

Chapter 1

Introduction to Linked Data



Abstract This chapter presents Linked Data, a new form of distributed data on the web which is especially suitable to be manipulated by machines and to share knowledge. By adopting the linked data publication paradigm, anybody can publish data on the web, relate it to data resources published by others and run artificial intelligence algorithms in a smooth manner. Open linked data resources may democratize the future access to knowledge by the mass of internet users, either directly or mediated through algorithms. Governments have enthusiastically adopted these ideas, which is in harmony with the broader open data movement.

Keywords Linked data · Semantic web · Democracy · Ontologies · Knowledge representation · eDemocracy

1.1 Introduction

More than half of the world's population has access to the Internet. Vast amounts of knowledge accumulated in roughly 2 billion websites are available to anyone who is able to read and can afford an internet connection.

Entertainment habits, interpersonal human relations and almost any conceivable aspect of human life have been profoundly transformed with the arrival of the internet. Yet modern democracies have remained relatively unaffected. It is true that propaganda techniques have undergone changes, political parties organize their campaign strategies differently and the idea of eDemocracy is perhaps about to hatch; but the public institutions, the habits of citizens and the overall political game are all apparently the same.

We have to indulge—Internet is a new thing. But a careful observation of the evolution of technologies and the new organizational forms they enable reveal discrete signs of change, now with little effect but potentially of much impact.

This chapter introduces some new technologies and ideas which may seem irrelevant today, but which will probably exert a powerful influence on the forthcoming transformations of the concept of democracy.

1.2 The World Wide Web as a Source of Data and Knowledge

1.2.1 *Data, Information and Knowledge*

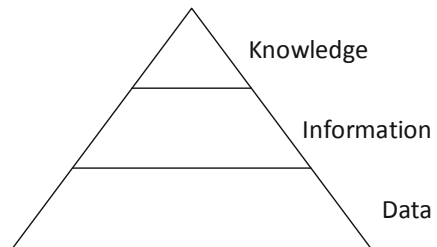
Marshall McLuhan described technology as extensions of man (McLuhan 1964), whereby our bodies and our senses are extended beyond their natural limits. Certainly, a shovel is an improvement of our hands when we dig a trench and telescopes are augmented eyes when we look at the stars. In top level chess tournaments, chess players prepare their games and study their opponents with a joint team of humans and machine—machines also extend human’s capabilities for thinking.

In order to make a value judgement, we need data—this is a truism. But today we also need machines which need data. Whenever we take an important decision, we usually google for some related information. Our decisions are mediated by information provided by a company, or a handful of companies, whose interests may not match our interests. Maybe in the future we will have a wider range of algorithms to apply to a common pool of open knowledge—both data and algorithms are essential extensions of our mind enhancing and rational processes.

This book is about linked democracy, a concept of democracy where knowledge plays a central role; and this relation between data, algorithms and knowledge has to be studied in more detail. One of the possible conceptual frameworks is the popular pyramid of data, information and knowledge, represented in Fig. 1.1 in a manner that suggests that data is abundant, information not so much and knowledge is scarce.

We can simply define *data* as ‘the symbols on which operations can be performed by a calculator, either human or machine’. Data conveys information about any conceivable entity—the stars, a unicorn, you. The following four types of data can be distinguished, as made by Floridi (1999): (a) *primary data* is the main sort of data an information system is designed to convey; (b) *metadata*, when data is about data. For example, the creation date or the creation place of another piece of data; (c) *operational data*, related to the usage, performance or command of the

Fig. 1.1 Data, information and knowledge



information system and (d) *derivative data*, when data has been extracted from the other types of data. The consideration of what is data and what is metadata is inseparable from the use that is going to be made of it; and what is metadata for one receiver may well be data, and a valuable one, for another receiver. The same blurred frontiers exist among the other types of data.

Data are grouped in messages that transmit some *information* in a communication channel from a sender to a receiver. One single piece of data has value inasmuch as it can represent a message with meaning in a context, that is to say, convey information. In other words, data can be seen as information without meaning. Extracting information from data is not always an obvious task.

Through the study and interpretation of data it is sometimes possible to extract valuable information. When this information is considered during the course of a decision process, then that information is called *knowledge*, at least under the most utilitarian gnosological dogma. If choice is an important element of democracy and decisions ultimately depend on data (processed either rationally or irrationally), we can conclude that data is at the base of democracy.

1.2.2 The Web as a Source of Data

In the World Wide Web, the pages that are visited when one does ‘internet surfing’ are a set of documents globally accessible and hosted in distantly located computers. These documents are text files in HTML format (richly formatted text), images, videos and small computer programs (scripts) among other file types. Documents are accessible because the variety of heterogeneous data transmission technologies, including optical fibre, radio links or network cables, observe the same standard protocols, thus enabling their interoperation.

The internet protocols determine that whenever somebody browsing a web (the client computer) types a web address in the web browser, like `http://site.com/page`, an internet address (IP address, from ‘Internet Protocol’ address) is returned from the name (`site.com`) and that computer (server) is contacted to retrieve the requested document. The protocol ruling the exchange of commands and documents between a client and a server in the web is the ‘HTTP’ (Hyper Text Transfer Protocol). The term *hypertext* makes reference to the fact that documents typically include links to other pages, either hosted locally in the same computer or remotely in another server.

The pieces of information in the Web are arranged as a complex network of interconnected *documents*, vaguely resembling the way neurons are connected in the human brain, or the way our ideas connect to other ideas. But the web is a source of extraordinarily valuable *data*. The documents in the web are rich in tables, diagrams, charts, infographics or simply numbers dropped among dull text paragraphs. These are all pieces of data. However, these data cannot be exploited in an efficient manner. First, because they are not always directly accessible. Some numbers may be given in a pie chart published as a raster image, and they can only

be extracted with OCR (optical character recognition) techniques and with much uncertainty. Second, because sometimes data is published as text, but then it lacks context—it is not information but a collection of meaningless raw numbers. These pieces of raw data are useless for computer algorithms because they cannot be systematically extracted and processed.

Different pieces of data referring to the same entity are totally disconnected in the web of documents and they lack any link that permits increasing the knowledge on specific entities. Pieces of relevant data in distant locations cannot be thus automatically related or compared. Whenever global identifiers for entities do not exist or they are not used, matching pieces of information becomes a cumbersome task (e.g. Shakespeare, W. vs. *William Shakespeare*) and is prone to errors. In other occasions data is well structured in large raw files using well established identifiers (e.g. ISBN for books), but then they are offered as a bulk file for download, without the ability to be queried in individual accesses. A large file has to be downloaded before it can be processed, rendering unpractical its use.

The task of extracting data from Web resources can also be a hard one because data is offered in a myriad of formats, sometimes described in closed specifications and in any case specific for different domains and requiring dedicated processing.

All these hurdles make it difficult to effectively use the billions of pieces of data that as today—in one way or another—are present on the web. In practice, the potential of the web as a source of data is lost.

Publishers on the Web (from web bloggers to public institutions) are in general interested in publishing content as fast as possible whereas possible consumers of data on the Web would like to find carefully described and well formatted, high-quality data. There is an evident mismatch between occasional data producers and data consumers with no easy solution. Two opposite approaches have been proposed.

