument for using semantic technologies in e-government, that even if there are high perceived costs, they are an important means to move toward transformation. If the aim is to produce horizontally and vertically integrated e-government, a semantic representation of government data, maximizing the use of external data and also releasing government data to the outside world will be important steps toward that goal. On the other hand, as Koné at al. argue [71], the SW cannot do everything. The semantics will help, but culturally, cooperation will still need to be fostered, while organizationally, different administrative environments will still need to be brought together to use, or at least to make reference to, standard reference ontologies. The W3C's interest group on e-government has published a working draft on how to publish open data [72], and recommends the following steps: (1) publish data in its raw form if it is structured and can be extracted from the document; (2) create a catalogue with documentation of what is available; and (3) convert the data so it is human- and machine-readable, with semantics, metadata, and URIs.

The particular case of small governments has also been studied [73]; such governments are thought to lack the management and reengineering resources to improve semantic interoperability of distributed e-government services and resources, partly because of the complex requirements highlighted earlier, where the creation and publication of information with machine-readable annotations is not a simple process. The level of support required is large, and the initial costs can also be very risky to take on. It follows that quick wins and a lowering of ambition (e.g., not using a single elaborate ontology, but multiple overlapping small-scale ontologies) will be important factors here [74]. Klischewski's work in Schleswig-Holstein, a small German state containing a thousand heterogeneous municipalities, revealed that maintaining up-to-date and standardized information bases locally for use by the state government is hard (small municipalities do not have the workforce), while central databases are inefficient. Therefore any central e-government application has to obtain the required information from heterogeneous local sources, and motivating such municipalities to cooperate requires a deep and sympathetic identification of their requirements and constraints, and some transfer of resources downward (e.g., for provision of new methodologies and tools).

Given that an opportunity for the SW to deliver benefits to government has been identified, and that some person or people within an appropriate institution are keen to implement the technology, it is important to understand and counter the common misconceptions held about the SW [74]. Common misconceptions are described in ➲ *Table 20.1*, along with a brief reality check.

20.2 Conclusion

The challenges outlined in this section are the sorts of challenge that present themselves to any organization attempting to solve similar problems or grasp similar opportunities for integration and upgraded information management. However, governments feel the difficulties more strongly, as the pressures on them are more acute and diverse. The responses outlined to the challenges above have stressed the pragmatic element, as well as the importance of pilot schemes and "quick wins." Progress has been patchy, and there is plenty more to be done. However, enough work has been surveyed to imply strongly that the technical issues are not intractable, even if managerial and political problems require a great deal of will and skill to solve. And as will be seen in the next section it is possible to launch national initiatives where the costs are dramatically smaller than those associated with traditional large-scale IT projects, in part because the main effort is in persuading the government to publish data in SW formats that support rapid reuse and exploitation outside government.

20.3 Examples: The Release of Public Linked Data in the UK

The challenges outlined in the previous section are pervasive throughout the public sector, and SW approaches must be alive to the dangers of ignoring them, especially as in ➲ 20.1.3.8 because inaccurate negative perceptions of the SW can be extremely damaging at the crucial early stages of a project. In this section, two example systems will be examined that look at two aspects of the same problem, the provision of public datasets

Table 20.1
Some common misconceptions about the Semantic Web (From [74])

Misconception	Reality
Everyone must agree to the same terminology to enable data and information sharing.	Different terminologies can be used by different departments, and linked to each other to ease sharing and communication.
Ontologies are typically large and complex.	Heavyweight and complex ontologies encode domain knowledge. Such ontologies are not always needed. Much can be done using relatively lightweight ontologies.
Ontologies are expensive to design, build, and maintain.	Some ontologies encode a great deal of domain knowledge and can be expensive to build. In these heavyweight ontologies the larger the potential user community the more the cost of construction is offset. Lightweight ontologies can have wide applicability and can be very cost-effective to build in terms of overall utility to the community [75].
Information and data must be taken out of current knowledge management practices, expensively converted to RDF, and then everything must be replaced with new standards and technology.	RDF creation can be automated, using simple scripts, APIs, or conversion languages (e.g., GRDDL). Data and information can be kept in their current formats, and cached or exported in RDF.
Providing access to data and information will benefit consumers and competitors, but there are no quick wins for the provider.	In the long run, exposing data and information will provide gains for the owner as well as for the whole network, just as exposing documents provided gains when the WWW took off. In the short term, much reuse of information is facilitated, which results in quick wins for any organization with a large quantity of distributed legacy data in heterogeneous formats.
The promiscuous release of data and information will be a privacy nightmare.	Standards are being developed to control access and reuse policies. In the meantime, as with conventional databases and Web technologies, organizations can pick and choose what data and information to expose and share.

for public use. Specifically, they are both addressing the challenge as in ➲ 20.1.3.3 of publishing information, yet it is clear to see that approaches to this challenge in this space will have ramifications to many if not all of the other challenges. *Publication* demands a stance on *representation* (an obvious question: should data be re-represented?). A sufficiently well-crafted representation, combined with the publication of data and the ability to link with other databases and knowledge bases, will allow bottom-up *integration* to be performed by people with specific information needs, whether they are members of the government, engineers of *services*, or simply citizens writing or using mashups to re-present information, empowering them in their local communities, or helping them hold their government to account. Conversely, a poor representation will render integration hard if not impossible. The facilitation of *search and retrieval* is of course the point of publication. Protecting *privacy* is clearly vital, and needs to be prominently addressed in any publication strategy. *Reengineering* will be required, and furthermore negative *perceptions* will have to be addressed to avoid government inertia.

The two examples discussed below have been selected to illustrate in particular two very specific points. The first example of AKTive PSI will demonstrate how the process of re-presenting information can be conducted in order to defuse negative perceptions of reengineering, change management and the SW itself; its focus is therefore primarily methodological. The second example, of data.gov.uk, focuses on the technical means for supporting representation and publication for the citizen to allow integration and provision of new inferred information, mashups, and services from existing information.

20.3.1 AKTive PSI

The first example is a detailed consideration of an early pilot or proof of concept for the integration of government information using reusable and linkable formats suitable for SW technology, the AKTive PSI project [39], which informed government thinking on information policy in the UK to an unusual degree in the SW world. The small-scale success of AKTive PSI in 2006–2008 paved the way for the more ambitious data.gov.uk site, which followed in 2009–2010. The ideas behind AKTive PSI were an important step in the "semanticization" of e-government, and understanding of how to represent government information to promote reuse in accordance with the Web's and the SW's model of value. The focus in this section is on the pragmatic reengineering of the systems and processes that use information, and how mandating good information representation, annotation, and publication policies is extremely important. Where possible, complexity of modeling is sacrificed to low-overhead practices, resulting in a tractable process even for small sectors of local government [74].

20.3.1.1 Context

The Office of Public Sector Information (OPSI) is responsible for the management of all of the UK government's intellectual property, including setting standards, delivering access, and encouraging the reuse of PSI. OPSI also has an important role as a regulator of holders of public-sector information (e.g., the Met Office, Ordnance Survey) for their information-trading activities.

Information policy developed rapidly in the UK in the early twenty-first century, with Freedom of Information legislation as well as adoption of EU Directives, but for a long time no large-scale work was done to research the potential for reuse using SW technologies and approaches. OPSI initiated a small project together with a UK-based project Advanced Knowledge Technologies (AKT – http://www.aktors.org), called AKTive PSI [39], as an exemplar to show what could be achieved if public-sector information was made available for reuse in an enabling way [40, 76]. Throughout the project, there were regular consultations with many governmental organizations, including the London Boroughs of Camden and Lewisham (local government), Ordnance Survey (the UK national mapping agency), The Stationary Office (the privatized but official publisher of the UK government), The Met Office (the UK's national weather service), The Environment Agency (the body with responsibility to protect the environment), The Office for National Statistics (ONS, the UK's provider and publisher of official statistics), and several others.

Some of the direct outcomes of AKTive PSI were: (a) the *London Gazette* (the official newspaper of public record in the UK – http://www.london-gazette.co.uk/) building OWL ontologies to represent parts of their data, and working toward publishing this data in RDF; (b) the development of a URI schema used to generate URIs for government official legislations and copyright statements; (c) Camden Borough Council added a SW engineer to their staff force to help the council in their effort to join the Web of Linked Data; and (d) the Ordnance Survey continuing their work and research in SW, building a number of ontologies and releasing several datasets.

The initial aims of the project were to draw together a sufficiently large set of heterogeneous information from a selection of public-sector organizations in order to: (a) explore how SW technology could help turn government information into reuseable knowledge to support e-government; (b) investigate the best practical approaches to achieve this goal, in terms of collecting data and constructing ontologies; (c) show how data can be integrated, and identify existing government taxonomies that are useful for this task; and (d) provide evidence of the added value from undergoing this process. Note the strong pragmatic bias in these goals.

To help focus the requests for content, information was collected from the geographical area covered by two of the participating London local authorities, Camden and Lewisham.

20.3.1.2 Public-Sector Datasets

Several organizations who participated in AKTive PSI made some of their databases available for the project (➋ *Table 20.2*). Note the heterogeneous nature of this data, and the standard formats in use. A number of scripts were developed to convert them to RDF automatically, in correspondence with their designated ontologies.

20.3.1.3 Ontology Construction

One of the AKTive PSI principles for building SW applications was to ensure the ontologies built for the provided datasets were of low complexity and limited in scope and size. Small ontologies are cheaper and easier to build, maintain, understand, use, and

Table 20.2
Datasets provided to AKTive PSI, the number of RDF triples generated for each dataset, and a description of what the data describe (From [39])

Camden Borough Council			
Land and property gazetteer	2.3 M	Excel	Properties in Camden, full address, coordinates, type (residential/nonresidential/mixed).
Food premises	84 K	Excel	Food related premises in Camden, their business names, hygiene inspection results, addresses, (e.g., restaurant, school, bar).
Local businesses	170 K	Excel	Businesses in Camden, names, addresses, contact info, and type of business.
Licenses	100 K	MSSQL	Licenses for businesses in Camden, their addresses, license types, and expiry dates.
Councillors and committees	29 K	Excel	Councillors and committees, subcommittees, who sits on which committee, councilor's personal information.
Meeting minutes	106 K	Text	Web pages of committee meeting's minutes.
Lewisham Borough Council			
Land and property gazetteer	4 M	Excel	Properties in Lewisham, their full addresses, and coordinates.
Property tax bands	10 K	Excel	Tax property references, description, rate payers, rate value, and one-string addresses.
Ordnance Survey (data for Camden and Lewisham only)			
Address layer 1	768 K	XML	Data about buildings, addresses, and coordinates.
Address layer 2	11.7 M	XML	Data about buildings, addresses, and coordinates, and building classifications (e.g., hospital, university).
PointX POI	467 K	XML	Various landmarks and businesses, with names, addresses, and coordinates.
The Stationery Office London Gazette (entire database was provided, but only that listed below was used)			
Administration notices	120 K	Text	Notices for the appointment of administrator for corporate insolvencies.
Deceased estates	3.2 M	Text	Decease notices of individuals, names, addresses, description, and date of death, address of representatives.

commit to. It transpired, as is often the case, that no databases held by the participating organizations required more than a relatively small number of concepts and relationships to represent the stored information.

When building these applications it was important to show that ontologies are not hard to build if their purpose is representing databases and information assets of circumscribed scope, and that it is not necessary for everyone to come to a common, agreed consensus on vocabulary, but that through ontology mapping techniques, local terminologies can also prove very useful. It is interesting to note that the average number of classes in the ontologies was 30, with a median of 10.

➲ *Figure 20.1* shows an example of an ontology from AKTive PSI that describes, in very simple terms, Camden's Land and Property Gazetteer. Each council in the UK has a database for the properties that exist in their administrative region, with simple data such as address, property type, ID, etc. The ontology in ➲ *Fig. 20.1* was built manually to model the concepts necessary to represent this data. Manually building ontologies for databases of such limited scope proved to be practical and cost-effective, especially as this ontology was reused for other similar databases held by other councils.

Representatives from the councils of Camden and Lewisham were given a 1-day course on ontologies, which covered ontology editing tools, and best practices and methodologies for ontology design and construction. The course was aimed at giving them the necessary basic skills and knowledge to start creating and exploring with ontologies. These measures of hand-crafted ontology building and small training courses are relatively low overhead, thereby helping meet the challenges of representing and integrating information and reengineering and change management, while simultaneously ensuring that negative perceptions are minimized.

20.3.1.4 Generating RDF

The knowledge representation language of choice for the SW is RDF, which supports linking via the use of URIs, and reuse across data stores. Representing new data in RDF is one thing, but of course much, probably most, government data will be in legacy formats, raising the important issue of re-representation in RDF without putting too much of an administrative overhead on the process. There are several ways in which this can be done. For example, it is often the case that the data is maintained in live relational databases as part of the company's information flow and data network. In such cases it becomes necessary to use technologies such as D2RQ (http://www4.wiwiss.fu-berlin.de/bizer/D2RQ/) and Virtuoso (http://www.openlinksw.com/dataspace/dav/wiki/Main/VOSRDF), which allow for the data in RDBMS to be accessed in RDF.

