

# Value Creation on Open Government Data

Judie Attard  
University of Bonn  
[attard@iai.uni-bonn.de](mailto:attard@iai.uni-bonn.de)

Fabrizio Orlandi  
University of Bonn  
[orlandi@iai.uni-bonn.de](mailto:orlandi@iai.uni-bonn.de)

Sören Auer  
University of Bonn  
[auer@iai.uni-bonn.de](mailto:auer@iai.uni-bonn.de)

## Abstract

*Governments are one of the largest producers and collectors of data in many different domains. As one major aim of open government data initiatives is the release of social and commercial value, we here explore existing processes of value creation on government data. We identify the dimensions that impact, or are impacted by value creation, and distinguish between the different value creating roles and participating stakeholders. We propose the use of Linked Data as an approach to enhance the value creation process, and provide a Value Creation Assessment Framework to analyse the resulting impact.*

## 1. Introduction

Open data initiatives, especially open government initiatives, gained an increasing amount of popularity and attention in recent years. Worldwide, a number of open data movements have been introduced, such as the Public Sector Information (PSI) Directive<sup>1</sup> in 2003 in Europe, U.S. President's Obama open data initiative<sup>2</sup> in 2009, the Open Government Partnership<sup>3</sup> in 2011, and the G8 Open Data Charter<sup>4</sup> in 2013. Collectively, there are three main reasons for opening government data<sup>5</sup>, namely *transparency*, *participatory governance*, and *releasing social and commercial value*. In order to have a well-functioning, democratic society, citizens and other stakeholders need to be able to monitor government initiatives and their legitimacy. Transparency means that stakeholders not only can access the data, but they also should be enabled to use, reuse and distribute it. The success to achieve transparency results in a considerable increase in citizen social

control. Furthermore, through the publishing of government data, citizens are given the opportunity to actively participate in governance processes, such as decision-taking and policy-making, rather than sporadically voting in an election every number of years. Through open government data initiatives such as portals, stakeholders can also be more informed and be able to make better decisions [1]. Finally, all data, whether addresses of schools, geospatial data, environmental data, weather data, transport and planning data, or budget data, has social and commercial value, and can be used for a number of purposes that could be different than the ones originally envisaged. Governments are one of the largest producers and collectors of data in many different domains [2]. Considering its volume (huge amount of data produced), velocity (frequent gathering of data, especially sensor data), variety (different domains), and veracity (uncertainty of data), government data can be considered to be Big Data. By publishing such data the government encourages stakeholders to innovate upon it, and create new services.

The main challenge in releasing social and commercial value is that open data has no value in itself, yet it becomes valuable when it is used [3]. There are many factors within an open government initiative that influence its success. In this paper we attempt to answer the following research question: *What are existing processes of value creation on open government data?* With this research question we aim to address the niche in existing literature with regards to the creation of value on open government data, and therefore provide relevant information to the involved stakeholders of open government data initiatives. Our intention is thus to aid these stakeholders in exploiting government data to its highest potential. In this paper we hence identify the various processes in a government data value chain, as well as dimensions that, in some way or another, have an impact on value creation upon government data. We also distinguish between the different value creating roles of participating stakeholders within the government data value chain, and identify the resulting impacts of value creation and of exploiting data as a product. As an extension to existing related

<sup>1</sup> <http://ec.europa.eu/digital-agenda/en/european-legislation-reuse-public-sector-information> (Accessed 31/08/2015)

<sup>2</sup> <http://www.whitehouse.gov/open/documents/open-government-directive> (Accessed 31/08/2015)

<sup>3</sup> <http://www.opengovpartnership.org/> (Accessed 31/08/2015)

<sup>4</sup> <https://www.gov.uk/government/publications/open-data-charter> (Accessed 31/08/2015)

<sup>5</sup> <http://opengovernmentdata.org/> (Accessed 31/08/2015)

work such as [4], we provide two additional contributions: we propose the use of Linked Data as a possible solution to some of the identified barriers to the creation of value on top of open government data, and we propose a Value Creation Assessment Framework to analyse the success of an open government data initiative. While we focus on government data, it is important to note that most of what we discuss is also valid for generic open data initiatives.