The first approach places the burden of work on the data consumer: content publishers are not going to make any effort without reward and data consumers have to assume they need intelligent tools and more clever search engines, capable of extracting information even from unstructured content. In a word, the first approach relies on Google being more intelligent every time. The second strategy consists of easing the task of high-quality publishing, providing a set of specifications and good practices for data to be on the web and trusting that at least a fraction of the data publishers will follow them.

None of these strategies has proved to be the ideal solution, but at least this second option offers the possibility of producing data within a larger web: the web of data. This chapter describes the new web of data relying on the specifications of the World Wide Web Consortium (W3C), and its most refined form, known as *linked data*.

1.3 Linked Data

Linked data is only the most refined form of publishing data on the web according to the W3C specs. The W3C describes 35 good practices for publishing data on the web (Farias et al. 2017), but only when networked in the web is its value fully realised. This data network is sometimes referred to as the ‘Web of Data’, a term with a more practical emphasis than the older but equivalent ‘Semantic Web’. The ‘*Semantic Web*’ was conceived in 1999 by Tim Berners-Lee, founder of the Web:

I have a dream for the Web [in which computers] become capable of analysing all the data on the Web – the content, links, and transactions between people and computers. A ‘Semantic Web’, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. (Berners-Lee and Fischetti 1999)

Soon after, new technical specifications appeared striving to implement the Tim Berners-Lee dream. These specifications were not, however, aimed at creating an independent web but at improving the existing one:

The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. (Berners-Lee et al. 2001)

The new data network thereby created has started to grow slowly and silently. First, enthusiastic researchers and computer scientists started dumping datasets, then, public institutions followed; finally, for-profit companies joined the effort.

The Web of Data shares with the World Wide Web the same problems, deficiencies and challenges: the information quality is highly irregular, its availability too unstable and the credibility of the sources uncertain. But few questions that the Web of Data is the seed of a new paradigm where humans are giving way to machines in the use of the internet; and a new sphere of communications where both senders and receivers are intelligent machines and humans play a lesser role.

1.3.1 *Universal Identifiers*

The first key idea of the Semantic Web is that every entity—animate or inanimate, particular or abstract—is liable to have an identifier: the universal resource identifier universal resource identifier or URI. Data in the Web of Data refer to entities very precisely identified.

URIs are sequences of characters with several parts separated by dots and slashes. For example, URLs (universal resource locator), which are the web addresses that are introduced in a web browser to get a page, are also a kind of URIs. This coincidence that makes URIs to be a superset of the URLs is not accidental: an expected behaviour of typing a URI in a web browser is that information on the identified object is retrieved. The string of characters used to identify a thing, magically retrieves more information on that thing if the HTTP

protocol is used to query the right computer. We say that a URI *resolves* when it has the form of URL and it can be navigated.

Perhaps we have not appraised well enough the importance of the URIs as identifiers and their ambitions, for URIs aim at naming every object in the world in a uniform manner. Some people claim ‘*there is nothing in a name*’, and a rose by any other name would smell as sweet. However, designating objects is not a neutral act—in other times this was a sacred act—and it reveals a specific worldview. URIs tend to assume simple hierarchical relations between authorities. For example, a fictitious domain `mydept.myorganisation.uk` actually embodies the idea of a Department (`mydept`) organically depending on a certain organisation (`myorganisation`) in turn located in the UK. Relations are not homogeneous (part-of vs. located-in) but suggest a tree structure. This tree structure is sometimes used to classify the type of resources described, in strings like `type_of_resource/identifier`.

The feature of URIs being at the same time both identifiers and the means to retrieve information in an easy manner is an invitation for information to be retrieved and the whole concept fosters fluent information flows.

1.3.2 *Linked Data and RDF*

The second key idea of the Semantic Web is that information can be given about any URI identifier. For example, Thomson Reuters, a giant company whose business is information, has collected a database of organizations from all over the globe called *permid*. In this database, each organization is identified by a URI. Thus, a fictitious company, let us say ACME Inc., is identified with the following URI: `https://permid.org/1-4296162760`.

If this URI (which is also a URL) is introduced in a web browser, a nicely laid out webpage will be displayed to the user. Actually, the web page is not impressive because there is not much information in the *permid* database: the headquarters address, the country where it is incorporated and a few other values. Other similar databases, like *crunchbase* or *opencorporates*, offer some more information, like the relevant shareholders or the people in executive jobs. However, *permid*'s ambition is big, as the ultimate purpose is ACME Inc. to be uniquely identified by the permid URI—replacing one of the functions of a public Commercial Registry. In some manner, this ambition is being fulfilled, as the acceptance of Thomson Reuters' ids has not stopped growing.

But there is more. When a machine resolves that URI, specifically demanding data, the retrieved answer is not the beautifully formatted HTML document in the figure above. Rather, a succinct dataset is returned, in a much more precise and structured format. The next figure reproduces the text message that would obtain a machine in whose HTTP request headers the proper code is given.

```

@prefix tr-common: <http://permid.org/ontology/common/> .
@prefix fibo-be-le-cb:
<http://www.omg.org/spec/EDMC-FIBO/BE/LegalEntities/CorporateBodies/.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
@prefix tr-org: <http://permid.org/ontology/organization/> .

<https://permid.org/1-4296162760>
  a tr-org:Organization ;
  tr-common:hasPermId "4296162760"^^xsd:string ;
  tr-org:hasActivityStatus tr-org:statusActive ;
  tr-org:isIncorporatedIn <http://sws.geonames.org/6252001/> ;
  fibo-be-le-cb:isDomiciledIn <http://sws.geonames.org/6252001/> ;
  vcard:organization-name "ACME Inc"^^xsd:string .

```

Details on the meaning of this piece of information are not relevant now, but the idea is that of having data describing an entity identified by an URI identifier. The URI <https://permid.org/1-4296162760> is special because it is being used deliberately to identify an entity and it is special because its resolution offers information suitable for both machines and humans.

The piece of data shown above is in a form known as linked data and it follows the best Web recommendations for publishing data online. It is not an Excel file, it is not an excerpt of a relational database. Instead, the piece of data above is RDF (*Resource Description Framework*). RDF is not a data format, but an information model which can be incarnated in different ways—for example XML or JSON. An RDF graph is a set of units of information known as RDF triples. Each of the RDF triples represents a sentence, an atomic unit of information linking three entities. These entities are known as subject, predicate and object, resembling the equivalent concepts in language studies.

In the daily use of language, however, we often use structures more complex than a subject, a verb and an object (like in *Heracles stole apples*). But we can always chain simple sentences to add information (and that *apples were golden*). Thus, using the constituents of one sentence in another sentence, arbitrarily complex pieces of information can be given. If we draw these relations, we see these RDF triples weave a web of connections. An example of RDF sentence, extracted from the ACME example, with a subject, a predicate and an object follows:

```

SUBJECT: <https://permid.org/1-4296162760>
PREDICATE: <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
OBJECT: <http://permid.org/ontology/organization/Organization>

```

The first line above is the subject, and it is a URI identifying ACME. The second line is the predicate meaning ‘is a kind of’. Finally the third line, the object, is URI representing the abstract concept of “organization”. We may understand this RDF triples means ‘ACME is an organization’.