The participating organizations provided AKTive PSI with data dumps extracted from their databases. The purpose was not to run live services off their networks, but to showcase SW technology as proof of concept. Therefore, in this project the aim was to transform the data into RDF and store it in triple-stores.

From an ontology it was possible to create instances by running simple scripts over the data/information to produce RDF. The scripts were hand-rolled specifically for the database and ontology which they were linking (reused across similar databases and ontologies). Although they were manually built, a framework for semiautomatic script generation was clearly conceivable. Most of the scripts were written using the Jena API, and were thus reusable and easy to tune for new datasets and ontologies. The participants were shown the relative ease of converting legacy data to RDF using free and simple technology.

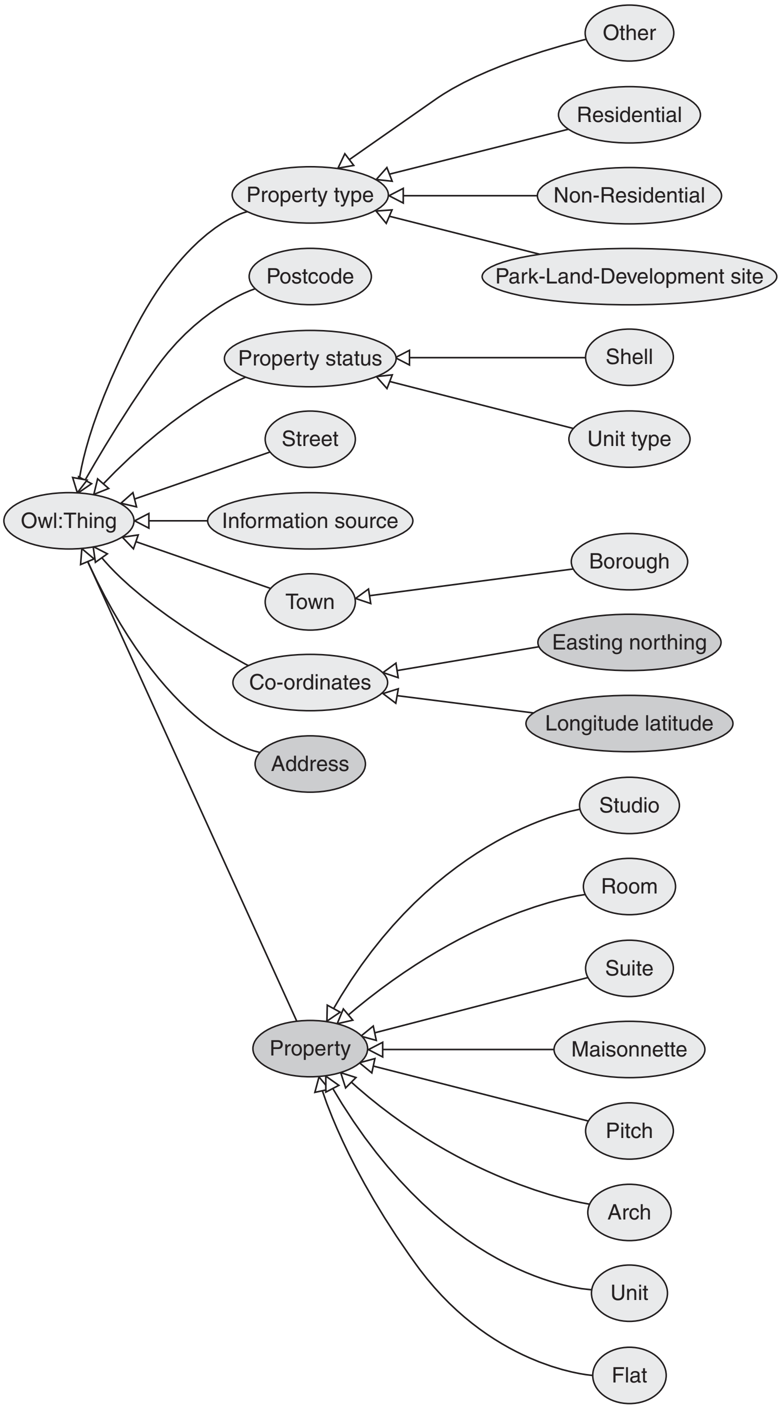


Fig. 20.1
Ontology for the Camden Land and Property Gazetteer, e-government

Although small ontologies were needed to represent the data, a scalable KB to hold the millions of RDF triples generated was also required. AKTive PSI used the 3Store (http://www.aktors.org/technologies/3store/), an RDF triple-store developed in the AKT project, to store the generated RDF files. This triple-store provides a SPARQL end point, a servlet that accepts SPARQL queries and returns results in XML. Publishing RDF in accordance with best practices [77] rendered the data viewable with general-purpose RDF browsers (e.g., Tabulator (http://www.w3.org/2005/ajar/tab) [78]). A key principle is that all entities of interest, such as information resources, real-world objects, and vocabulary terms should be identified by URI references. Once these are in place one can insist that they should be de-referenceable, meaning that an application can look up a URI over the HTTP protocol and retrieve RDF data about the identified resource.

20.3.1.5 Migrating to the Web of Data

Reuse of URIs increases connectivity between the published data and thus facilitates discovery of related data, addressing the search and discovery challenge discussed above [79]. Ontologies support integration using "soft" mappings between concepts and instances that queries or data browsers can follow to find similar or duplicated entities. AKTive PSI used the *owl:sameAs* property to link any mapped entities. By connecting KBs in this way much greater flexibility and querying power was available than the original data structures could provide.

This measure helped demonstrate the added value of using SW technology for publishing and using data. Forming a bigger semantic network by integrating the KBs containing all the data is clearly important in this context, easing communication and data/information exchange between the partners.

Two levels of mappings were performed.

- Mapping of local ontologies. Automatic ontology mapping has been the focus of much research for a number of years [80], and many tools have been developed for this purpose. However, to ensure accuracy, human supervision and intervention is still a necessity when mapping ontologies. Because of the relatively small size of the ontologies, this was not a difficult task in AKTive PSI. In fact, it was easier to map them manually than to correct automated mappings – an important by-product of the effort to address negative perceptions of the costs of implementing SW technology. Because of their expertise in the domain, the individual organizations provided important input to this process; this again was made possible because of the relatively simple level of the ontologies in question.

  As will be shown later, mapping does not have to be complete to be useful. Much value can be drawn from mapping even a small number of concepts.
- Mapping of instances. Because of the data-centric approach it was important to map the instance data to each other as well. Instance mappings have to be done automatically as even in simple domains there will be a lot of instances to map. Automation is

done with simple scripts that search for duplicates of specific types of instances (e.g., postcodes, airplane models). An *owl:sameAs* link can be automatically added between the corresponding instances once such a mapping is found.

These processes create several files that contain RDF *owl:sameAs* triples linking various parts of the data. These files are stored separately from the data, and invoked when querying. To retrieve data from the knowledge base, the applications use SPARQL queries. Because the ontologies and data have been linked as described, it is possible to extract information from multiple data sources.

20.3.1.6 Mappings

Two ontologies for datasets from Lewisham Borough Council were developed, each with classes representing *property*, *address*, and *postcode* (i.e., equivalent to US zip codes). These concepts were linked with *owl:sameAs* to indicate that they represented the same concepts. There were also many simple mappings, such as between the concept *Premises* from the Food Premises ontology of Camden to the *Property* class in the Land and Property ontology. The CROSI mapping library (http://www.aktors.org/crosi/) was used for automatically generating these mappings.

But even simple mappings can be powerful, such as between instances of postcodes, for example, the instance *postcode_N6_6DS* in one KB maps to the instance *pc_N66DS* in another. Since these instances really do refer to the same object it is possible to infer far more by noting the identity. In fact, simply linking to one data object (the postcode) was generally enough to glean useful information from various datasets for the creation of interesting mashup applications.

20.3.1.7 Mashing up Distributed KBs

Once data and information is available in easily parsable and understandable formats such as RDF, mashups become much easier to generate by searching RDF KBs and mashing up data on the fly, one of the advantages of linked data. Two such mashups were created in AKTive PSI. The aim of building these mashups was to demonstrate the benefit of exposing this data to the consumer, and the relative ease with which they can be constructed from semantically represented knowledge.

The Camden Food Premises database gives information about the hygiene check results and health risk of various premises around the Camden area that handle food. The risk categories are given a level between A, which is high risk, to E that is low risk, and is based on the cleanliness of the premises, compliance with regulations, type of preparation that is performed, etc. The Food Premises database contains lots of information on these properties, but displaying this information on a map is difficult because the geographical coordinates are missing from this particular data set.

However, the Ordnance Survey's Address Layer and Points of Interest (PointX) datasets contain easting and northing coordinates for businesses and properties. The instance mapping of postcodes performed earlier helped to cut down the search space for finding matching addresses in the datasets. Indeed, once matches had been found it was possible to assert them as being the same, thereby avoiding the need for searching again.

To create the mashup, a number of SPARQL queries were written that searched for each premise's address from the Food Premises dataset in each of the two OS datasets and once a match is found the coordinates are retrieved and the premise displayed on a Google map. The information from Food Premises together with the mapping between the datasets provides extra context to instances from either dataset. The PointX dataset gains access to the risk level of the food premises (as well as the implicit knowledge that the premises are used for preparing food), and the food premises dataset garners exact coordinates for the premises. ➲ *Figure 20.2* shows a simple Google Maps mashup that uses the mapping to provide a visual display of the food premises dataset.

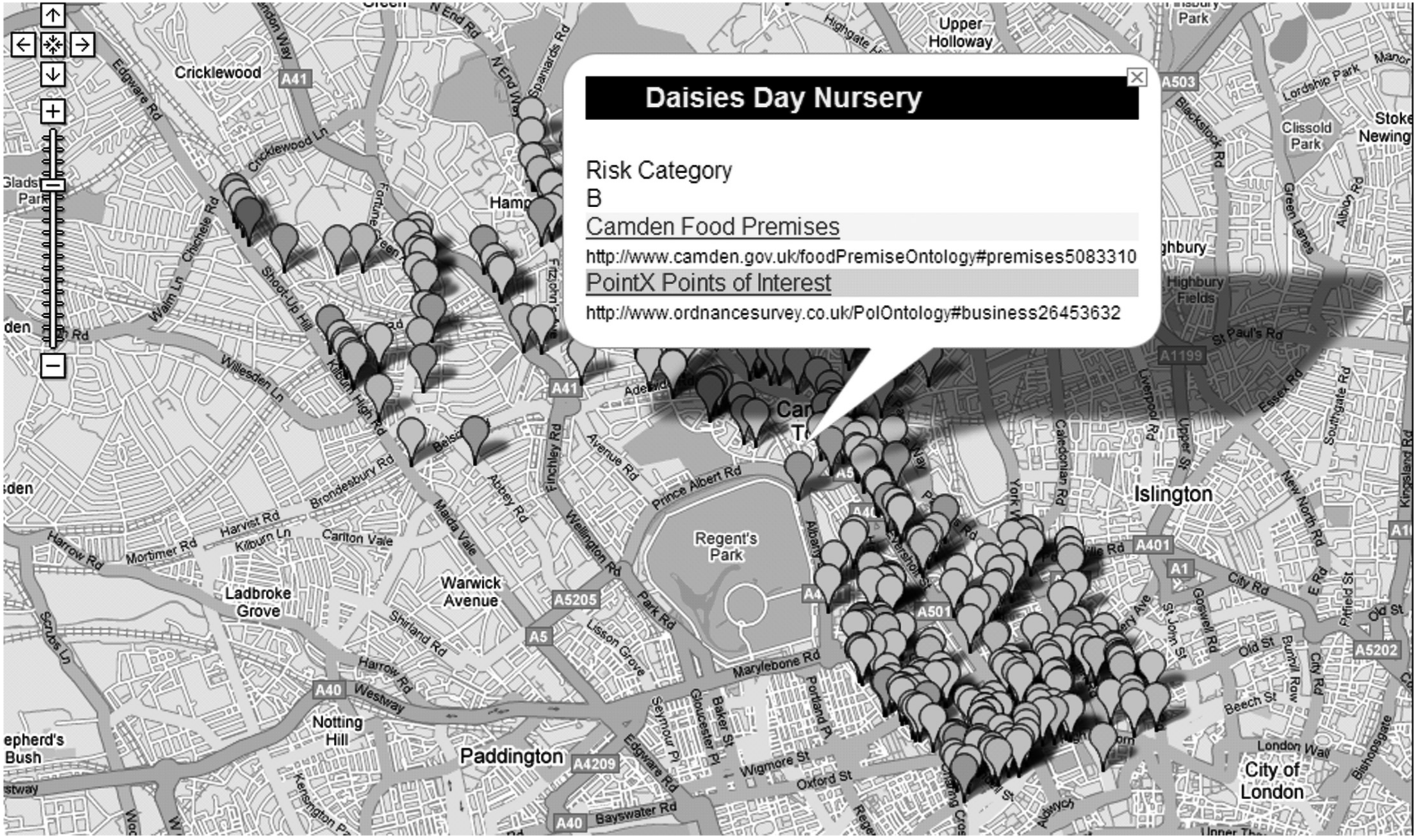
This type of mashup promotes public awareness and indeed commercial competition. For example, it became evident that one particular business that scored within the high-risk category has glowing customer reviews on restaurant review sites across the Internet.