## 2. Methodology

In order to analyse existing approaches undertaken for creating value, based on open government data, we review existing literature on open government data initiatives. We implement a systematic approach, where we define a number of search terms and perform a search on a number of digital libraries. Thereafter, we select which literature to include in our study by applying inclusion and exclusion criteria.

The search terms we defined are a combination of the following keywords: *government*, *data*, *portal*, *open*, *publishing*, *consuming*, and *public*. The latter were selected with the aim of obtaining results relevant to the research question defined in Section 1, or, any initiative that exploits government data in order to create value. We here stick to the definition of **Government Data** to entail any data that is government-related. It may or may not be produced or published by a governmental entity, and it may or may not be made openly available (it can have varying degrees of openness).

In order to cover the largest spectrum of relevant publications possible, we identified and used the most extensively used electronic libraries, namely: ACM Digital Library, Science Direct, Springer Link, IEEE Xplore Digital Library, and ISI Web of Knowledge. To achieve relevant results that are sufficiently comprehensive and precise, we apply these search terms on both the title and the abstract search fields.

After the systematic search was completed, we led a manual study selection based on exclusion and inclusion criteria. Basically, we only considered literature to be relevant if it regarded the actual exploitation of government data. This resulted in 74 publications that form our set of primary studies<sup>6</sup>.

With the research question in mind, we analysed the 74 publications with the aim of identifying

current practices of creating value using government data. We hence provide our observations, comments, guidelines, and conclusions in the rest of this paper.

## 3. Background Literature

Data is increasingly becoming a commodity in our information society. It is steadily becoming the basis for many products and services, such as open data, Linked Data, or Big Data applications. Using open data, specifically open government data, has the potential of not only resulting in economic benefits, but also has good social and governmental impacts. Releasing government data will impact transparency and accountability factors, while the release of specific datasets can encourage stakeholders to create innovative services and boost economic growth.

In order to reflect such a data-centric society, the concept of *value chains* [5] is being used to identify how value is created in order to achieve a product. Authors such as [6][7][8] provide a newer definition of the concept, where they cater for digital dimensions and take into account factors and activities that set a non-tangible data product apart from a more physical one.

The data value chain is at the centre of a knowledge economy [9], where data products provide digital developments to more traditional sectors, such as transport, health, manufacturing, and retail. Being one of the largest producers and collectors of data in so many domains, governments play a vital role in data value chains. Essentially, the strategy beneath a data value chain is to extract the maximum value from data by building on the intelligent use of data sources [9]. The authors of [10] add that value-adding is one of the most important properties for information, where the objective of adding value to information is to develop information products and provide information service with social and economic value.

Essentially, a data value chain forms part of a *data life cycle*. In [11] we specifically define a government data life cycle, based on more generic data life cycles<sup>7</sup>. In this paper, following the research question defined in Section 1, we focus on the processes that involve value creation, namely:

- Data Creation
- Data Harmonisation
- Data Publishing
- Data Interlinking

<sup>6</sup> All primary studies can be accessed here: <http://mnd.ly/1LFgFQJ>

<sup>7</sup> [http://www.w3.org/2011/gld/wiki/GLD\\_Life\\_cycle](http://www.w3.org/2011/gld/wiki/GLD_Life_cycle) (Accessed 31/08/2015)

- Data Exploitation
- Data Curation

Within all of these processes an entity invests efforts into adding value to a data product. In the specific use case of government data, value creation involves *making the data more usable, or otherwise more fit for the intended use.*

#### 4. Value Creation Techniques

Table 1 shows various value-creating techniques within the mentioned processes, as identified through the led systematic literature review. While not comprehensive, we included the most popular and frequently-used techniques from various stakeholders participating in the data value chain.