Let us imagine that the Thomson Reuters’ *permid* database of organizations exactly devotes 6 RDF triples to ACME. These 6 triples are represented in the following code excerpt; each of the RDF triples has been shown separated by a

blank line. The subject in all the triples is `<https://permid.org/1-4296162760>`, which is the URI of ACME in the permid database. The predicate is also a URI in each of those 6 cases, including words like `type`, `hasActivityStatus`, or `isIncorporatedIn`—words follow each other without blank space because they are not allowed in URIs. Finally, the object in each of the RDF triples is either a URI or a value, the former given between angle brackets and the latter given between quotation marks. Values are also known as constants or literal values.

```
<https://permid.org/1-4296162760>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://permid.org/ontology/organization/Organization> .

<https://permid.org/1-4296162760>
<http://permid.org/ontology/organization/hasActivityStatus>
<http://permid.org/ontology/organization/statusActive> .

<https://permid.org/1-4296162760>
<http://www.omg.org/spec/EDMC-FIBO/BE/LegalEntities
/CorporateBodies/isDomiciledIn>
<http://sws.geonames.org/6252001/> .

<https://permid.org/1-4296162760>
<http://permid.org/ontology/organization/isIncorporatedIn>
<http://sws.geonames.org/6252001/> .

<https://permid.org/1-4296162760>
<http://permid.org/ontology/common/hasPermId>
"4296162767"^^<http://www.w3.org/2001/XMLSchema#string> .

<https://permid.org/1-4296162760>
<http://www.w3.org/2006/vcard/ns#organization-name>
"ACME Inc"^^<http://www.w3.org/2001/XMLSchema#string> .
```

The six RDF triples above can be represented in an informal, visual manner in Fig. 1.2. Resources are represented as ovals, literals with rectangles. Every triple is represented as an arrow, where the subject of the triple is the origin and the object the destination. Prefixes have been used to shorten the URIs.¹

There are some rules, a few, determining how a RDF triple can be built—the minimal information unit in the web of data. One of these rules determines that subjects and predicates in the RDF triples must be URIs, whereas objects can be either URIs or literal values. Nothing prevents a URI appearing in a triple as subject to be part of another RDF triple as object, or vice versa. In the sentence ‘Heracles stole the apples’, ‘the apples’ are the direct object (object in RDF terminology), but the same apples are the subject in the second exemplary sentence (the apples are gold). Given that URIs can represent any conceivable entity (resource) and given that RDF triples can be chained once and again, we can say that RDF can express any thing about anything. Humans are able to convey much more information with hardly a few words, but this is due to the fact that we humans share an implicit

¹<https://www.w3.org/TR/xml-names>.

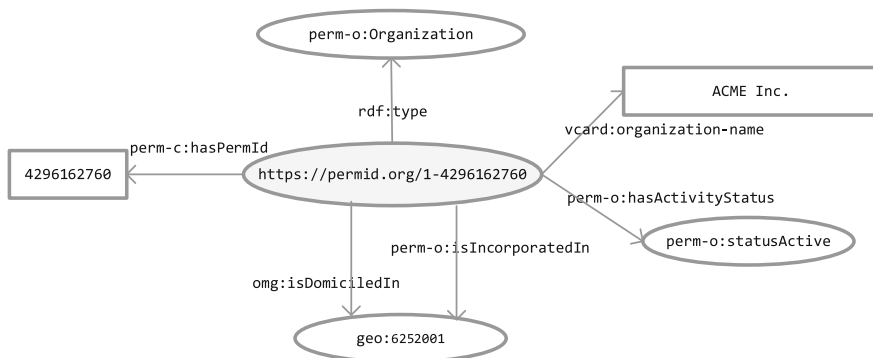


Fig. 1.2 Six RDF triples represented in a diagram

context, a background knowledge known to both emitter and receiver. But nothing, at least in theory, would prevent this context to be codified with another set of RDF triples.

Entities mentioned in an RDF graph can refer to both general ideas and specific individuals. The following code excerpt displays two out of the six RDF triples mentioned before, in the same format where each RDF triple is a set of three lines (S-P-O) separated from the next RDF triple by a blank line.

```
<https://permid.org/1-4296162760>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://permid.org/ontology/organization/Organization> .

<https://permid.org/1-4296162760>
<http://www.w3.org/2006/vcard/ns#organization-name>
"ACME Inc"^^<http://www.w3.org/2001/XMLSchema#string> .
```

The meaning of the first triple is ‘ACME is an organization’. The second triple means ‘ACME has by organization name ACME Inc.’. ACME can be a real and concrete organization, whereas organization is just an abstract concept. In fact, ‘organization’ is a common noun while ACME is a proper noun.

Some philosophers in the past debated about the real existence of these abstract concepts—like organization—and posed the so called problem of the universals. Thus, the realist school claimed that universals were real, they existed and they were different from the particulars instantiating them. On the contrary, the nominalists denied the existence of universals both in an immanent manner (in the particulars) and in a transcendent manner (out of the particulars). In RDF, which is nothing but a language, both universals and particulars are in the same plane and there is no specific difference: a URI can identify both abstract concepts (organization, city) and concrete concept (Heracles, Japan) without any explicit reference to their nature.

1.3.3 *Data Models, Ontologies and Ontology Design Patterns*

The distinction between concrete things (the zip code of the ACME headquarters) and abstract concepts (the idea of organization) is syntactically non-existent in RDF. However, we shall distinguish between pieces of *data* and the *terms* of a vocabulary.

Any URI can be used in any RDF triple without further limitation. However, URIs with general ideas such as ‘organization’ are usually URIs which have been attributed more properties somewhere else, such as a definition, its relation with other similar concepts, its constituents or other properties inherent to its nature. Very often, the person or entity specifying the knowledge about a concept proceeds in the same manner with other concepts in the same domain, covering a specific area of interest and building one domain vocabulary. The complexity of vocabularies varies between a mere list of concepts and a complete ontology with a large amount of knowledge having been specified.

Gruber defined ontology as an ‘explicit specification of a conceptualization’ (Gruber 1993), Studer as ‘a formal, explicit specification of a shared conceptualization’ (Studer et al. 1998). Both definitions speak about conceptualizations made explicit, and the language to make them explicit today is OWL. An OWL (Web Ontology Language) ontology is asserted as a set of RDF triples, and it is, in fact, an ontology in the philosophical sense of the word, for it describes a collection of beings and their properties and relations. Ontologies can cover the whole universe of knowledge, or they can be limited to a specific area. In the latter case they are known as domain ontologies. Ontologies aiming at describing any piece of human knowledge can become huge: for example CYC keeps one of the largest knowledge base in the world and it describes several hundreds of thousands of terms carefully organized (Matuszek et al. 2006), competing with Yago (Suchanek et al. 2007) and others. On the contrary, domain ontologies can be as small as a few dozen triples. Some of these ontologies are mere catalogues of lexical resources. For example, WordNet (Miller 1995) comprises one hundred thousand terms, including nouns, verbs and adjectives. Nouns are related to other nouns that are hyperonyms, hyponyms, meronyms or holonyms.