20.3.1.8 Inconsistencies

Data and information integration from multiple sources adds the value of knowledge augmentation and verification. Integrating datasets can provide useful insights into the quality of the dataset for the data provider involved. For example, the Ordnance Survey's *Address Layer 2* dataset provides a list of businesses, including their addresses and their geo-locations, and similarly so does the PointX dataset. However, it was found that the two lists of businesses do not match, for instance some being present in one dataset but not the other. In some examples, the PointX dataset contained several businesses listed at the same address, while only one was listed in the OS Address Layer 2. Was this an error? Perhaps one business took over the building from another, but the lack of temporal information concealed the fact, or perhaps one business is sited in the same building on a different floor to another business. It is difficult to infer an answer, but the integration has at least provided some information about the quality of the datasets and made such comparisons and cross-matchings possible. As noted above with the more complex examples of OntoGov for instance, such models can promote reengineering by helping identify inconsistencies.

20.3.2 data.gov.uk

AKTive PSI was an important proof of concept that led the UK government thinking on data management and integration. As the concept of *linked data* became increasingly important as a stepping stone to the development of the fully machine-readable SW



■ **Fig. 20.2**
Google Maps mashup of the Camden Food Premises dataset made possible by mapping the data to the OS Address Layer II and PointX datasets, e-government

[79, 81], allowing data from multiple and heterogeneous sources to be linked, integrated, and reused, the quantity of linked data available on the Web has grown dramatically.

Where AKTive PSI had demonstrated the viability of using SW technology to add value to government information by integrating it using RDF, ontologies, and SPARQL, a potential extension was to add government data to the Linked Data Web by the use of resolvable URIs. If resolvable URIs could be used for reference in public-sector data, then it could be added to the Web of Linked Data with all the benefits that entail [82]. This is the approach underlying data.gov.uk.

20.3.2.1 Introduction and Context: First Steps

The early history of data.gov.uk is outlined in [83]. In June 2009 the then UK Prime Minister asked Tim Berners-Lee and Nigel Shadbolt to act as Government Information Advisors. Their terms of reference included:

1. Overseeing the creation of a single online point of access and working with departments to make this part of their routine operations
2. Helping select and implement common standards for the release of public data
3. Developing Crown Copyright and "Crown Commons" licenses and extending these to the wider public sector
4. Working with the Government to engage with the leading experts internationally working on public data and standards

The project was made public with an online call for help from the UK Cabinet Office that established one of the key principles of data.gov.uk, citizen participation: "From today we are inviting developers to show government how to get the future public data site right – how to find and use public sector information" [84]. By setting up a Google Group the Cabinet Office started to collect, and still collects, Web user opinions about the data.gov.uk site (http://groups.google.com/group/uk-government-data-developers). In this online e-mail group there were more than 2,400 members (as of August 2010) to participate in daily message exchanges about all sort of topics related to Open Public Sector Information (PSI) and Web technologies. During the first phase of data.gov.uk access to the portal was restricted only to members of the Google Group. The goal was to use this online group as input to improve the future site. During the following 3 months opinions and questions were gathered and the group listed valuable issues – technical, social, organizational, cultural, etc. – that needed to be understood and confronted in order to achieve the program's ambitions to put government data in a position where it will be:

- Easy to find
- Easy to reuse
- Easy to license

The result of this collaborative process was made public in January 2010 when data.gov.uk was launched [85]. From that moment on, the portal allowed unrestricted access and any Web user could make use of the data.gov.uk services and access more than 3,763 datasets.

Much of the work behind the scenes was around policy as much as technology. There were significant issues to be resolved to enable a permissive license for anyone to reuse data for any purpose – this has led to a new Crown Licence (http://data.gov.uk/terms-and-conditions). There was work to release significant amounts of Ordnance Survey UK Mapping data (http://www.ordnancesurvey.co.uk/oswebsite/opendata/) using the same permissive license and free of derived data constraints (i.e., the constraint that if one contributes data to another underlying data source – say a map – the owner of that underlying data has the rights to the contributed data). There was an early recognition that much important data lay in local government and not in the central government's hands. This led to the establishment of a Local Public Data Panel (http://data.gov.uk/blogs/local-data-panel) charged with coordinating and promoting transparency and open data policies in local government bodies. And there has been a continuing recognition that this is all work in progress.

The data.gov.uk project's first premise is to open data for reuse in any processable format – spreadsheets, database dumps and XML files, etc. The project has also a clear commitment to Web standards in general and Semantic Web technologies in particular (cf. [82]). Each dataset is to be transformed into Linked Data format after being released, a strategy corresponding to Berners-Lee's call for "Raw. Data. Now!" [86] and his five-point scheme for publishing data [79]. Hence the data.gov.uk site combines Semantic Web and traditional Web technologies giving a single point of access to citizens for finding and reusing data through the following services:

- Semantic Web Features:
  – SPARQL end points organized by the government sector.
  – Linked datasets.
  – Practices and strategies for publishing linked data.
- Non-semantic Web features
  – Forum and Wiki to continue promoting collaboration around the datasets.
  – Searching/browsing datasets. The site acts as a single point of access allowing users to quickly access data by areas of interest such as "crime," "education," "economy," etc.
  – Ideas and applications repository, where users can submit and find applications that are already using PSI data.

One of the key efforts in data.gov.uk is to develop best practice and strategies for publishing PSI UK Linked Data. These practices are directed to any data publisher – part of the government or not – that wants to transform or create a PSI dataset into Linked Data format. Consistent with the earlier work on AKTive PSI, this process compares cost to benefit for adopting Linked Data as suitable technology for publishing PSI data. So far, the project has identified a number of practices as crucial [87], including URI design, versioning, provenance and the development of core ontologies.

Note that the data.gov.uk approach (like the data.gov approach in the USA) addresses the challenge of integration in effect by outsourcing. By releasing data, and by supporting the creative use of mashup technology to bring data together in novel, informative, and surprising ways, these sites open up government data to the power of the Web's scale [82]. Many people can get together to find creative ways of amalgamating databases, and so the bottom-up integration is demand-driven, while actual government involvement need only be limited to the mandating of the representation or re-representation of data in formats designed to support linking. The rest of this section will discuss the formats in use.

20.3.2.2 Government URI Structure

One of the first documents released by the Cabinet Office as part of the data.gov.uk effort was *Designing URI Sets for the UK Public Sector* [88]. This document describes the structure of a Government URI and, together with the Cool URI definition from the W3C [89], represents a guideline for minting URIs based on established and emerging good practices in the Linked Data community. They also meet specific needs for the UK public sector. In summary, these practices include:

1. Use of data.gov.uk as the domain to root those URI sets that are promoted for reuse.
2. Organization of URI sets into "sectors" (e.g., education, transport, health) with a lead department or agency.
3. Consistent use of metadata to describe the quality characteristics of each URI set.

➜ *Table 20.3* shows a summary of the URI structures with examples for various types of URI.

Table 20.3
UK government URI structures and examples [88]

URI Type	URI Structure	Examples
Identifier	http://{domain}/**id**/{concept}/{reference} or http://{domain}/{concept}/{reference}#id	http://education.data.gov.uk/**id**/school/78 http://education.data.gov.uk/school/78#id http://transport.data.gov.uk/id/road/M5/junction/24
Document	http://{domain}/**doc**/{concept}/{reference}	http://education.data.gov.uk/**doc**/school/78
Representation	http://{domain}/**doc**/{concept}/{reference}/{doc.file-extension}	http://education.data.gov.uk/**doc**/school/78/doc.rdf
Definition of the scheme concept	http://{domain}/**def**/{concept}	http://education.data.gov.uk/**def**/school
List of scheme identifiers	http://{domain}/doc/{concept}	http://education.data.gov.uk/**doc**/school
Set	http://{domain}/**set**/{concept}	http://education.data.gov.uk/**set**/school

As can be seen, the domain of the data is indicated by the first fragment of the URI to be minted. The URI scheme defines entities for both the instance and the schema level establishing a clear separation for individuals and definitions by using the "def" and "id" nomenclature. To differentiate documents from information resources it introduces the "doc" suffix. For easy dereferencing of lists of instances – for instance lists of schools – the document defines the URI "set" structure. When resolving a URI "set" one expects to retrieve a collection of the entities represented by that set.

20.3.2.3 Versioning

Legislation, geographical boundaries, and local authorities are just some examples of PSI datasets that change over time. Statistical datasets contain an implicit versioning mechanism, because time is normally one of the axes in multidimensional datasets. This means that the time dimension is treated as a series and the statistical observation has validity only over a temporal instance. For example, one does not tend to create versions of the number of tons of CO_2 emitted in the South West of England for different years; instead that knowledge is represented with a multidimensional dataset where geography and time are dimensions that give contextual meaning to the observation (ontologies to describe such statistical information are given in the "ontologies" subsection below).

Even though most of the PSI data are statistics, there are also important nonstatistical datasets where it is very important to relate to pieces of information that were valid in the past. For instance, it is important to track shifting scheme classifications, like the UK Standard Industrial Classification of Economic Activities (SIC) that released different versions in 1992, 2003, and 2007. The approach to implement such type of versioning in Linked Data is to use Named Graphs [90]. A graph asserts a set of statements and the graph gets annotated with the temporal validity of the information in it. The ontologies used to annotate the validity of a graph and the relations between different versions of the same resources are mainly FOAF [36] and Dublin Core [91].

➋ *Figure 20.3* shows an example of a Linked Data resource, a UK school, which has changed its name. The end of validity of the older graph is stated by the assertion of the Dublin Core predicate *isReplacedBy*. This solution also asserts metadata about the version in each of the graphs so that a software agent that visits one of these versioned graphs will be aware of the temporal validity of the information and how to navigate to other versions of the same data.

20.3.2.4 Provenance

Alongside with versioning it is important to provide information about the provenance of the data. Provenance is not just about the source of the information but also about the ways in which the data has been manipulated in the process of publishing it. The approach adopted is the same as in versioning, where named graphs play a key role in stating

Fig. 20.3
Example of versioning, recording the change of name of a school, e-government

metadata about a dataset. The W3C Provenance Incubator group (http://www.w3.org/2005/Incubator/prov/wiki/W3C_Provenance_Incubator_Group_Wiki) is investigating the state-of-the-art and developing a roadmap in the area of provenance and Semantic Web Technologies. The Open Provenance Model (OPM) [92] is a model that has been embraced by both the W3C Provenance Group and the data.gov.uk project as a standard to represent the provenance of the data. This model among other things consists of an RDF vocabulary, the Open Provenance Model Vocabulary (OPMV), which provides terms to enable practitioners of data publishing to publish their data responsibly.

The Core OPMV predicates are represented in *Fig. 20.4*. The main entities that make up this model are *agents*, *artifacts*, and *processes*. As in almost any other workflow model agents represent the actors that trigger and control the processes; Processes refer to any action performed over artifacts; and artifacts are the input and outputs of the processes.

In the data.gov.uk project this model has been adopted to represent all the different actions that take place in the process of publishing linked data.

20.3.2.5 Core data.gov.uk Ontologies

Another activity in the data.gov.uk project is the development of core ontologies, which act as references for publishing linked data for the UK government. Two efforts within this activity have attracted attention from the linked data community: the SCOVO and SDMX schemas for describing statistical data, and an ontology for describing organizational structures.

SCOVO/SDMX: Statistics are probably the most common type of information in PSI. Almost every PSI dataset contains a multidimensional data structure with temporal series as one of the axes, from traffic flows and CO_2 emissions to indices of deprivation and government expenses. SCOVO (http://sw.joanneum.at/scovo/schema.html) was the first ontology describing multidimensional structures that was explicitly designed to describe statistics. First versions of the statistical data in data.gov.uk used this ontology as main schema.