Government Data Life Cycle Processes	Value Creation Techniques
Data Creation	Data Generation Data Collection
Data Harmonisation	Licensing Format Conversion
Data Publishing	Data Accessibility
Data Interlinking	Link Discovery Data Interlinking Data Integration
Data Exploitation	Analysis Reasoning Visualisation Data Interpretation
Data Curation	Data Updating Data Cleaning/Repairing Metadata Enrichment

**Table 1: Value Creation Techniques categorised according to Government Data Life Cycle Processes**

Data is produced in the day-to-day administration of a governing entity. The simple **generation** of this data is the first step towards its (re) use as a data product. As opposed to data generation, data **collection** involves the specific gathering of data for a particular purpose, for example weather data to evaluate the severity of a drought. Feedback is a good example of both data generation and data collection. While citizens might generate feedback on say, budget data, a governmental entity can collect this data in order to improve future budget allocation. In order for the best value potential, both generated and collected data need to be complete. This means a record has all the information required for an accurate representation of the described data.

The value creation techniques falling under the data harmonisation process have the purpose of making the data more usable through the application

of standards. **Licensing** has the purpose of declaring if and how data can be used. In the case of government data preferable licences would be of an open nature. The **conversion** of formats is required when the original data exists in a non-reusable format, such as PDF, or a proprietary format, such as Microsoft Word. The conversion of such data into an open and reusable format such as RDF would encourage its reuse. The ideal format of the data should be in a machine-processable format.

In order for existing data to be available for reuse, it obviously needs to be publicly available. Thus, data publishing creates value by making data available for reuse. Therefore, data that is provided in a timely manner (data is provided in a reasonable amount of time after creation/generation), without discrimination on its consumers (not requiring any registration), and made **accessible** for all, has the best value creation potential. Moreover, the addition of metadata enables the data to be more discoverable, thus enhancing this potential. Popular methods of publishing data include SPARQL<sup>8</sup> endpoints and Application Program Interfaces (APIs).

Data interlinking is an essential process in value creation that paves the way for further data exploitation. **Link discovery** involves finding the connection between different sets of data, for example a bus timetable dataset could have links to a map dataset through common geo-location data in both sets. The **interlinking** of such datasets could then provide us with invaluable insight, and the potential innovative use of data that is different than the original envisaged use. Linked Data can then be added to the LOD cloud<sup>9</sup>. Data **integration**, as opposed to interlinking, involves the incorporation of two different datasets. Generally, this requires that there is some sort of link existing between the datasets. In this way, value is created through the resulting dataset that is enriched with further knowledge/information.

Data **analysis** comprises a number of methods, such as data mining, pattern identification, and trend analysis, which are used to inspect the data with the aim of discovering useful information. As opposed to data analysis, **reasoning** has the goal of identifying implicit knowledge, for example through the use of first order predicate logic. Therefore, through the use of existing data/information, this value-creating technique would provide us with new information or knowledge. **Visualisations** involve a visual representation of data that, similar to data interlinking

<sup>8</sup> <http://www.w3.org/TR/rdf-sparql-query/> (Accessed 31/08/2015)

<sup>9</sup> <http://lod-cloud.net/> (Accessed 31/08/2015)

and data analysis, can provide us with a new insight. Visualisations can also be used to provide ‘stories’, since they are more easily interpreted than raw data. Data **interpretation** can be used in conjunction with the techniques within the data exploitation process. This technique involves the manual (non-machine) understanding of data in the provided context.

The **updating** of existing data enables the data to remain valid throughout its existence. Data **cleaning/repairing** involves a number of different approaches, including removing data inconsistencies, correcting inaccurate data, correcting data format (e.g. date in day-month-year rather than year-month-day). This technique gives the data a higher quality level and encourages its reuse. Similarly, **metadata** also enhances a datasets reuse potential. By enriching a dataset’s metadata, a dataset is made more easily discoverable by potential users [12].

#### **4.1. Stakeholders: Beneficiaries, Contributors, and Their Roles**

Government data, or public sector information, is a resource holding great potential for a large number of stakeholders. Governmental agencies, citizens, non-profit organisations, and businesses, are but a few of the potential stakeholders who, through the exploitation of open government data, can reap substantial benefits. Since the efforts of the latter stakeholders remain largely uncoordinated, their motivations, levels of expertise, and priorities differ. In this section we proceed to identify and explore the various stakeholders who, either through value creation or other means of consumption, use open government data.