Ontologies can cover different needs, from representing the consensus in a certain domain (namely, keep a list of definitions), to determining the execution of a computer application. In the latter case, the knowledge base is conceptually divided into two large blocks: the block with terminological information (or T-Box) and the block with information about the individuals instantiating those abstract concepts (A-Box).

There are multiple ways of modelling a reality with ontologies. Likewise, there are multiple ways of implementing an algorithm or designing a relational database. However, it is a good practice to solve recurrent problems with common solutions, because the solutions will have been tested, because others can better understand one’s work and because there is no need to reinvent the wheel. Much like using

design patterns is a common practice among software engineers, *ontology design patterns* (ODP) should be a common practice among ontologists.

Ontology Design patterns were proposed in 2005 by Gangemi, and since their inception a few dozen have been described and published online, with the sole purpose of being reused as building blocks. Their influence, however, is unquantifiable, as the use of patterns is never acknowledged and probably less than what was expected. The reuse of individual terms have been fostered more actively by search engines (<http://vocab.cc>) and ontology repositories (<https://lov.linkeddata.es>). Indeed, ontologies have been defined in the legal domain.

1.3.4 Features of the Semantic Web

In linguistics, semantics is the science that studies the meaning of symbols. If we hear an ambulance siren wailing, we will probably interpret that a sick person is traveling inside. If we use now the ACME identifier (<https://permid.org/1-4296162760>), the careful reader will know that the headquarter is in the state of Michigan in the United States. The communication of a set of RDF triples from one agent (man or machine) to another agent is an act of communication and therefore words like ‘syntax’ or ‘semantics’ have full validity in this context.

The ACME URI is a linguistic sign, a signifier, which evokes a meaning (the idea of the company ACME). Computers with access to the web of data have a precise image of ACME, which can be accounted for, and it is indeed the information in the Thomson Reuters database and other possible mentions in the web of data. Computers might quantify how many facts they know about ACME. We, humans, are not able to determine what we know about ACME, nor the reactions that it provokes in us. Some will recall a bad experience with ACME, some will recall their favourite product from ACME; but no one will be able to know the subconscious.

The semiotic triangle of Ogden and Richards, applied to the human language, links three entities: the mental image (my idea of the ACME) with the sign (the sound of the word ACME Inc.) and both with the real object (the entity ACME). We may draw an equivalent semiotic triangle for the Semantic Web, as Sowa first suggested (Sowa 2010). Both are represented in Fig. 1.3, which adapts the figure in ‘The Meaning of Meaning’ by Ogden and Richards (1923).

The symbol invokes a meaning, the meaning refers to a referent. There is no direct relation between the symbol and the referent other than through the signified, which is an idea. Making it simpler and applied to the spoken language, the triangle puts in relation words with worlds with ideas—stressing the influence of language upon thought. In the human language, Saussure’s concept of arbitrariness holds, and there is no direct connection between signifier and signified. Save for onomatopoeias, words do not resemble real objects. In the Semantic Web, the language

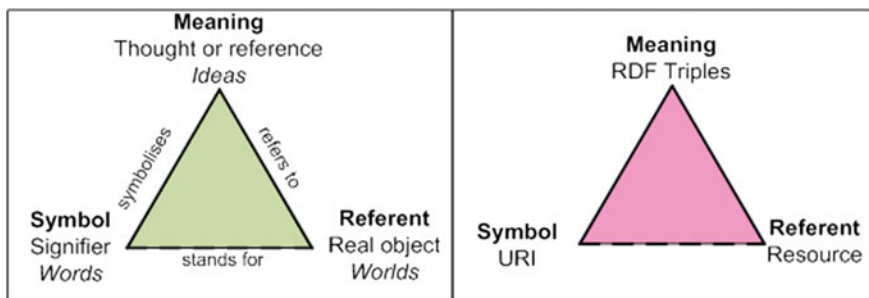


Fig. 1.3 Ogden and Richard’s triangle adapted to the semantic web

is not entirely addressed to computers and symbols are not pure numbers but URIs with some words meaningful to humans in it.

Between humans, the relation between symbol and meaning is a complex one: the word “rain” may denote “drops of water falling” in its primary meaning, but it may connote “sadness and melancholy” in subjective meanings. As of today, machines can only denote primary meanings, and no computer has managed to emulate the richness of a human spoken communication, with all its ambiguities, double meanings and implicit connotations. Computers have not reached lyricism.

Syllogisms are structures of valid reasoning that were studied by Aristotle. Thus, if ‘all men are mortal’ and ‘Socrates is a man’, then we can derive that ‘Socrates is mortal’. These two premises entail a conclusion. Each of the sentences represents some knowledge. We might say, that if we represented each of the two first sentences with a single RDF triple (and their simple structure favours that), we may deduce the third one. These kind of sentences are categorical propositions and their conveyed knowledge is limited to sentences of the sort ‘some (or all) members of category A belong to category B’. But other kinds of reasoning are also possible.

In general, symbolic logic is the branch of science that studies valid forms of reasoning. Logic systems define a language, with an alphabet with symbols and some syntactic rules that determine which combinations of symbols are well formed. Logic systems also define inference rules, which can be applied to produce new formulas from existing ones. Valid reasoning only grants that false premises will never be derived from true premises—please note that truth and falsehood are attributes exclusive of the language, and the logical languages are simply languages. The concepts of truth and falsehood would not exist if there were no languages at all—the observation was made by Hobbes.

Computer ontologies have a logical foundation that enable some reasoning tasks. In particular, OWL ontologies are formalized as one of the Description Logics well described by Baader et al. (2003). The RDF triples can be the proposition in logical arguments that produce new RDF triples. One RDF triple may say that ‘ACME is an organization’. One ontology may say that ‘Organizations have Agents as members’. One reasoner may derive that Action has Agents as members. The millions of triples in the CYC ontology mentioned before might be used in complex

reasoning. To date, there are not many computer applications using this powerful knowledge base, but the potential is huge.

The atomic units of information in the Semantic Web, namely, the RDF triples, do not live in a single location but they are distributed in computers all around the globe. With a uniform technology one can get access to either ontological assertions (‘all men are mortal’, ‘organizations have agents as members’) or to mere data (‘Socrates is mortal’, ‘ACME has headquarters in Michigan’) published by heterogeneous entities. Data is usually published as datasets, namely collections of registers about a topic in particular. Thus, the Thomson Reuters dataset on organizations has a dozen triples for each of the three million organizations they consider.

The peculiarity of the Semantic Web is that data are interconnected at a global level. The concept of publishing a collection of data is not a novelty, but the concept of publishing a collection of data massively connected to data and vocabulary terms published by others, certainly is. Let us consider as an example one of the RDF triples mentioned above:

```
<https://permid.org/1-4296162767>
<http://www.omg.org/spec/EDMC-FIBO/BE/LegalEntities/CorporateBodies/isDomiciledIn>
<http://sws.geonames.org/6252001/> .
```

This RDF triple can be interpreted as ACME is a legal entity domiciled in a certain place identified by geonames. The idea of ‘domicile’ is invoked using an identifier published by a third entity. This triple thus links three entities whose definitions are given by computers in London, Massachusetts (USA) and Bayern (Germany). Two of them belong to private companies (Thomson Reuters and Unxos), the third one to a not-for-profit technology standards consortium (OMG). In the other triples describing ACME, some more vocabulary terms and data are referred (defined for example by W3C).