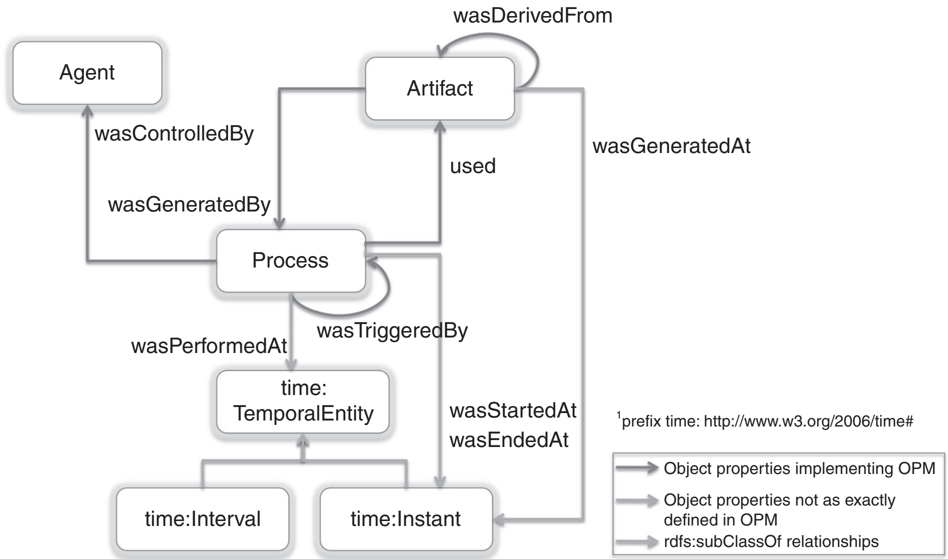


Fig. 20.4
Core definitions of the Open Provenance Model Vocabulary (http://open-biomed.sourceforge.net/opmv/ns.html), e-government

Unfortunately the semantics in SCOVO are not powerful enough so as to describe standards like the Statistical Data and Metadata eXchange standard (SDMX – http://sdmx.org/), supported by numerous institutions and by the UK Office for National Statistics (www.statistics.gov.uk/). As part of a consultation process started by data.gov.uk, a different project was launched to bring together SDMX and SCOVO. This seeks to provide a forum to agree on a SDMX representation for the RDF information model. The result of that effort is a vocabulary, which, by using RDFS and OWL, extends SCOVO to represent statistical observations. Moreover, this vocabulary uses SKOS (http://www.w3.org/2004/02/skos/) to represent classification schemes and code lists (see *Fig. 20.5*).

The main differences between using SDMX/SCOVO and using just SCOVO are the representation of code lists (SKOS), time series, and data groups within a dataset [93]. SCOVO and SDMX are compatible and in fact SDMX extends SCOVO. It is important to notice that both ontologies will coexist in the Linked Data cloud. SCOVO is a lighter ontology than SDMX, easier to understand and more suitable for simple statistics. On the other hand SDMX works better for complex cases and for datasets already in the XML representation of SDMX.

Organizational Structures Ontology: data.gov.uk has been also immersed in the definition of an organizational ontology. A survey on previous ontologies to describe organizational structures (http://www.epimorphics.com/web/wiki/organization-ontology-survey) discovered that most of them were designed to fit their own purposes and that generalized definitions of organizations are minimal. The organizational ontology (http://www.epimorphics.com/public/vocabulary/org.html) that the data.gov.uk group, led by

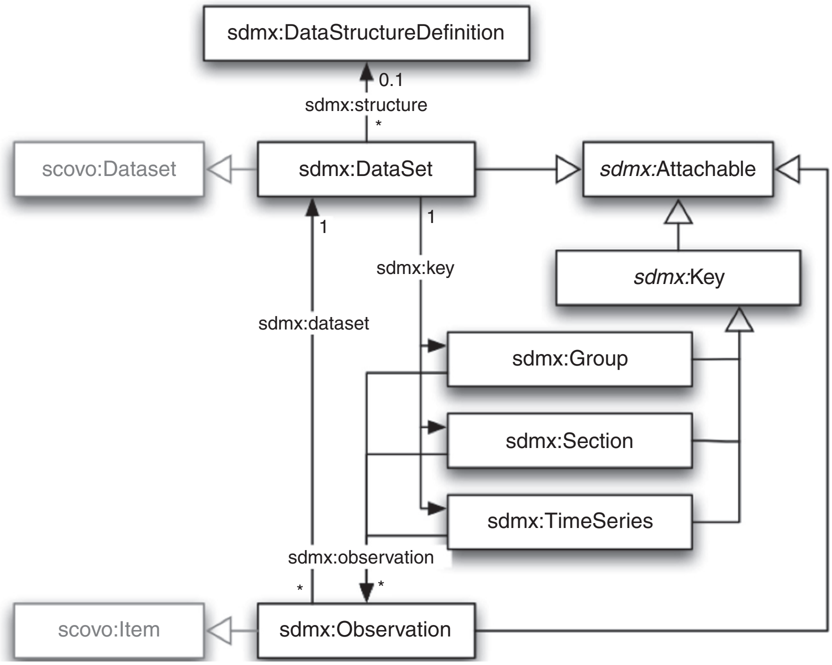


Fig. 20.5
Part of the SDMX RDF Information Model (From [93]), e-government

Epimorphics, developed was critiqued by the Linking Open Data (LOD) community and the schema went through several iterative drafts (http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData). The result is a simple ontology suitable to represent government bodies such as departments, ministries and secretaries, and the relationships between them.

The ontology reuses other vocabularies such as FOAF, Dublin Core, and OPMV. The core of the ontology is represented by *organizations*, *organizational units*, and *agents* (FOAF). It enables the representation of more complex structures by specialization of these concepts. The other important part of the ontology is the representation of relationships between agents and organizations: who reports to whom, and which units are part of others. For this, the ontology has predicates like *memberOf*, *hasMember*, *reportsTo*, *hasMembership*, etc. The ontology, in essence, is simple (see *Fig. 20.6*) but it can be extended to fit complex structural organizations. In fact, this ontology can be reused for representing not just government organizations but also other types of structured entities.

Other components of this ontology provide schema definitions for:

- *Locations*: classes to represent *sites* and *addresses* together with other predicates (baseAt, hasSite, siteOf, etc.) to link these with organizations, agents, etc.
- *Projects and Other Activities*: the class *organizational collaboration* (a subclass of organization) represents a collaborative action between one or more organizations.

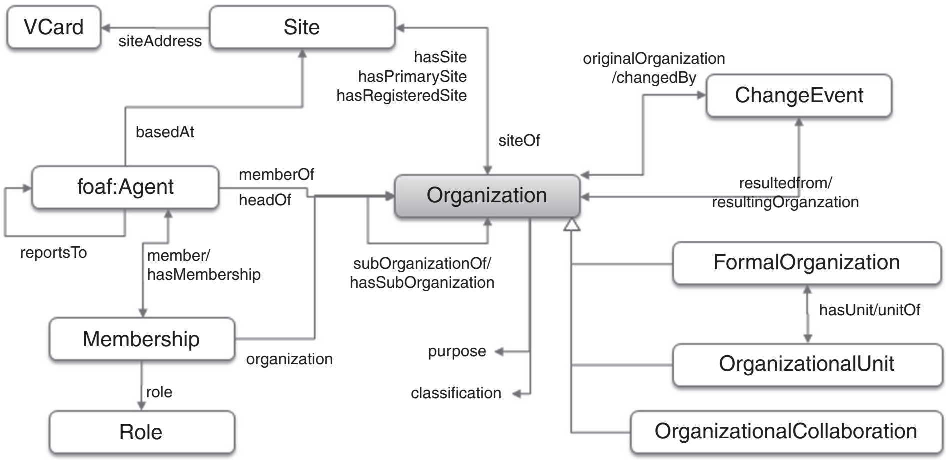


Fig. 20.6
Organizational ontology overview, e-government

20.3.2.6 Linked Datasets

The imperative for any Open Data Government initiative is to have the data available on the Web, ideally in machine-readable form but even better in linked data format. One needs to recognize that for many releasing data is a journey and that the "best should not be the enemy of the good." This is one of the reasons behind Berners-Lee's star system for describing Government data sets (http://www.youtube.com/watch?v=ga1aSJXCFe0) – the rating system simply stated is:

- Make your stuff available on the Web (whatever format)
- Make it available as structured data (e.g., Excel instead of image scan of a table)
- Use an open standard, not a proprietary format (e.g., CSV instead of Excel)
- Use URLs to identify things, so that people can point at your stuff
- Link your data to other people's data to provide context

To this end data.gov.uk has begun to release datasets in linked data format. The ambition is that ultimately this would be the publication format of choice for all government data. In these datasets previous practices have been tested to prove their applicability focusing so far on two domains: *education* and *transport*. A third domain, *geography*, has been also put in place thanks to the collaboration of the Ordnance Survey, which has been an important actor in the UK PSI Linked Data participating in research and releasing key geographical datasets, and as noted earlier was prominent in AKTive PSI. Most PSI datasets contain a geographical dimension so it is extremely important to have access to an authoritative source in that domain. To date (2010) the main Linked Datasets under the data.gov.uk umbrella are:

1. *Education*: The education dataset is an RDF transformation of the UK Edubase database (http://www.edubase.gov.uk/). This dataset is available through an SPARQL end point and the resources in it are also exposed as linked data. It contains information about all the educational establishments across England and Wales, including the schools' type, whether they are religious, their numbers of pupils, and their geographical positions linked to the Ordnance Survey Administrative Geography.
2. *Ordnance Survey*: Ordnance Survey has released the Administrative Geography of Great Britain in linked data format. This is an important data hub for data integrators. In [94] this is demonstrated in the context of a case study where linked data from various PSI sources were integrated. The administrative geography from OS describes each of the types of administrative areas of the UK: European Regions, Unitary Authorities, Counties, Boroughs, Districts, and Parishes; and, more important, the spatial containments between them.
3. *Transport*: Two important transport databases are in the process of transformation into linked data: the public transport network database and traffic flows. The public transport data sources used are the National Public Transport Access Node (NaPTAN) and the National Public Transport Gazetteer (NPTG) databases. The former contains all the public transport stops from airports and ferries to buses and trains; and the latter is a topographic database of towns and settlements in the UK, providing a common frame of reference for NaPTAN.
4. *Multiple Indexes of Deprivation*: In collaboration with the JISC-funded Open PSI project (http://www.openpsi.org/), the Multiple Indexes of Deprivation database (http://www.communities.gov.uk/documents/communities/) has been transformed into linked data. This database contains rankings of different types of social, economic, and cultural indexes in the UK.
5. *Ministers*: The ministers' dataset contains a reference set of the current UK government, their ministers and secretaries as well as the relationships with the different UK government organizations. It uses the Organizational Structure ontology to represent such information in linked data format.
6. *NUTS Geography*: NUTS (Nomenclature of Territorial Units for Statistics) is a standard developed and regulated by the EU members and it is an instrument for reporting statistics. The data.gov.uk project transformed the latest version of the UK NUTS Geography into linked data format.

As a separate effort but in collaboration with data.gov.uk the EnAKTinG project has also released key datasets in linked data format (http://www.enakting.org/gallery/). The EnAKTinG catalogue of datasets contains information about population, CO_2 road emissions, energy consumption, mortality, Parliamentary data (MPs, Lords, and their recorded expenses), and crime offenses. The SPARQL end points of the data.gov.uk datasets (as of August 19, 2010) are given in ➲ *Table 20.4*.

In the same context, but not part of data.gov.uk, it is important to mention the openlylocal project (http://openlylocal.com/). This project is an effort to make UK local

Table 20.4
data.gov.uk accessible SPARQL end points

Dataset	SPARQL end point	De-referenceable URIs
Education	http://services.data.gov.uk/education/sparql	Yes
Ordnance survey	http://api.talis.com/stores/ordnance-survey/services/sparql	Yes
Transport	http://gov.tso.co.uk/transport/sparql	No
Multiple indexes of deprivation	No	Yes
Ministers	http://services.data.gov.uk/reference/sparql	No
NUTS geography	http://services.data.gov.uk/statistics/sparql	No

governments more transparent by accessing their information. As their home page declares they hold information about:

- 158 councils
- 9,946 councillors
- 5,793 committees
- 47,386 committee meetings
- 309 hyperlocal sites
- 29,205 documents
- 52,443 financial transactions

The data available from openlylocal.com can be retrieved in RDF format (see example in *Fig. 20.7*) and the site provides interesting links to other linked datasets like data.gov.uk, Ordnance Survey, and DBPedia. Despite the fact that UK is one of the leading countries in the PSI open data effort and that there is a clear commitment from the UK government to keep improving, the statistics from openlylocal (see *Fig. 20.8*) show that a very small portion of local authorities can claim to have published open data.

20.3.2.7 Linked Data API

The Linked Data API (http://code.google.com/p/linked-data-api/) gained much attention in the course of the data.gov.uk project. This open source project tries to bring together Web developers and linked data technologies. Despite the simplicity of linked data technologies – RDF and de-referenceable URIs – some Web developers have raised issues regarding its adoption as mainstream technology. The linked data API provides tools to overcome this barrier by enabling access to the RDF data model through more developer-friendly technologies.