The most obvious role of *governments* in open government data initiatives is the role of a data provider. Yet, public entities are also the direct beneficiaries of their own published data. Through transparency as a motivation, the publishing of data can increase accountability, and moreover inhibits corruption. In turn this increases citizens’ trust in their government. The analysis of government data, such as budget data, has the potential of increasing efficiency and influencing decision-making. Innovations based upon such data can also be used to provide more personalised public services, thus increasing the quality of the interactions between governments and their citizens.

Through publishing government data, *citizens* are given the possibility of participating in governance processes. Apart from being able to make more informed decisions, citizens are sometimes given the opportunity to take part in participatory governance, where they are given a say as to how, or for what,

budget should be prioritised. Citizens can also participate in open government initiatives by being data *prosumers*. By this we mean citizens who both produce and consume data. For example, the Fix My Street<sup>10</sup> application provides a platform where anyone can submit an existing problem in a street, in order to indicate the problem areas to the government. In this crowdsourced co-production of value, we have geographical data consumption, and street issues data production. Open government data certainly has the potential of increasing citizens’ quality of life.

*Non-profit organisations*, such as non-governmental organisations (NGOs) or Civil Society initiatives, can have a huge difference in their goals. Examples of such organisations include the Sunlight Foundation<sup>11</sup> and the Open Knowledge Foundation<sup>12</sup>, present in various countries. Organisations such as the latter usually share the goals of demonstrating the benefits of opening governmental data both to the general public and to the governments themselves. They also play a vital role as intermediaries who can identify key datasets that have the potential of being very valuable if published as open data.

Private companies, small to medium enterprises (SMEs), entrepreneurs, and other *businesses*, have the potential of not only making an economic profit through using government data, but can also create more jobs, and (depending on the nature of the service) also provide innovative services that increase the beneficiaries’ quality of life and indirectly impact job creation in this field. While the sole access to data does not provide competitive advantage, private entities can innovate upon the available data to provide value-added services.

Whatever the stakeholder’s nature (citizen, governmental entity, NGO, etc.), we identify four roles in which they can participate to create value:

- **Data Producer/Publisher:** A data producer is the entity that creates or generates the data, while a data publisher makes it openly available. These roles can be fulfilled by the same entity.
- **Data Enhancer:** This role involves creating value through the actual manipulation of the data. This can be achieved through the techniques described in the data harmonisation, interlinking, exploitation, and curation processes.
- **Service Creator:** A service creator entity has the task of using open government data to provide a service. This can take the shape of a website, a

<sup>10</sup> <https://www.fixmystreet.com/> (Accessed 31/08/2015)

<sup>11</sup> <http://sunlightfoundation.com/> (Accessed 31/08/2015)

<sup>12</sup> <https://okfn.org/> (Accessed 31/08/2015)

mobile application, information access points, etc.

- **Facilitator:** This role involves entities that, in some way or another, aid the other stakeholders in using, reusing, or exploiting, open government data. This can be done through the provision of software, services, or other technologies.

## 4.2. Barriers, Enablers, and Impacts of Value Creation.

Within the data value chain, value creation is both dependent on a number of dimensions, and also results in impact on other dimensions. Based on efforts in the primary studies (See Section 2), and other literature such as [3][4][13][14], we identify the dimensions with the strongest impact. Figure 1 maps their relationship, where a number of dimensions act as *enablers* or *barriers* towards value creation. In turn, the value creation process impacts a number of other dimensions. The stakeholders, while they give input for value creation, are also impacted through the results of their efforts.

**4.2.1. Value Creation Enablers/Barriers.** The latter dimensions have a great impact on value creation in that they control to what extent value is created.

The **Technical Dimension** mostly regards aspects concerning the data itself. The format of the data is an essential aspect. Two of the eight Open Government Data Principles, in fact, regard the format in which data is made available to the public. They state that such data should be available in a *machine-processable* format which is *non-proprietary*. Such data would enable easier and unrestricted use of the data for value creation. Furthermore, if a format such as Resource Description Framework (RDF) is used, data ambiguity is reduced due to the format's expressivity, making the data more *understandable*. Additionally, the use of common schema aids to reduce interoperability issues caused by the large heterogeneity of the existing data. In order to encourage its use, data must also be easily *discoverable*. This is possible through the use of good quality metadata. The implementation of agreed-upon standards would aid reduce some, if not most, of the issues within this dimension.