The data published by geonames about the referred entity available under <http://sws.geonames.org/6252001/>, namely the USA, happens to be exactly 167 RDF triples, with information like the name of the place in different languages, or different coordinates with geolocation. Some of these 167 triples declare that the entity (USA) matches other records in other datasets, like DBpedia (Auer et al. 2007). DBpedia is a dataset published by an association (located in Leipzig) which publishes data extracted from Wikipedia as RDF. DBpedia is the only link that geonames makes to external data source, but the metadata refers to other external data, such as the Creative Commons license. The Creative Commons license is expressed with 90 RDF triples, and it is a dead end in the sense that no further datasets are linked from it. The information about the USA in DBpedia consists of 260 triples densely linked to other datasets published by different sources, like Eurostat, CYC or Freebase (Bollacker et al. 2008)—the number of accessible triples in a second level is already high.

DBpedia is actually massively linked by other datasets—even permid in some of their registers. Given that datasets (and vocabularies) reference each other, we may

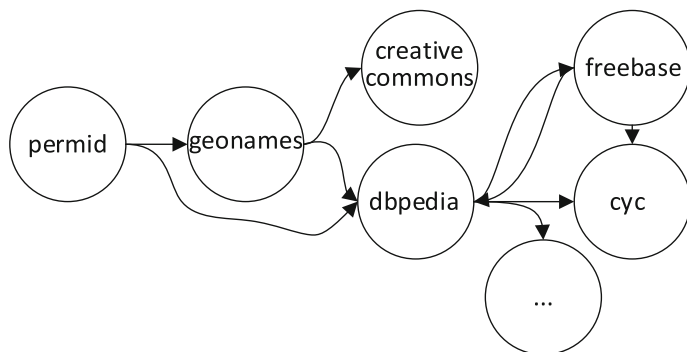


Fig. 1.4 Reference relations between several datasets

think of a graph, a data structure defined by nodes and edges that link them. These edges are directed, namely, they have a direction (for example from permid to geonames, but not vice versa). If we draw each dataset as a node and each connection between two datasets as an edge, we may create a figure as follows (Fig. 1.4).

If a dataset has their entities dereferenceable (the URIs identifying entities resolve with data when properly browsed with the HTTP protocol) and if these entities are linked to other datasets, then the dataset qualifies to be part of the Linked Open Data cloud (LOD). This data is better known then as linked data (the O in the LOD making reference to the idea of open standards, rather than data being openly licensed).

The datasets linked in the LOD cloud has not stopped growing in the last few years. The number of datasets is now so high that it cannot be comfortably fitted into a sheet of paper or a slide in a presentation—like the Tim Berners-Lee presentation shown in the next image. DBpedia is still represented at the center of these diagrams. Not in vain, many see in Wikipedia the Universal Encyclopedia that was at the core of H. G. Wells’s book the World Brain (Wells 1938).

1.3.5 Rights in the Web of Data

RDF is mostly used to represent facts, positive assertions that can be either true or false, like the sentence ‘Heracles stole apples’. But we humans also use other types of expressions, referring to what can be done, to what must be done and to what must not be done. These permissions, obligations and prohibitions are called deontic expressions and can be also represented in RDF with proper vocabularies.

For example, the well-known Creative Commons licenses have been also represented as RDF by the Creative Commons foundation using their own vocabulary, which declares some terms such as `cc:Prohibition` or `cc:Permission`.

The Creative Commons vocabulary defines the necessary terms to represent the most important concepts in Creative Commons licenses, but it does not aim any further. Other vocabularies are more general, and rights can be in general represented as linked data (Rodríguez-Doncel et al. 2013). Thus, the Open Digital Rights Language (ODRL) is a more versatile policy language intended to be used in different domains: financial information, content in mobile devices, ebooks, news and others. ODRL was first specified in 2000 as an XML language, but more recently, the W3C has extended the language and has included a RDF serialization based on an ontology (Ianella et al. 2018) in its latest version.

The ODRL language permits representing permissions, possibly subject to certain restrictions (‘you have access to this file but only in France’), prohibitions (‘do not make derivative works’) and duties (‘you must inform the licensor’); with remedies if rules are not satisfied and a complete suite of policy types suitable for agreements, offers of assets (possibly at a certain price) or privacy policies.

ODRL does not provide any mechanism to digitally enforce the rights, mostly because this operation is not usually feasible beyond the mere access control. Yet, the value of ODRL should not be underestimated, as it enables the automated processing and administration of rights, making easier the search-by-license feature (when looking for images in a Google search, images can be filtered by rights information), the reasoning on rights expressions (it is possible to compute whether two licenses are compatible or not, as shown by Governatori et al. in 2013).

Moreover, the mere existence of policy languages with regulatory power and their acceptance by internet users, is transforming the mere nature of law. The *pragmatic turn* (Casanovas et al. 2017), which considers users’ needs and contexts to facilitate the automated interactive and collective management of knowledge, is likely to become an element of growing importance in a future linked democracy as described in the forthcoming chapters.

1.3.6 *Government of the Semantic Web*

The Semantic Web does not have a different physical infrastructure to the Web. Linked datasets, ontologies, vocabularies and other resources are said to be in the Semantic Web as long as they are published following the best recommendations of the W3C. There is no centralized authority for the Semantic Web other than the W3C as the editor of purely technical specifications.

Participants in the Semantic Web are companies, public institutions and individuals alike, and this does not seem to be problematic. Let us consider one of the RDF triples mentioned before.

```
<https://permid.org/1-4296162767>
<http://permid.org/ontology/organization/isIncorporatedIn>
<http://sws.geonames.org/6252001/> .
```

As we have seen before, the ACME location is given with a reference to an entity managed by geonames. Geonames.org is a website created by the effort of a single engineer, Marc Wick, that is now maintained by the Swiss company Unxos GmbH. Indeed, Unxos might stop providing the service, but this would eventually be a relatively small problem for Thomson Reuters (publisher of this RDF triple), as they would change the reference in a short time (possibly to a location in DBpedia).

Above geonames, there is only the upper domain manager, the one in charge of .org names. The .org domains depend directly on ICANN (Internet Corporation for Assigned Names and Numbers), who also manage the top-level domains in the hierarchical namespace of the Domain Name System (DNS) of the Internet. As ICANN is the entity who ultimately manages IP addresses and names on the Internet, it is a key institution for the internet and consequently for the Semantic Web too.

Legally, ICANN is a non-profit organization, with a mandate to implement from the US Department of Commerce. After 18 years, as of October 2016, changes have happened in order to transfer some of its management duties to multisector agents of the global community. This model, known as MSG or MSI (from multi-stakeholder governance model or initiative) and described by Savage and McConnell (2015) tries to involve the different stakeholders in the internet government, much like technical specifications on the internet that are often written collectively: this is the case of IETF (Internet Engineering Task Force) or the W3C (World Wide Web consortium), where a large community of companies, researchers and public institutions coexist in a relatively peaceful and productive relation. The MSG is further described in Chap. 5.