In the last years simple RESTful APIs have succeeded in getting adopted by Web developers, while key players on the Web like Yahoo, Google ,or Amazon offer access to

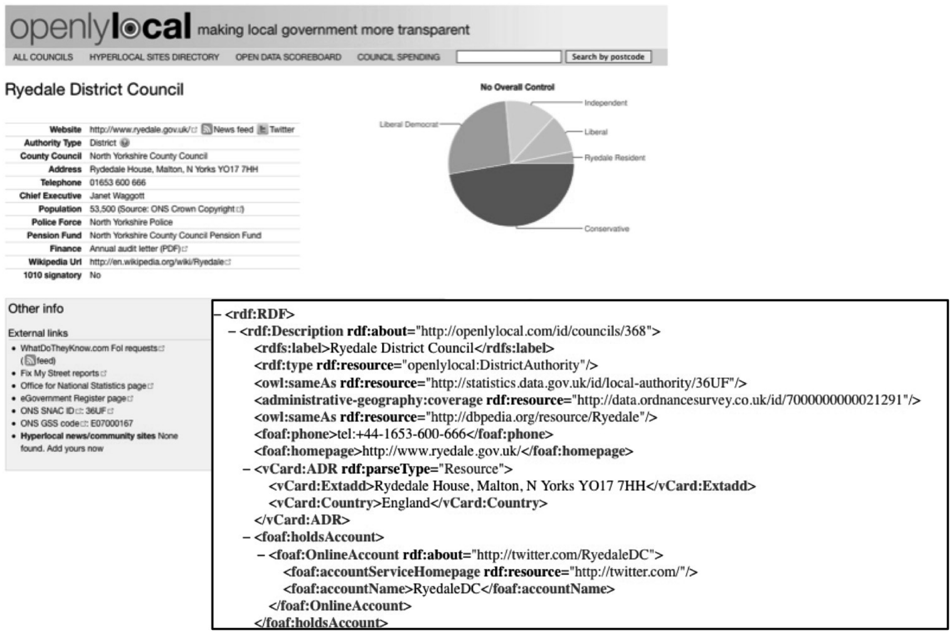


Fig. 20.7
Openlylocal representation of Ryedale District Council showing HTML and RDF versions, e-government

Fig. 20.8
Openlylocal Open Data Scoreboard as of August 2010 (http://openlylocal.com/councils/open), e-government

their data through them. The Linked Data API is a specification that describes how to expose the RDF model using RESTful services and, enables clients to process the data in various formats like JSON, XML, and RDF. The API is intended to be deployed as a proxy in front of a SPARQL query to support:

1. Generation of documents (information resources) for publishing linked data.
2. Provision of sophisticated querying and data extraction features, without the need for end users to write SPARQL queries.
3. Delivery of multiple output formats from these APIs, including a simple serialization of RDF in JSON syntax.

The API maps URL patterns into SPARQL queries. The results of a SPARQL query are passed through two components, a Viewer and a Formatter, before giving a response to the data consumer. In essence the Viewer and the Formatter accommodate the answer in an adequate form. For instance, the following URL pattern exposes a search of schools by district name and different views for the result can be parametrized.

http://gov.tso.co.uk/education/api/school/district-name/{**NAME**}?_view = {**view**}

A Web developer can do a RESTFul request to this service and by binding {**name**} and {**view**} variables he can use the request for his requirements.

Figure 20.9 shows part of the output for the given URL using "Vale Royal" as district name. The parameter **_view** is an API parameter that enables different output customizations for the same type of search – if this parameter is not given the default view is used. The Linked Data API provides pagination for search results. As can be seen at the top of *Fig. 20.9*, references to first and next pages of results are provided in the XML output. For each result of the search a link to the linked data resource is given. The result of the search keeps the original linked data URI so that the client application can always retrieve the RDF representation if needed.

The Linked Data API also provides parameters to customize the selection by changing the query. The requester of the service can modify the select, where, or sort clauses of the


```
<result format="linked-data-api" href="http://gov.tso.co.uk/education/api/school/district-name/Vale%20Royal?_view=short" version="0.1">
  <type href="http://purl.org/linked-data/api/vocab#Page"/>
- <isPartOf href="http://gov.tso.co.uk/education/api/school/district-name/Vale%20Royal">
    <type href="http://purl.org/linked-data/api/vocab#List"/>
    <hasPart href="http://gov.tso.co.uk/education/api/school/district-name/Vale%20Royal?_view=short&_page=1"/>
  </isPartOf>
  <first href="http://gov.tso.co.uk/education/api/school/district-name/Vale%20Royal?_view=short&_page=1"/>
  <next href="http://gov.tso.co.uk/education/api/school/district-name/Vale%20Royal?_view=short&_page=2"/>
  <itemsPerPage datatype="integer">10</itemsPerPage>
  <startIndex datatype="integer">0</startIndex>
- <items>
  - <item href="http://education.data.gov.uk/id/school/111059">
      <label>Cuddington Primary School</label>
      <uniqueReferenceNumber datatype="int">111059</uniqueReferenceNumber>
      <establishmentNumber datatype="int">2196</establishmentNumber>
    - <typeOfEstablishment href="http://education.data.gov.uk/def/school/TypeOfEstablishment_TERM_Community_School">
        <label>Community School</label>
      </typeOfEstablishment>
    - <phaseOfEducation href="http://education.data.gov.uk/def/school/PhaseOfEducation_Primary">
        <label>Primary</label>
      </phaseOfEducation>
    - <gender href="http://education.data.gov.uk/def/school/Gender_Mixed">
        <label>Mixed</label>
      </gender>
    - <religiousCharacter href="http://education.data.gov.uk/def/school/ReligiousCharacter_Does_not_apply">
        <label>Does not apply</label>
      </religiousCharacter>
    </item>
  - <item href="http://education.data.gov.uk/id/school/111056">
      <label>Wimboldsley Community Primary School</label>
      <uniqueReferenceNumber datatype="int">111056</uniqueReferenceNumber>
      <establishmentNumber datatype="int">2190</establishmentNumber>
    - <typeOfEstablishment href="http://education.data.gov.uk/def/school/TypeOfEstablishment_TERM_Community_School">
```


Fig. 20.9
Example of Linked Data API response, e-government

query. For instance, by adding "**_sort** = phaseOfEducation.label,religiousCharacter.label," one would sort the result set by the phase label – primary, secondary, etc. – and the religious character of the school.

As a complementary function to search, the Linked Data API also enables the retrieval of single database items like http://gov.tso.co.uk/education/api/school/100869, which represents a school. It implements content negotiation through HTTP content negotiation. Therefore, a request with "Accept: application/json" in the HTTP header would retrieve the JSON format. As an example, the following command with **curl** would give back the JSON object represented by *Fig. 20.10*.

```
curl -L -H ''Accept: application/json'' http://gov.tso.co.uk/education
/api/school/100869
```

In *Fig. 20.10*, three different sections for the document are shown. The first contains the current data for the selected view. The other two provide links to other formats and views. By changing the **_view** parameter in the request a different representation of the data section could be obtained. For instance the following command produces a different view:

```
curl -L -H ''Accept: application/json'' http://gov.tso.co.uk/education
/api/school/100869?_view = location
```

It has "location" as **_view** parameter, so the output includes the location of the school. As shown in *Fig. 20.11* the data section of this view includes the easting and northing coordinates and the detailed address of the school.

The Linked Data API is independent of programming languages or Web servers; at the time of writing (2010) two implementations – one in Java and another in PHP – coexist.

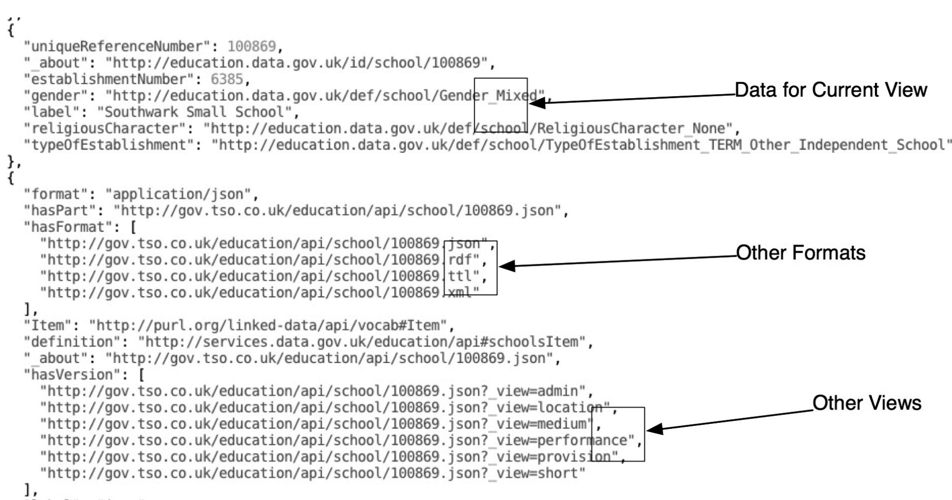


Fig. 20.10
JSON representation of a UK school with the Linked Data API, e-government

```
{
      "administrativeWard": "http://statistics.data.gov.uk/id/local-authority-ward/00BEGH",
      "LSOA": "http://statistics.data.gov.uk/id/lsoa/E01003953",
      "_about": "http://education.data.gov.uk/id/school/100869",
      "establishmentNumber": 6385,
      "LLSC": "http://statistics.data.gov.uk/id/llsc/GL140",
      "gender": "http://education.data.gov.uk/def/school/Gender_Mixed",
      "parliamentaryConstituency": "http://statistics.data.gov.uk/id/parliamentary-constituency/142",
      "urbanRural": "http://education.data.gov.uk/def/school/UrbanRural_Urban_10k_less_sparse",
      "districtAdministrative": "http://statistics.data.gov.uk/id/local-authority-district/00BE",
      "typeOfEstablishment": "http://education.data.gov.uk/def/school/TypeOfEstablishment_TERM_Other_Independent_School",
      "religiousCharacter": "http://education.data.gov.uk/def/school/ReligiousCharacter_None",
      "localAuthority": "http://statistics.data.gov.uk/id/local-education-authority/210",
      "uniqueReferenceNumber": 100869,
      "long": -0.078969999999999999,
      "easting": 533552,
      "address": {                              location of the school
        "town": "London",
        "address1": "14 Derwent Grove",
        "address2": "East Dulwich",
        "postcode": "SE22 8EA"
      },
      "lat": 51.460700000000003,
      "label": "Southwark Small School",
      "MSOA": "http://statistics.data.gov.uk/id/msoa/E02000834",
      "censusAreaStatisticWard": "http://statistics.data.gov.uk/id/cas-ward/00BEGH",
      "northing": 175275
},
```


Fig. 20.11
Location view with the Linked Data API in JSON format, e-government

20.3.2.8 PSI Linked Data Mashups and Applications

It is now possible to see Semantic Web mashups and applications where some of the practices described above have been applied, and where the datasets also played an important role, either by use of their linked data resources or via querying the SPARQL end points.

Most of the PSI linked data applications show how to integrate several linked data sources using a map as key part of their functionalities. The EnAKTing project studied how much Linked Data can be retrieved using UK postcodes as input [94], in an application integrating Parliamentary data, crime, hospital waiting times, and mortality rates. This use-case study brought up numerous issues that a Web or Semantic Web developer would need to face when developing this type of application. Other applications like "How Good is my area?" (http://myarea.psi.enakting.org/) or schools.openpsi.org have shown mashups around the Indexes of Multiple Deprivation dataset. The former ranks an area based on multiple rankings on a series of parameters and compares it with its surroundings. The latter is an interesting study that visually integrates school performances with the different sociocultural and economic aspects for a given location (see *Fig. 20.12*).

As part of the EnAKTing research, similar applications have been developed that consume linked data sources providing visualizations of integrated datasets. The "UK CO2 Emissions Visualization" application (*Fig. 20.13*) is a good example of such work and shows an application made with the integration of three Linked Data sources: Geonames, Ordnance Survey Administrative Geography, and CO_2 Emissions.