The **Policy/Legal Dimension** regards issues with existing laws or policies that, through their ambiguity or due to being out-of-date, prevent data from being used to create value. On the other hand, well thought out policies encourage and enforce the creation of value, for example the publishing of data as Linked Data. Fortunately, there are growing efforts towards amending such laws and policies, but there is still a long way to go. Copyright and licensing of data can inhibit its unrestricted use. The incompatibility of licences, due to the data being created by various entities, further aggravates the issue. Privacy and data protection is another important aspect. Data providers need to strike a balance between making data freely available, whilst respecting the right to privacy.

The **Economic/Financial Dimension** is about aspects related to monetary issues and mainly concern the data provider and the data publisher roles. Being a relatively new concept, there might not be any budget allocation specifically for open government data efforts. In order to foster value creation, governmental entities cannot solely rely on existing data created in their day-to-day functionalities. Commitment is required, and hence also finances, for identifying and opening datasets with a high value creation potential.

The **Organisational Dimension** is concerned with the strategic aspects of the involved stakeholders. This dimension is especially relevant for governmental institutions. Considering there probably isn't an institution specifically in charge of open government data initiatives, data can get lost in the various hierarchical levels of a government. Adequate workflows need to be put in place for all the processes within a government data life cycle.

Finally, the **Social/Cultural Dimension** regards the feeling of the public towards open government data. While efforts are well under way to increasing

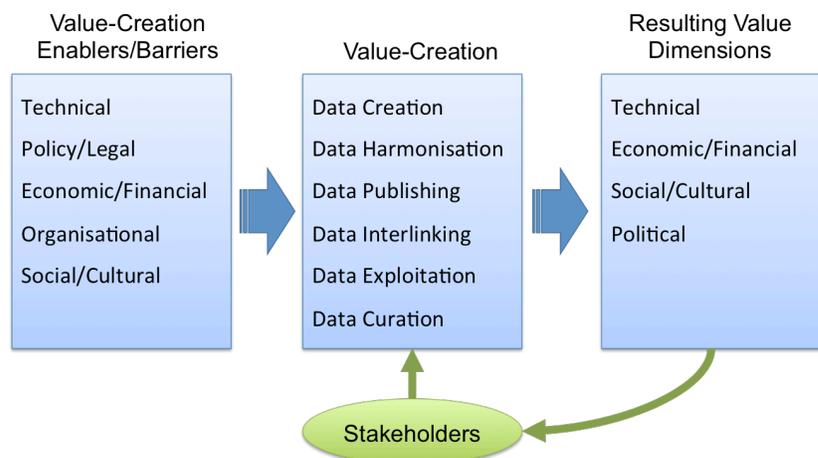


Figure 1: Dimensions impacting, and impacted by, Value Creation

awareness about the potential of open government data, not all stakeholders are ready to jump on the bandwagon. Workers within governmental entities might not understand the value of the data they are gathering/creating. This results in lack of motivation towards providing this data to the public. Stakeholders can also have misconceptions about the opening of public data. While open data can be considered as unfair competition for private entities (who invested to create their own data), public entities might consider the commercial appropriation of public open data unfair. The public also needs to be further informed on the advantages of public participation in creating value.

**4.2.2. Impacts of Value Creation.** As already discussed in the previous sections, value creation has a number of different dimensions of impact, which in turn affect the stakeholders. The term *public value* is used to define “what adds value to the public sphere” [15], where the public sphere is used to broadly indicate all of the following dimensions.

**Technical Value** is simply generated through the implementation of standards and the creation of services. As more value is created upon government data, the available data will be of better quality, and value creating services will increase.

**Economic Value** is defined as the worth of a good or service as determined by the market [16]. Value creation upon data enables the data itself to be considered as a product. Therefore, opening government data encourages its reuse in value creation, in turn stimulating competitiveness in the participating stakeholders and also encourages economic growth. For example, Mastodon C (a big data company) used open data to identify unnecessary spending in prescription medicine<sup>13</sup>. This will result in potentially huge savings from the National Health Service in the UK.