This wide and coordinated participation in the edition of rules is quite a rare case. If we make an analogy with the road traffic regulations, we should imagine taxi drivers, truck transporters, local police and Royal Automobile Club members discussing together and deciding on the traffic regulations that will apply next year.

The role of individuals is not minor in the Semantic Web. Many well-known vocabularies and ontologies are the result of the work of researchers working alone or crowd sourced by individuals. Sometimes two vocabularies overlap in scope, covering the same domain. Over time some will survive and some will fall into disuse, being ultimately abandoned and their publication discontinued. It is a notable fact that authority (whether the vocabulary is published by the W3C or by a single individual) is important but not totally determining in this struggle. Technical quality and popularity of the resources are sometimes more important factors than the pure argument of authority. For example, despite the huge investment made by Cycorp, manager of the CYC knowledge base, CYC is secondary to DBpedia, created by a collective effort of internet users. This parallels the case of Wikipedia and the Encyclopædia Britannica, the former being the fourth website most visited in the world and the latter having fallen into a relative digital oblivion.

1.4 Government and the Web of Data

1.4.1 *Open Government Data*

Governments are relevant but not dominant stakeholders in the Web of Data and their role has been so far more about producing than about exploiting it. The term Open Government Data (OGD) is often defined as ‘data produced or commissioned by government or government controlled entities’.

Open Government Data is published in government Open Data portals, which offer thousands of datasets in an organized manner. These portals either actively request data from the different government departments and agencies or passively wait for them to send the datasets. Some of the most relevant portals are the US data portal (<https://www.data.gov/>) and the UK data portal (<https://data.gov.uk/>). Stemming from Obama’s US Open Government initiative in 2009, the US data portal collects almost 200,000 datasets with the purpose “to increase public access to high value, machine readable datasets generated by the Executive Branch of the Federal Government”. The UK data portal, which maintains about 40,000 datasets, was created “to help people understand how government works and how policies are made” and embraces very warmly the linked data principles for putting data on the web.

At least three reasons have been identified for opening government data: transparency (for citizens to know that the government is doing), releasing social and commercial value (assuming again the idea that data is an asset) and participatory governance (engaging citizens in decision making). The use of OGD for the latter purpose has also been studied. For example Davies (2010) takes a theoretical and empirical look to assess who is using OGD and for what purposes, in order to determine the possible implications for different models of democratic change and public sector reform. Shadbolt and O’Hara (2013) also evaluated the UK OGD portal, but participatory governance played a minor role. It is generally agreed that Open Data Government must satisfy at least the eight principles written in December 2007 by thirty open government advocates (including Lawrence Lessig, Tim O’Reilly or Aaron Swartz): that data must be complete, from primary sources, timely, accessible, machine processable, non-discriminatory, non-proprietary and license-free.

1.4.2 *Linked Open Government Data*

Indeed, not every piece of OGD follows the linked data principles. But some relevant datasets in the Linked Open Data cloud have been produced directly by public institutions and many others have been re-formatted by third parties. And

even these third parties have been very often partners in publicly funded research projects—governments have been supporting the development of the Semantic Web, especially in Europe.

Besides OGD, there are many other datasets relative to local or national governments which have also been published. Actually, the whole scope of OGD has been questioned as to whether OGD should stand for “(open government) data” or “open (government data)” (Yu and Robinson 2011). In the latest radiography of the Linked Data Cloud, in 2014, 183 datasets were classified as “government-related”, amounting to 18% of the total (Schmachtenberg et al. 2014). Some of the datasets include the Brazilian politicians (de Souza et al. 2013), the debates in the Italian legislative cameras, data from the Greek police (Bratsas et al. 2011) or European Parliament debates (van Aggelen et al. 2017), to name a few.

We may define the term *linked open government data* as the intersection between government data (which is itself only a fraction of eGovernment), linked data and open data.

1.4.3 *eGovernment and eDemocracy*

The concept of eGovernment is about the better provision of services by public sector organisations by using digital technologies. In a wider sense, and according to the World Bank, “e-Government refers to the use by government agencies of information technologies [...] that have the ability to transform relations with citizens, businesses, and other arms of government”. eGovernment can benefit from Semantic Web technologies in many ways. As an example, RDF vocabularies for the definition of public services offered by municipalities in Europe may help migrants to recognize the same service that is differently named in different regions.

As of 2018, about 49% of European citizens have used at least once an online service offered by a public institution,² and there is a public determination towards increasing this rate. The implementation of eGovernment is systematically evaluated by the public authorities in Europe. For example, the eGovernment Benchmark Study³ monitors the development of eGovernment in Europe, evaluating indicators such as the number of services online, their degree of transparency, or the ability to make administrative processes fully available online. Important declarations have been signed (like the *Tallinn Ministerial Declaration on eGovernment*) as well as specific action plans (*EU eGovernment Action Plan 2016–2020*). Thus, on March 2017, the *European Interoperability Framework* (EIF) was adopted, focused on making digital public services more interoperable. Interoperability in that EIF was understood at four different levels: legal interoperability (which exists when

²Eurostat Information Society Indicators, April 2018, <https://ec.europa.eu/digital-single-market/en/graph-european-society-indicators>.

³http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=48587.

legislation does not impose unjustified barriers to the reuse of data), organizational interoperability (which exists when formal agreements rule cross-organisational interactions), semantic interoperability (which exists when there is a common understanding of exchanged data) and technical interoperability (which exists when information systems allow the free flow of bytes). Whereas this chapter has focused on semantic interoperability, the overall schema is reviewed with more detail in Chap. 5.

When information and communication technologies are specifically applied to empower deliberative democracy, the term eDemocracy is used instead. In eDemocracy, citizens go online to communicate opinions or complaints to the public administrations. The term eDemocracy is the preferred one when information technologies are used in one of the following cases: (i) as tools to strengthen deliberative democracy; (ii) as tools to communicate to the public institutions any kind of complaints, preferences or incidents or (iii) as a space to exercise political rights and participate in the political life. We will also suggest in Chap. 5 to place these different regulatory dimensions under the provisions of the rule of law (i.e. the meta-rule of law).⁴

The Semantic Web technologies have been postulated as a helpful tool to retrieve some meaning out of the online chatter about politics (Hilbert 2009), and it has been said to support the self-organization of people with joint political goals. For example, Belák and Svátek (2010) provided a core ontology for the description of political programs, commitments and trust between people. This work helps people to analyze, compare and discuss political programs, already in great databases like the Manifesto Project. The Manifesto Project offers the policy positions of parties derived from a content analysis of their electoral manifesto, covering over 1000 parties from 1945 until today in over 50 countries (Volkens et al. 2016). Similarly, the Constitute Project⁵ (Elkins et al. 2014), offers as RDF, almost every constitution which has been in force anywhere in the world in the past 200 years. Legislation is offered as linked data in the UK and the Netherlands, with partial engagement also in the USA (Casellas et al. 2011a, b) and Canada (Desrochers 2012) and non-official support in many other countries.