In the USA, the data.gov project has declared a clear commitment to Semantic Web technologies in part thanks to the promotion of these technologies from the Tetherless

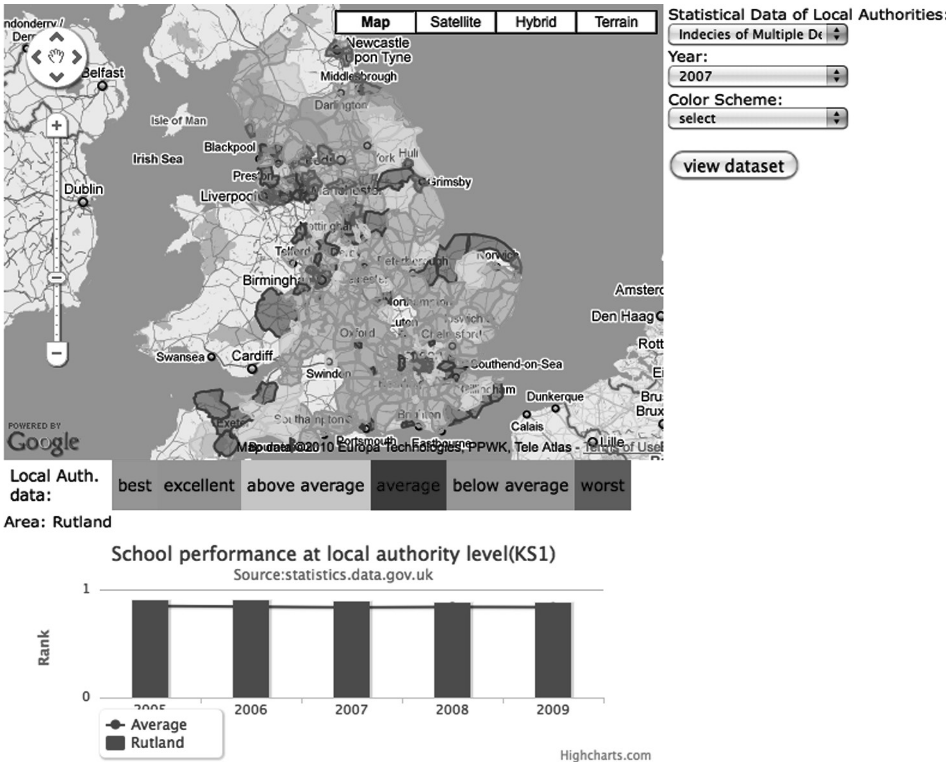


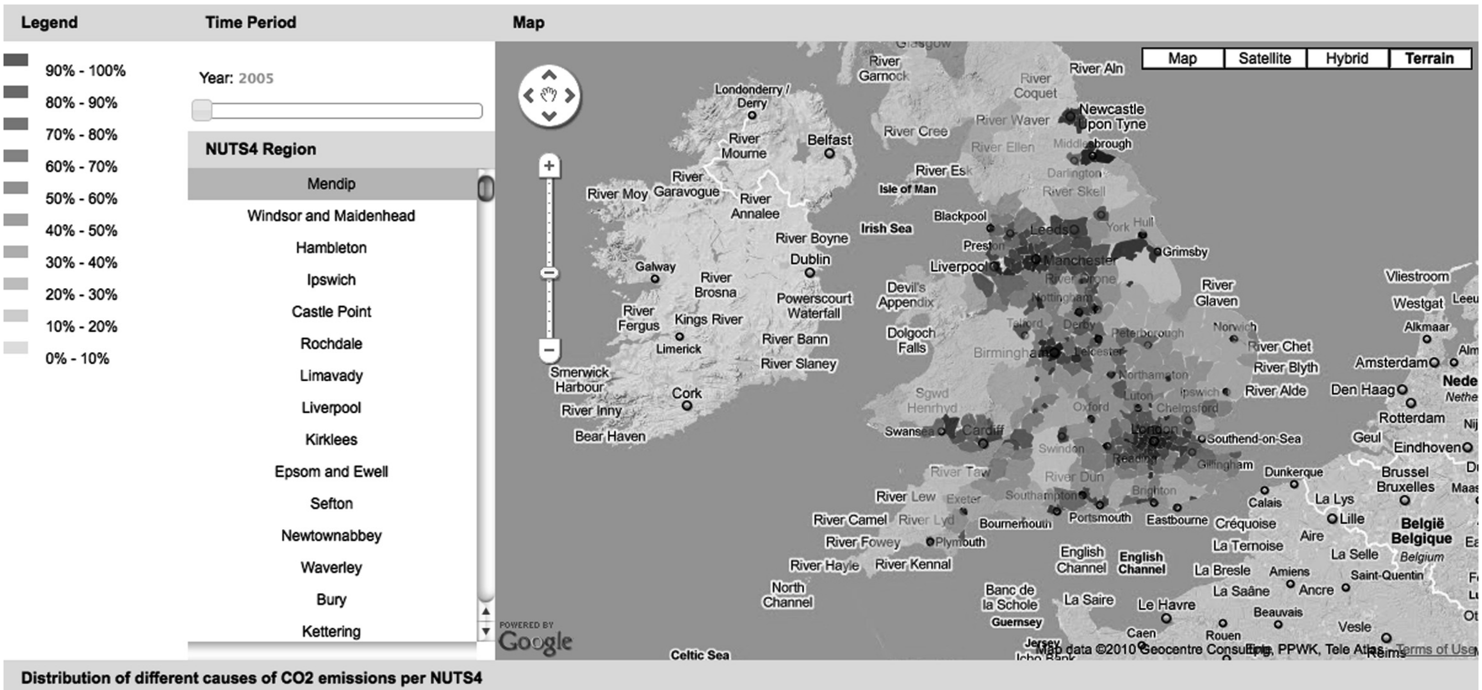
Fig. 20.12
OpenPSI school application, e-government

World Constellation at Rensselaer Polytechnic Institute (http://rpi.edu/research/constellations/tetherlessworld.html). In data.gov one can see the important role of these technologies with interesting mashups that show the use of linked datasets in a geographical context (see *Fig. 20.14* – http://www.data.gov/semantic/).

20.4 Related Resources

Important resources for the student of e-government and the SW include:

- http://www.recovery.gov/ is the address of Obama's Website to monitor the stimulus. The US site that releases public-sector data is http://www.data.gov/
- The UK equivalent is http://data.gov.uk/
- The UK government's Public Sector Transparency Board can be found at http://writetoreply.org/publicsectortransparencyboard/. The UK transparency program is described in a letter from the UK Prime Minister David Cameron to Cabinet



■ Fig. 20.13
EnAKTinG UK CO_2 emissions linked data app., e-government

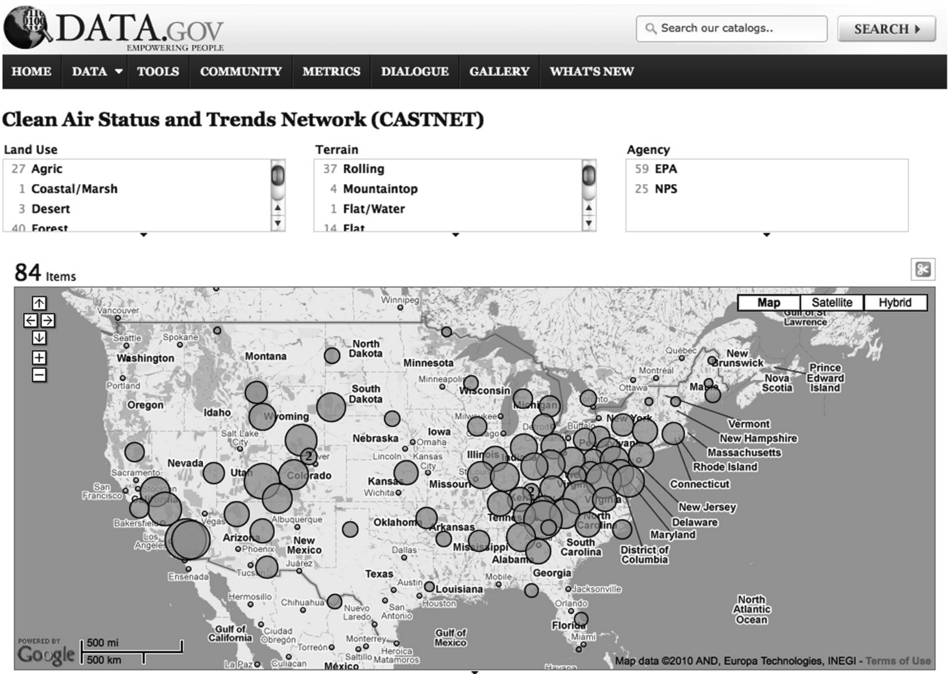


Fig. 20.14
Clean Air Status and Trends Network (CASTNET) application from data.gov, eGovernment

Ministers in 2010, available at http://www.cabinetoffice.gov.uk/newsroom/statements/transparency/pm-letter.aspx. A position paper written by OPSI reflecting its information policy and the influence of AKTive PSI is at http://www.w3.org/2007/06/eGov-dc/papers/opsi-position-paper

- The EnAKTinG project can be found at http://www.enakting.org/
- SEMIC.EU is at http://www.semic.eu/semic/view/index.xhtml
- For OntoGov, see http://www.hsw.fhso.ch/ontogov/
- For DIP, see http://dip.semanticweb.org/
- The WSMO working group is at http://www.wsmo.org/
- Access-eGov is at http://www.accessegov.org/acegov/web/uk/index.jsp
- The Tetherless World Constellation is at http://rpi.edu/research/constellations/tetherlessworld.html

A very useful volume containing a series of essays describing EU projects, and the general European approach of reengineering is Tomas Vitvar, Vassilios Peristeras, and Konstantinos Tarabanis (eds.), *Semantic Technologies for E-Government*, Berlin: Springer, 2010. It contains papers by most leading Continental European researchers from a range of projects, and broadly covers architectures and process integration, ontologies and interoperability, and portals and user interaction.

20.5 Future Issues

Infrastructure development is an important issue. For the e-government domain, the trust layer is all-important, and at the current rate of the SW's progress that is a matter for research. The description of privacy policies, and discovery of trustworthy services and resources is essential for the future take-up of SW technologies in this space. Nevertheless, it is a common assumption that the SW's reasoning capabilities should allow assessment of indicators of trustworthiness, and of the requirements (e.g., for privacy) of clients.

Other aspects of infrastructure are also lacking. Some of the most promising approaches for supplying services are based around WSMO, but that is still something of a work in progress. However, it has been argued (e.g., [12]) that WSMO has a great deal of potential for supporting e-government services.

Full integration of SW technologies and services with the complex environment of e-government remains a difficult issue and Gugliotta et al. [12] argue that a complex semantic layer providing a framework explicitly for e-government is required. On the other hand, such a complex layer risks adding to the complexity which e-government practitioners already perceive. Standards for semantic technologies are important, in order to remove risk from the development process, as well as improving trust in the services provided. The perennial issue of the markup of legacy data, of which of course there is an enormous amount in most government data stores, is also extremely important in this area.

Trust and privacy have been neglected in this field, at least partly because of the requirement to create the functionality able to deal with a complex space. Research in the Policy Aware Web is ongoing, transferring the imperative from curtailing data access to the more tractable one of ensuring fair use of data, via transparency of use and accountability for action (enforcement of privacy policies). Privacy sits alongside, and often in tension with, freedom of information and the need for data protection – for data protection to be properly implemented, a strong recommendation is that users are able to interact with data, by being able to check it for accuracy, amend where necessary, and apply their own privacy policies where this is admissible. Developing systems for presenting data to users, and allowing flexible semantic search, is very important. Once data is out there, other social processes can be brought in to help semanticization – for example, Web 2.0 style tagging, which would allow the harnessing of emergent semantics [9].

Structuring and release of information is a vital early step for increasing data access. AKTive PSI has shown how this does not need heavy-duty ontologies or major and immediate buy-in from a large number of units. Ambition should be reserved for the long run, not the short. Quick wins will be gained from providing the means to link data, rather than applying the full panoply of SW technologies.

But culture change will also be required. Data are often unavailable or hard to process for a number of reasons. First, it can be technically difficult or impossible to access, represented as proprietary databases, inaccessible spreadsheets, or embedded in semi- and unstructured Web Pages. Indexes and directories of content are often poor. There is little decentralization and single database models predominate that are not always maintained. Second, organizationally and socially there is little incentive for departments or individuals in them to publish data. Fear of inappropriate release, the lack of standards, the invisibility of the benefits and fear of disruption to existing legacy systems dominate thinking. Privacy is a particular concern of the cautious here (and sometimes is an excuse for inertia). Third, licensing and regulatory regimes can add complexity – for instance, the terms of a license might be irrelevant to many aspects of potential reuse.

The lines of a general SW approach to the representation of government data, and making it available either across government departments, or to nongovernmental users (especially citizens) are becoming clear. In more detail, the following are useful and practical steps to ensure that technical standards are met and that institutional culture does not impede progress.

1. Access to data should be supplied using an HTTP browser and SPARQL end points to a single point of online access for public datasets.
2. There should be a standard publication format using agreed W3C Linked Data Standards – ideally using standards adopted by other countries.
3. Lightweight integrative ontologies should be used, and some effort put into selecting and developing common terms where necessary.
4. Copyright and commons standards should be developed that are easy to understand and implement.
5. There should be support for community-based indexing, categorization, and annotation of data sets.
6. There should be support for the exploitation and publication of distributed and decentralized information assets by a wide variety of users.
7. There should be attention to make the above part of departments' routine operations.
8. The information regulatory regime should be adjusted to support the proactive publication of government information.
9. Governments should work to promote international liaison and global standards setting, investing in future international data sharing.
10. Governments should adopt an assumption of total publication for anonymous data using open standards.