However, there is also a reasonable concern about these tools and datasets remaining at a technical level, without actually reaching the masses. For example, the Linked Leaks datasets, containing information about 200,000 offshore entities that were part of the Panama Papers investigation, were released in 2016 as richly linked data; yet they have not been widely used.

⁴See Casanovas (2015), and Casanovas et al. (2016) for a comprehensive state of the art of Semantic Web applications in the legal domain.

⁵<http://www.constituteproject.org/>.

1.4.4 The Open Data Principles

Most of the data in the Linked Open Data cloud has been published as open data, namely, licensed under very liberal terms. This is the most natural option, as in the Web of Data building on others' resources is the most common practice.

The limits for what is considered open data and what is not open data have been well defined. Open data is data that anyone can access, use and share, according to the Open Data Institute (ODI) whereas openness is defined by Open Knowledge Foundation (OKFN) as situations when anyone can freely access, use, modify and share for any purpose (subject, at most, to requirements that preserve provenance and openness). Both OKFN and ODI have listed the well-known licenses (e.g. from Creative Commons or Open Data Commons) that comply with their definitions and have created visual labels to be easily recognized. In essence, open licenses grant that data can be used without legal barriers.

In the collective conscious, the open software movement has been associated with individual champions such as Linus Torvalds, Richard Stallman or Aaron Swartz. However, the open data movement is being promoted by global institutions. For example, the Group of Eight (G8) has signed in 2013 the "G8 Open Data Charter" outlining a set of five core open data principles to be followed by governments, the World Bank has devoted large resources to promote the adoption of the open data principles and the United Nations has drafted a development agenda called UN Data Revolution largely based on open data. Consistently, governments of most countries have enacted laws for publishing public sector information as open data, under the general principle that data produced with public funding must be openly published.

Open data has some downsides, though. First, it might favour inequality as the strongest become stronger. In theory, individual citizens have free access to information. In practice, only large companies with data science teams can extract actual value from it. These companies will leverage the open data resources for their own benefit, to the detriment of the rest. Second, the risk of re-identifying individuals in anonymised personal data is higher. The fact is that whereas the open data movement is energetically supported by public institutions, internet users and citizens in general have shown little enthusiasm.

1.4.5 Business Intelligence in the Public Sector

In the last decades, the relevance of data has increased as more and more decisions have been entrusted to computers and decision support systems.⁶ Many large

⁶For a review of decision support systems using the Web of Data as presented in this book, see Blomqvist (2014).

companies make their biggest corporate decisions based on the results of complex computer processes that chew tons of apparently worthless data —this is known as *business intelligence* (BI).⁷

Decisions are usually taken as a result of a process in which different sorts of questions have to be answered. First, *descriptive* questions portray a certain reality (what is happening? how much? when? where?). In a second place, *diagnostic* questions look for explanations (why something has happened?). Then, *predictive* questions help forecasting the future (what will happen if I don't do anything?). Finally, *prescriptive* questions determine the best possible action (what should I do?). These answers can feed either a *decision support system*, where the ultimate decision is taken by a human, or a *decision automation system* where actions are executed without human intervention. Some data analytics applications stop at the descriptive stage, some power fully automated systems, like the management systems of the stocks of a retailer.

Most of the data that a company bases its decisions upon (like figures of sales or the customers' location) have an intrinsic value that is zealously protected—they are an intangible asset and its dissemination may favour other competitors. Data, as a commodity, can be also traded in a data market (“*i.e. the marketplace where digital data is exchanged as products or services derived from raw data*”)⁸ in exchange for money. Data markets are being fostered by governments.⁹

But for several reasons, data can also be publicly available under open licensing modalities. Many of the datasets relevant to BI processes had been always available, although not digitally—only as printed statistical yearbooks or in other non-digital forms. In the last few years, data has been massively dumped in the web and its full potential is yet to be realized.

Public administrations lag behind in the application of business intelligence to their decisions and there is not much literature in the area.¹⁰ However, intelligent analyses are quietly being used by public administrations for the better provision of the services they offer (e.g. a municipality optimally planning the transport system). The growth of the amount of available data and the advances of the Artificial Intelligence (AI) algorithms will enable business intelligence to play a more important role in the decisions taken by public administrations in the years to come.

These techniques introduce a slight novelty in a long-standing question: the relation between experts in possession of scientific knowledge and politicians. In the most simplistic approach, the politician takes decisions and the expert provides

⁷The term predates computers. Business intelligence was defined as “*the ability to collect information and react to it accordingly*” in Cyclopaedia of Commercial and Business Anecdotes, R. Millar Devens (1865).

⁸European Data Market study, SMART 2013/0063, IDC 2016.

⁹As an example, see “Building a European Data Economy”, COMM(2017), 9 (final).

¹⁰See the overview of possible applications by Wowczko (2016) or how business intelligence is being applied by the public institutions in the UK (National Archives 2016).

technical advice on how to execute them.¹¹ But the progress of technology not only rationalizes the means to implement the decisions, but also reduces the scope of politics: some of the questions, originally entrusted to the political sphere, can be optimized as well. The space of pure political decision-making is thus reduced by technological advances. The novelty in this question is that experts are also being replaced, in many of their functions, by intelligent machines. Further, the role of *professional experts* is even further diminished, as the expertise of a crowd of non-professionals is now available in the Internet era.

1.5 Conclusion

Data plus the right algorithms equals *information*, the right information used in a decision-making process is *knowledge*—at least according to the data/information/knowledge pyramid model. The power of algorithms is not usually in the hands of individuals, but of large corporations with server farms and dedicated professionals. These algorithms, as almost any other modern technology, are no longer used to control the natural world, but to control other humans. In particular, political campaigns all over the world have allegedly been in recent years strongly influenced by intense data analytics processes powerful enough to tilt the scale.

Before this gloomy scenario, an unexpected actor can still play a role: a cloud of linked data enabling distributed knowledge and facilitating collective intelligence. The Web of Data, and more specifically, the linked data cloud, is a growing universe of connected information published about any matter in any language and accessible by anyone. The open data movement, initially sparked to increase the transparency of public administrations, has gained momentum and its economic and social value is now fully revealed. Public administrations, large and small enterprises, foundations, universities and individuals alike are contributing to creating a web of data, sharing the features of the web that we know is heterogeneous and diverse.

Much of the open source and free software movements have yielded first-class, high quality operating systems such as Linux, and the idea of open content has led to the release of millions of works now published under Creative Commons licenses, the open data movement combined with the semantic web technologies is creating a new data resource available to all. Maybe in the future, machine learning and data mining algorithms running over this pool of data will be also standard tools in the hands of individuals or self-organised collectives.

¹¹This decisionist model of the relation between politicians and experts has been compared with the technocratic model and Dewey's pragmatist one by Habermas (1974).

References

- Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z (2007) DBpedia: a nucleus for a web of open data. In: *The semantic web*. Springer, Berlin, pp 722–735
- Baader F, Calvanese D, McGuinness D, Patel-Schneider P, Nardi D (eds) (2003) *The description logic handbook: theory, implementation and applications*. Cambridge University Press, Cambridge
- Belák V, Svátek V (2010) Supporting self-organization in politics by the Semantic Web technologies. *IFIP ePart*, pp 1–8
- Berners-Lee T, Fischetti M (1999) *Weaving the web: the original design and ultimate destiny of the World Wide Web by its inventor*, 1st edn. Harper, San Francisco
- Berners-Lee T, Hendler J, Lassila O (2001) The semantic web. *Sci Am* 284(5):34–43. <https://doi.org/10.1038/scientificamerican0501-34>
- Blomqvist E (2014) The use of semantic web technologies for decision support—a survey. *Semant Web* 5(3):177–201
- Bollacker K, Evans C, Paritosh P, Sturge T, Taylor J (2008) Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD Int. Conf. on Management of data*. ACM Press, Vancouver, Canada, pp 1247–1250. <https://doi.org/10.1145/1376616.1376746>
- Bratsas C, Alexiou S, Kontokostas D, Parapontis I, Antoniou I, Metakides G (2011) Greek open data in the age of linked data: A demonstration of LOD internationalization. *Proc ACM WebSci* 11:1–4
- Casanovas P (2015) Conceptualisation of rights and meta-rule of law for the web of data, *Democracia Digital e Governo Eletrônico* (Santa Caterina, Brazil), vol 12. 2015: 18–41; repr. *J Gov Regul* 4(4):118–129
- Casanovas P, Palmirani M, Peroni S, van Engers T, Vitali F (2016) Semantic web for the legal domain: the next step. *Semant Web* 7(3):213–227. <https://doi.org/10.3233/SW-160224>
- Casanovas P, Rodríguez-Doncel V, González-Conejero J (2017) The role of pragmatics in the web of data. *Pragmatics Law*. Springer, Cham, pp 293–330
- Casellas N, Vallbé JJ, Bruce TR (2011) From legal information to open legal data: A case study in us federal legal information. In: *POGK 2011: The AAAI Fall symposium on open government knowledge: AI opportunities and challenges*
- Casellas N, Vallbé J-J, Bruce TR (2011) From legal information to open legal data: a case study in U.S. Federal legal information. *SSRN Electron J*. <https://doi.org/10.2139/ssrn.1959931>
- Davies T (2010) *Open data, democracy and public sector reform*. M.Sc. dissertation, Oxford University
- de Souza JF, Siqueira SWM, de Ramos Araújo L, Melo RN (2013) Providing information from Brazilian politicians using linked data. In: *Cases on open-linked data and semantic web applications*. IGI Global, 39–57
- Desrochers P (2012) Recordkeeping and linking government data in Canada. *IEEE Intell Syst* 27(3):50–53
- Elkins Z, Ginsburg T, Melton J, Shaffer R, Sequeda JF, Miranker DP (2014) Constitute: the world’s constitutions to read, search, and compare. *Web Semant Sci Serv Agents World Wide Web* 27:10–18. <https://doi.org/10.1016/j.websem.2014.07.006>
- Farias B, Burle C, Calegari N (2017) Data on the web best practices. W3C Recommendation, 31 Jan 2017
- Floridi L (1999) *Philosophy and computing: an introduction*. Routledge, London
- Gangemi A (2005) Ontology design patterns for semantic web content. In: *International semantic web conference*. Springer, Berlin, pp 262–276
- Goodwin T (2015) The battle is for the customer interface. *Crunch Network*, 3 Mar
- Governatori G, Rotolo A, Villata S, Gandon F (2013) One license to compose them all. In: *International semantic web conference*. Springer, Berlin, pp 151–166
- Gruber T (1993) A translational approach to portable ontologies. *Knowl Acquisition* 5(2):199–229

- Habermas J, Lennox L, Lennox F (1974) The Public sphere: An Encyclopedia Article (1964). *New German Critique* (3):49. <http://doi.org/10.2307/487737>
- Hilbert M (2009) The maturing concept of e-democracy: from e-voting and online consultations to democratic value out of jumbled online chatter. *J Inform Technol Politics* 6(2):87–110. <https://doi.org/10.1080/19331680802715242>
- Ianella R, Steidl M, Myles S, Rodríguez-Doncel V (2018) ODRL vocabulary and expression 2.2. W3C Recommendation, 15 Feb 2018
- Matuszek C, Cabral J, Witbrock MJ, DeOliveira J (2006) An introduction to the syntax and content of cyc. In: AAAI Spring symposium: formalizing and compiling background knowledge and its applications to knowledge representation and question answering, pp 44–49
- McLuhan M (1964) *Understanding media: the extensions of man*. The New American Library, New York
- Miller GA (1995) WordNet: a lexical database for English. *Commun ACM* 38(11):39–41
- National Archives (2016) The digital landscape in government 2014–15 Bus Intell Rev. As of Apr 2017. Available at: <http://www.nationalarchives.gov.uk/>
- Ogden CK, Richards A (1923) *The meaning of meaning*. Harcourt, New York
- Rodríguez-Doncel V, Gómez-Pérez A, Mihindukulasooriya N (2013) Rights declaration in linked data. In: Hartig O et al (eds) *Proceedings of the 3rd international workshop on consuming linked data* (CEUR 1034)
- Savage JE, McConnell BW (2015) *Exploring multi-stakeholder internet governance*. Brown University Bruce W. McConnell, EastWest Institute
- Schmachtenberg M, Bizer C, Paulheim H (2014) Adoption of the linked data best practices in different topical domains. In: *International semantic web conference*. Springer, Cham, pp 245–260
- Shadbolt N, O’Hara K (2013) Linked data in government. *IEEE Internet Comput* 17(4):72–77. <https://doi.org/10.1109/MIC.2013.72>
- Sowa JF (2010) The role of logic and ontology in language and reasoning. *Theory and applications of ontology: Philosophical perspectives*. Springer, Dordrecht, pp 231–263
- Studer R, Benjamins VR, Fensel D (1998) *Knowledge engineering: principles and methods*. *Data Knowl Eng* 25(1):161–198
- Suchanek FM, Kasneci G, Weikum G (2007) Yago: a core of semantic knowledge. In: *Proceedings of the 16th international conference on world wide web*. ACM, pp 697–706. <https://doi.org/10.1145/1242572.1242667>
- Van Aggelen A, Hollink L, Kemman M, Kleppe M, Beunders H (2017) The debates of the European Parliament as linked open data. *Semant Web* 8(2):271–281
- Volkens A, Lehmann P, Matthieß T, Merz N, Regel S (2016) *The Manifesto data collection*. Manifesto Project (MRG/CMP/MARPOR)
- Wells HG, Mayne AJ (1938) *World brain*. Methuen & Company
- Wowczko I (2016) Business intelligence in government driven environment. *Int J Infonomics* 9(1). <https://doi.org/10.20533/iji.1742.4712.2016.0134>
- Yu H, Robinson DG (2011) The new ambiguity of ‘Open Government’. *UCLA L Rev Discourse* 59:178. <https://doi.org/10.2139/ssrn.2012489>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