These principles currently remain a wish list, with variable application across the range of countries that promote e-government. What is encouraging is the increasing signs that countries are looking to follow the examples of the USA and UK.

Much of the success of these endeavors will ultimately depend on political will and leadership. The UK has been fortunate to have successive governments promote and support Government Open Data. Following the formation of the UK Coalition Government in May 2010 Berners-Lee and Shadbolt were asked to join the Public Sector Transparency Board (http://writetoreply.org/publicsectortransparencyboard/) whose terms of reference seek to extend and consolidate public rights to data, transparency, and accountability through data release, setting open data standards across the whole public sector and the ongoing development of data.gov.uk. An early output is the set of Public Data Principles (http://data.gov.uk/blog/new-public-sector-transparency-board-and-public-data-transparency-principles), one of which states:

Release data quickly, and then re-publish it in linked data form – Linked data standards allow the most powerful and easiest re-use of data. However most existing internal public sector data is not in linked data form. Rather than delay any release of the data, our recommendation is to release it 'as is' as soon as possible, and then work to convert it to a better format.

These principles are intended to change attitudes and behavior – in this way one can hope to drive culture change in the UK government administration toward an assumption of total publication for anonymous data using open standards. Government information out in the open will drive innovation, as more people are able to interrogate the data for their own purposes. Semantic technologies provide great promise, but pragmatic considerations loom large. It may be that open standards, freely available data, small ontologies, quick wins, and modest models are the way to drive the use of SW in e-government, rather than implementing large IT systems entirely from the top down. As has been argued, a judicious mix of top-down and bottom-up strategies is required.

It has been observed that SW technologies are playing a key role in the evolution of eGovernment. In particular a world of linked open government data offers not just a win for the technology. As the UK's *Guardian* newspaper wrote in a leading editorial:

- It is . . . hazardous trying to envision how freer data will redraw the boundaries between different communities or recast their relationship with power. But it is reasonable to speculate that the uncovering and unlocking of so much information will drive improvements in public policy. It will level the territory on which voters meet politicians, and could prove a powerful brake on campaigning hyperbole in the coming election. Without the printed word there would have been no informed electorate, no demand for accountability from our leaders – and indeed no democracy at all. Open data will surely revive it, and in time could transform it too [95].

Acknowledgments

This work was supported in part by the EPSRC project EnAKTinG (http://www.enakting.org/about.html), grant no. EP/G008493/1, LiveMemories – Active Digital Memories of Collective Life, Bando Grandi Progetti 2006, Provincia Autonoma di Trento, the EU FET project Living Knowledge (http://livingknowledge-project.eu/), contract no. 231126, and the EU project WeGov (http://wegov-project.eu/), contract no. 248512. Our thanks to an anonymous reviewer for extremely helpful comments.

References

1. Accenture: eGovernment leadership: High performance, maximum value. http://www.accenture.com/NR/rdonlyres/D7206199-C3D4-4CB4-A7D8-846C94287890/0/gove_egov_value.pdf (2004). Accessed 8 Oct 2010
2. U.K. Cabinet Office: Transformational government: enabled by technology. http://www.cabinetoffice.gov.uk/media/141734/transgov-strategy.pdf (2005). Accessed 8 Oct 2010

3. U.S. Office of Management and Budget: Expanding e-government: improved service delivery for the American people using information technology. http://www.whitehouse.gov/omb/budintegration/expanding_egov_2005.pdf (2005)
4. Peristeras, V., Mentzas, G., Tarabanis, K.A., Abecker, A.: Transforming e-government and e-participation through IT. IEEE Intell. Syst. **24**(5), 14–19 (2009)
5. Layne, K., Lee, J.: Developing fully functional e-government: A four stage model. Gov. Inform. Q. **18**, 122–136 (2001)
6. Berners-Lee, T., Shadbolt, N.: Our manifesto for government data. Guardian DataBlog. http://www.guardian.co.uk/news/datablog/2010/jan/21/timbernerslee-government-data (2010). Accessed 8 Oct 2010
7. Shadbolt, N.: Letter from Professor Nigel Shadbolt to Rt Hon Francis Maude MP, Minister for the Cabinet Office. http://data.gov.uk/sites/default/files/Transparency%20Board%20-%20letter%20from%20Nigel%20Shadbolt%20to%20MCO%2014.06.10.pdp__0.pdf (2010). Accessed on 14 June 2010
8. Orthofer, G., Wimmer, M.A.: An ontology for eGovernment: Linking the scientific model with concrete projects. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web Meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 96–98 (2006)
9. Ilgar, E., Oldenhuizing, J., Mika, P.: Let the citizen speak: a demand-driven e-government portal using Semantic Web technology. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web Meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 54–55 (2006)
10. Klischewski, R.: Semantic Web for e-government. In: Traunmüller, R. (ed.) EGOV2003, pp. 288–295. Springer, Berlin (2003)
11. Wagner, C., Cheung, K.S.K., Ip, R.K.F., Böttcher, S.: Building Semantic Webs for e-government with wiki technology. Electr. Gov. **3**, 36–55 (2006)
12. Gugliotta, A., Tanasescu, V., Domingue, J., Davies, R., Gutiérrez-Villarías, L., Rowlatt, M., Richardson, M., Stinčic, S.: Benefits and challenges of applying Semantic Web services in the e-government domain. Semantics 2006, Vienna, http://projects.kmi.open.ac.uk/dip/resources/Semantics2006/Semantics2006_DIP_Camera_Ready.pdf (2006)
13. Walden, I.: Privacy and data protection. In: Reed, C., Angel, J. (eds.) Computer Law: The Law and Regulation of Information Technology, pp. 459–504. Oxford University Press, Oxford (2007)
14. Alani, H., Kalfoglou, Y., O'Hara, K., Shadbolt, N.: Towards a killer app for the Semantic Web. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) The Semantic Web – ISWC 2005, pp. 829–843. Springer, Berlin (2005)
15. Berners-Lee, T., Hendler, J.A., Lassila, O.: The Semantic Web. Sci. Am. **284**(5), 34–43 (2001)
16. Shadbolt, N., Berners-Lee, T., Hall, W.: The Semantic Web revisited. IEEE Intell. Syst. **21**(3), 96–101 (May–Jun 2006)
17. Berčič, B., Vintar, M.: Ontologies, Web services and intelligent agents: Ideas for further development of life-event portals. In: Traunmüller, R. (ed.) EGOV2003, pp. 329–334. Springer, Berlin (2003)
18. Corradini, F., Sabucedo, L.Á., Polzonetti, A., Rifón, L.A., Re, B.: A case study of semantic solutions for citizen-centered Web portals in eGovernment: The Tecut portal. In: Wimmer, M.A., Scholl, H.J., Grönlund, A. (eds.) EGOV2007, pp. 204–215. Springer, Berlin (2007)
19. Furdik, K., Klischewski, R., Paralič, M., Sabol, T., Skokan, M.: e-Government service integration and provision using semantic technologies. In: Scholl, H. J., Janssen, M., Traunmueller, R., Wimmer, M. A. (eds.) Electronic Government. Proceedings of Ongoing Research, General Development Issues and Projects of EGOV 09, 8th International Conference, pp. 273–280, 2009, Linz. http://web.tuke.sk/fei-cit/furdik/publik/egov09_aeg.pdf
20. Furdik, K., Sabol, T., Bednar, P.: Framework for integration of eGovernment services on a semantic basis. Presented at EGOV2007. http://www.accessegov.org/acegov/uploadedFiles/webfiles/cffile_10_29_07_9_50_10_AM.pdf (2007)
21. Skokan, M., Bednár, P.: Semantic orchestration of services in e-government. Znalosti 2008, Bratislava, pp. 215–223. http://znalosti2008.fiit.stuba.sk/download/articles/znalosti2008-Bednar.pdf (2008)
22. Maltese, V., Giunchiglia, F., Denecke, K., Lewis, P., Wallner, C., Baldry, A., Madalli, D.: On the interdisciplinary foundations of diversity. Dipartimento di Ingegneria e Scienza dell'Informazione, 2009, University of Trento, technical report DISI-09-040.

http://eprints.biblio.unitn.it/archive/00001651/01/040.pdf
23. Klischewski, R.: Top down or bottom up? How to establish a common ground for semantic interoperability within e-government communities. In: Traunmüller, R., Palmirani, M. (eds.) E-government: Modelling norms and concepts as key issues. 1st International Workshop on e-Government at ICAIL 2003, 2003
24. Klischewski, R.: Information integration or process integration? How to achieve interoperability in administration. In: Traunmüller, R. (ed.) EGOV2004, pp. 57–65. Springer, Berlin (2004)
25. Guijarro, L.: Semantic interoperability in eGovernment initiatives. Comp. Stand. Inter. **31**, 174–180 (2009)
26. Mach, M., Sabol, T., Paralic, J.: Integration of eGov services: back-office versus front-office integration. In: Abecker, A., Mentzas, G., Stojanovic, L. (eds.) Proceedings of the Workshop on Semantic Web for eGovernment 2006, 3rd European Semantic Web Conference, pp. 48–53, 2006, Serbia & Montenegro. http://www.imu.iccs.gr/semgov/final/SemanticWebForEGovernemnt-Proceeding.pdf
27. Weinstein, P.: Model unification in support of political process. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web Meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 138–140
28. Della Valle, E., Cerizza, D., Celino, I., Estublier, J., Vega, G., Kerrigan, M., Ramírez, J., Villazon, B., Guarrera, P., Zhao, G., Monteleone, G.: SEEMP: A semantic interoperability infrastructure for e-government services in the employment sector. In: Franconi, E., Kifer, M., May, W. (eds.) The Semantic Web: Research and Applications: Proceedings of the 4th European Semantic Web Conference, pp. 220–234, 2007, Springer, Berlin
29. Herborn, T., Wimmer, M.A.: Process ontologies facilitating interoperability in e-government: A methodological framework. In: Hinkelmann, K., Karagiannis, D., Stojanovic, N., Wagner, G. (eds.) Proceedings of the Workshop on Semantics for Business Process Management, pp. 76–88, 2006
30. Van Elst, L., Klein, B., Maus, H., Schöning, H., Tommasi, A., Zavattari, C., Favaro, J., Giannella, V.: Business Register Interoperability Throughout Europe: BRITE. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web Meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 70–79
31. Fraser, J., Adams, N., Macintosh, A., McKay-Hubbard, A., Lobo, T.P., Pardo, P.F., Martínez, R.C., Vallecillo, J.S.: Knowledge management applied to e-government services: the use of an ontology. In: Wimmer, M.A. (ed.) Knowledge Management in Electronic Government: Proceedings of 4th IFIP International Working Conference, KMGov 2003, pp. 116–126, 2003, Springer, Berlin
32. Barnickel, N., Fluegge, M., Schmidt, K.-U.: Interoperability in eGovernment through cross-ontology Semantic Web service composition. In: Abecker, A., Mentzas, G., Stojanovic, L. (eds.) Proceedings of the Workshop on Semantic Web for eGovernment: 3rd European Semantic Web Conference, pp. 37–47, 2006, Montenegro
33. Maedche, A., Motik, B., Silva, N., Volz, R.: MAFRA – A MApping FRAmework for distributed ontologies. In: Gómez-Pérez, A., Benjamins, V.R. (eds.) Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web: 13th International Conference EKAW 2002, pp. 235–250, 2002, Springer, Berlin
34. Czajkowski, M., Ashpole, B., Hughes, T., Le, T.: Bridging semantic eGovernment applications using ontology-to-ontology message translation. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web Meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 4–6
35. Salvadores, M., Correndo, G., Szomszor, M., Yang, Y., Gibbins, N., Millard, I., Glaser, H., Shadbolt, N.: Domain-specific backlinking services in the Web of Data. Web Intelligence 2010, Toronto. http://eprints.ecs.soton.ac.uk/21258/ (2010)
36. Breslin, J.G., Harth, A., Bojars, U., Decker, S.: Towards semantically-interlinked online communities. In: Gómez-Pérez, A., Euzenat, J. (eds.) The Semantic Web: Research and Applications: 2nd European Semantic Web Conference ESWC 2005, pp. 500–514, 2005, Springer, Berlin
37. Tullo, C.: Information age. Civil service network. http://www.civilservicenetwork.com/features/features-article/newsarticle/information-age/ (2008). Accessed on 24 Oct 2008
38. Pullinger, D.: Building the public sector information infrastructure. Talk presented at Conference on Unlocking the Potential of Public Sector Information, London. http://www.epsiplus.net/

media/files/3_14october2008_david_pullinger (2008)
39. Alani, H., Dupplaw, D., Sheridan, J., O'Hara, K., Darlington, J., Shadbolt, N., Tullo, C.: Unlocking the potential of public sector information with Semantic Web technology. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) The Semantic Web: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC2007 + ASWC2007, pp. 708–721, 2007, Springer, Berlin
40. U.K. Office of Public Sector Information: The United Kingdom implementation of the European directive on the re-use of public sector information – the first two years. http://www.opsi.gov.uk/advice/psi-regulations/uk-implementation-first-years.pdf (2007)
41. Comte, F., Leclère, M.: A semantical reasoning framework for eGovernment of the French social welfare system. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web Meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 7–9
42. Sidoroff, T., Hyvönen, E.: Semantic e-government portals – A case study. In: Proceedings of SWCASE05: Semantic Web Case Studies and Best Practice for eBusiness, International Semantic Web Conference (ISWC), 2005, Galway, Ireland. http://www.seco.tkk.fi/publications/2005/sidoroff-hyvonen-semantic-e-government-2005.pdf
43. Peristeras, V., Goudos, S.K., Loutas, N., Tarabanis, K.: Ontology-based search for eGovernment services using citizen profile information. J. Web Eng. **8**(3), 245–267 (2009)
44. Markellos, K., Sakkopoulos, E., Sirmakessis, S., Tsakalidis, A.: Semantic Web search for e-government: the case study of Intrastat. J. Internet Technol. **8**, 1–12 (2007)
45. Apostolou, D., Stojanovic, N., Anicic, D.: Responsive knowledge management for public administration: an event-driven approach. IEEE Intell. Syst. **24**(5), 20–30 (2009)
46. Stojanovic, N., Mentzas, G., Apostolou, D.: Semantic-enabled agile knowledge-based e-government. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web Meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 132–134
47. Sacco, G.M.: User-centric access to e-government information: e-citizen discovery of e-services. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web Meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 114–116
48. Medjahed, B., Bouguettaya, A.: Customized delivery of e-government Web services. IEEE Intell. Syst. **20**(6), 77–84 (Nov–Dec 2005)
49. Bouguettaya, A., Medjahed, B., Rezgui, A., Ouzzani, M., Liu, X., Yu, Q.: WebDG – A platform for e-government Web services. In: Wang, S., Tanaka, K., Zhou, S., Ling, T.W., Guan, J., Yang, D., Grandi, F., Mangina, E., Song, I.-Y., Mayr, H.C. (eds.) Conceptual Modeling for Advanced application domains: ER2004 Workshops CoMoGIS, CoMWIM, EDCM, CoMoA, DGOV and eCOMO, pp. 553–565, 2004, Springer, Berlin
50. Chun, S.A., Atluri, V., Adam, N.R.: Policy-based Web service composition. In: Proceedings of the 14th International Workshop on Research Issues in Data Engineering: Web Services for e-Commerce and e-Government Applications, pp. 85–92, 2004, Boston
51. Medjahed, B., Bouguettaya, A., Elmagarmid, A.K.: Composing Web services on the Semantic Web. VLDB J. **12**, 333–351 (2003)
52. Medjahed, B., Rezgui, A., Bouguettaya, A., Ouzzani, M.: Infrastructure for e-government Web services. IEEE Internet Computing. **7**(1), 58–65 (Jan–Feb 2003)
53. Cabral, L., Domingue, J., Galizia, S., Gugliotta, A., Tanasescu, V., Pedrinaci, C., Norton, B.: IRS-III: A broker for Semantic Web services based applications. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, M. (eds.) The Semantic Web – ISWC 2006, pp. 201–214. Springer, Berlin (2006)
54. Gugliotta, A., Domingue, J., Cabral, L., Tanasescu, V., Galizia, S., Davies, R., Gutierrez Villarias, L., Rowlatt, M., Richardson, M., Stincic, S.: Deploying Semantic Web services-based applications in the e-government domain. J. Data Semant. **10**, 96–132 (2008)
55. Gugliotta, A., Cabral, L., Domingue, J., Roberto, V., Rowlatt, M., Davies, R.: A Semantic Web service-based architecture for the interoperability of e-government services. In: Proceedings of the Web Information Systems Modeling Workshop

(WISM 2005), 5th International Conference on Web Engineering, 2005, Sydney. http://oro.open.ac.uk/3005/1/SemanticWeb.pdf
56. Wang, X., Vitvar, T., Peristeras, V., Mocan, A., Goudos, S.K., Tarabanis, K.: WSMO-PA: Formal specification of public administration service model on Semantic Web service ontology. In: Proceedings of the 40th Annual Hawaii International Conference on System Science (HICSS07), 2007. http://portal.acm.org/citation.cfm?id = 1255783
57. Peristeras, V., Tarabanis, K.: Providing pan-European e-government services with the use of Semantic Web services technologies: a generic process model. In: Wimmer, M.A., Traunmüller, R., Grönlund, Å., Andersen, K.V. (eds.) Electronic Government: 4th International Conference EGOV 2005, pp. 226–236, 2005, Springer, Berlin
58. Peristeras, V., Tarabanis, K.: Reengineering public administration through semantic technologies and a reference domain ontology. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 99–110
59. O'Hara, K., Shadbolt, N.: The Spy in the Coffee Machine: The End of Privacy As We Know It. Oneworld, Oxford (2008)
60. Klischewski, R., Jeenicke, M.: Semantic Web technologies for information management within e-government services. In: Proceedings of the 37th Annual Hawaii International Conference on System Science (HICSS04), 2004. http://www2.computer.org/portal/web/csdl/doi/10.1109/HICSS.2004.1265305
61. Weitzner, D.J., Abelson, H., Berners-Lee, T., Hanson, C., Hendler, J., Kagal, L., McGuinness, D.L., Sussman, G.J., Waterman, K.K.: Transparent accountable data mining: New strategies for privacy protection. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web Meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 141–150
62. Stojanovic, L., Abecker, A., Stojanovic, N., Studer, R.: On managing changes in the ontology-based e-government. In: Meersman, R., Tari, Z., van der Aalst, W., Bussler, C., Gal, A., Cahill, V., Vinoski, S., Vogels, W., Catarci, C., Sycara, K. (eds.) On the Move to Meaningful Internet Systems 2004: CooPIS, DOA, ODBASE, vol. 2, pp. 1080–1097. Springer, Berlin (2004)
63. Klischewski, R., Ukena, S.: Designing semantic e-government services driven by user requirements. Presented at EGOV2007. http://www.accessegov.org/acegov/uploadedFiles/webfiles/cffile_10_29_07_9_46_36_AM.pdf (2007)
64. Stojanovic, L., Abecker, A., Apostolou, D., Mentzas, G., Studer, R.: The role of semantics in e-government service model verification and evolution. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web Meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, 117–128
65. Stojanovic, L., Stojanovic, N., Apostolou, D.: Change management in e-government: OntoGov case study. Electr. Gov. **3**, 74–92 (2006)
66. Hinkelmann, K., Probst, F., Thönssen, B.: Agile process management: Framework and methodology. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web meets e-government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 33–41
67. Stojanovic, N., Stojanovic, L., Hinkelmann, K., Mentzas, G., Abecker, A.: Fostering self-adaptive e-government service improvement using semantic technologies. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web meets e-government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 129–131
68. Apostolou, D., Stojanovic, L., Lobo, T.P., Thoenssen, B.: Towards a semantically-driven software engineering environment for eGovernment. In: Böhlen, M., Gamper, J., Polasek, W., Wimmer, M.A. (eds.) e-Government: Towards Electronic Democracy: Proceedings of TCGOV2005, pp. 157–168, 2005, Springer, Berlin
69. Tambouris, E., Gorilas, S., Kavadias, G., Apostolou, D., Abecker, A., Stojanovic, L., Mentzas, G.: Ontology-enabled e-gov service configuration: An overview of the OntoGov project. In: Wimmer, M.A. (ed.) Knowledge management in Electronic Government: Proceedings of 5th IFIP International Working Conference KMGov2004, pp. 122–127, 2004, Springer, Berlin
70. Fishenden, J., Bell, O., Grose, A.: Government interoperability: Enabling the delivery of e-services. Microsoft white paper (2005)

71. Koné, M.T., Jaafar, F.B., Saïd, A.M.: A critical step in eGovernment evolution. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web Meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 64–69
72. Bennett, D., Harvey, A.: Publishing open government data. W3C working draft. http://www.w3.org/TR/gov-data/ (Sept 2009). Accessed 8 Oct 2010
73. Klischewski, R.: Migrating small governments' websites to the Semantic Web. In: Abecker, A., Sheth, A., Mentzas, G., Stojanovic, L. (eds.) Semantic Web Meets e-Government: AAAI Spring Symposium 2006, technical report SS-06-06, pp. 56–63
74. Alani, H., Hall, W., O'Hara, K., Shadbolt, N., Chandler, P., Szomszor, M.: Building a pragmatic Semantic Web. IEEE Intell. Syst. **23**(3), 61–68 (May–June 2008)
75. Berners-Lee, T.: The Fractal Nature of the Web. http://www.w3.org/DesignIssues/Fractal.html (1998/2006). Accessed 8 Oct 2010
76. U.K. Office of Public Sector Information: Unlocking PSI potential: The United Kingdom report on the re-use of public sector information. http://www.opsi.gov.uk/advice/psi-regulations/uk-report-reuse-psi-2008.pdf (2008)
77. Bizer, C., Cyganiak, R., Heath, T.: How to publish linked data on the Web. http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/ (2007)
78. Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., Sheets, D.: Tabulator: Exploring and analyzing linked data on the Semantic Web. In: 3rd International Semantic Web User Interaction Workshop (SWUI06), 2006, Athens. http://swui.semanticweb.org/swui06/papers/Berners-Lee/Berners-Lee.pdf
79. Berners-Lee, T.: Linked data. http://www.w3.org/DesignIssues/LinkedData.html (2007–2009). Accessed 8 Oct 2010
80. Kalfoglou, Y., Schorlemmer, M.: Ontology mapping: the state of the art. Knowl. Eng. Rev. **18**(1), 1–31 (2003)
81. Bizer, C., Heath, T., Berners-Lee, T.: Linked data – The story so far. In: Heath, T., Hepp, M., Bizer, C. (eds.): International Journal on Semantic Web and Information Systems: Special Issue on Linked Data, 5(3) 1–22, (2009). http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf
82. Acar, S., Alonso, J.M., Novak, K.: Improving access to government through better use of the Web. http://www.w3.org/TR/egov-improving/ (May 2009). Accessed 8 Oct 2010
83. Crabtree, J.: Whitehall's Web revolution: the inside story. Prospect, 166. http://www.prospectmagazine.co.uk/2010/01/whitehalls-web-revolution-the-inside-story/ (Jan 2010)
84. T, J.: Calling open data developers – we need your help. Cabinet Office Digital Engagement Blog. http://blogs.cabinetoffice.gov.uk/digitalengagement/post/2009/09/30/Calling-Open-Data-Developers-We-need-your-help.aspx (Sept 2009)
85. Arthur, C.: UK's free data website 'is a world showcase'. Guardian DataBlog. http://www.guardian.co.uk/technology/datablog/2010/jan/21/government-free-data-website-launch (21 Jan 2010). Accessed 8 Oct 2010
86. Berners-Lee, T.: Tim Berners-Lee on the next Web, TED talk. http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html (2009)
87. Sheridan J., Tennison J.: Linking UK Government Data. Linked Data on the Web (LDOW2010), WWW2010 Workshop. http://events.linkeddata.org/ldow2010/slides/ldow2010-slides-tennison.pdf (Apr 2010)
88. Chief Technology Officer Council: Designing URI sets for the UK public sector. http://www.cabinetoffice.gov.uk/media/301253/puiblic_sector_uri.pdf (Oct 2009)
89. Sauermann L., Cyganiak R.: Cool URIs for the Semantic Web, W3C Interest Group Note 03. http://www.w3.org/TR/cooluris/ (Dec 2008)
90. Carroll, J.J., Bizer, C., Hayes, P.J., Stickler, P.: Named graphs, provenance and trust. In: Proceedings of the World Wide Web Conference 2005, pp. 613–622, 2005, New York
91. Baker, T.: Maintaining Dublin Core as a Semantic Web vocabulary. From Integrated Publication and Information Systems to Virtual Information and Knowledge Environments (2005), 61–68
92. Moreau, L.: The foundations for provenance on the Web. Found.Trends Web Sci. **2**(2) (2010)
93. Cyganiak, R., Field, S., Gregory, A., Halb, W., Tennison, J.: Semantic statistics: bringing together SDMX and SCOVO. Linked Data on the Web (LDOW2010), WWW2010 Workshop (Apr 2010)

94. Omitola, T., Koumenides, C. L., Popov, I. O., Yang, Y., Salvadores, M., Szomszor, M., Berners-Lee, T., Gibbins, N., Hall, W., schraefel, M., Shadbolt, N. (2010) Put in your postcode, out comes the data: a case study. In: 7th Extended Semantic Web Conference, May 2010, Greece

95. The Guardian: Government information: Creative Commons. Guardian newspaper leader column. http://www.guardian.co.uk/commentisfree/2010/jan/23/government-information-creative-commons-internet (23 Jan 2010)