**Social\Cultural Value** is created first and foremost through the engagement of the public in open government data initiatives. The opening of data allows stakeholders to scrutinise the data and provide feedback on it. If the governmental entities exploit this feedback, it can result in improvement of citizen services. This sort of participation also increases citizen social control. Social value is also generated through creating innovative services based on open government data. For example, the Walkonomics Application<sup>14</sup> uses open data to enable

users to identify potential dangers in a street, such as fear of crime or road safety.

**Political Value** is created through the stimulation of democratic dialogue. Through participatory governance, citizens can gain a better insight as to how the governing process works. Stakeholders can possibly also participate in improving the policy-making process. Besides, the efforts of governmental entities to be more transparent and accountable increases citizens’ trust in their government.

## 5. Linked Data

In recent open government data initiatives, Linked Data practices are being followed by an increasing number of data publishers/providers such as [data.gov.uk](http://data.gov.uk) and [data.gov](http://data.gov). Yet, the use of Linked Data in open government initiatives is still quite low [17]. This might be due to a number of reasons, as the use of Linked Data is a process involving a high number of steps, design decisions and technologies [18]. We here investigate the advantages and benefits of using Linked Data practices in an open government data initiative.

The term *Linked Data* is used to refer to a set of best practices for publishing and connecting structured data on the Web [19]. Therefore, Linked Data is published on the Web in a machine-readable format, where its meaning is explicitly defined. It is also linked to and from external datasets. This has the potential of creating the *Web of Data* (also known as Semantic Web); a huge distributed dataset that aims to replace decentralized and isolated data sources [20]. The benefits of applying Linked Data principles to government data as covered in literature include [21] [22]:

- Simpler data access through a unified data model;
- Rich representation of data enabling the documentation of data semantics;
- Reuse of existing vocabularies;
- Use of URIs allow fine-grained referencing of any information;
- Related information is linked, allowing its unified access.

While significant efforts in literature cover advantages of using Linked Data (for example [23] [24] [17] [25]), there is no evident effort targeted towards the benefits of using Linked Data specifically in open government data value creation. We here therefore proceed to focus on the value creation techniques described in Section 4 and the benefits provided through the use of Linked Data. While still having similar barriers, enablers, and

<sup>13</sup> <http://theodi.org/news/prescription-savings-worth-millions-identified-odi-incubated-company> (Accessed 31/08/2015)

<sup>14</sup> <http://www.walkonomics.com/> (Accessed 31/08/2015)

impacts, as described in Section 4.2, the use of Linked data can result in different levels of impact, since the use of Linked Data techniques directly reduces some barriers of the technical level.

### 5.1. Linked Data as a Basis for Value Creation

Linked Data and Semantic Web technologies have the potential of solving many challenges in open government data, as well as possibly lowering the cost and complexity of developing government data-based applications.

Starting from the most common starting point of creating value, in general, data **generation** is the least impacted from the use of Linked Data since essentially the data is still being created. Data **collection** is similarly not impacted to a high level. Yet, the data creation process can be enhanced through the use of Linked Data. Consider the example of providing feedback based on a linked open dataset, for example budget data. The use of Linked Data enables feedback providers to have further context on the available data through the links. This would aid them in making a more informed decision. Furthermore, the high level of granularity of Linked Data has the potential of providing a deeper insight on the resource at hand. Also, since the data publisher is not necessarily the data provider, Linked Data will enable the access to primary data through the use of provenance information located within the metadata.

The Government Data Life Cycle Processes (Section 2) do not necessarily occur in the provided order. This means that the **linking** of the data can be done before data is published. This is actually ideal, as it would aid in the subsequent **harmonisation** of the data. If an entity is working with Linked Data, we can safely assume the data is represented in a semantically rich, machine-processable, format. Such a format has the greatest potential for reuse.

When data is published using Linked Data standards, it is made more **accessible**, discoverable, and ultimately, reusable. Since Linked Data standards demand the use of a semantic representation such as RDF, Linked Data is automatically more accessible than other standards such as CSV or PDF. RDF can be accessed in a variety of manners, such as RESTful APIs and public endpoints (queryable through SPARQL). This means that while Linked Data alternatives might require a consumer to download a data dump, the use of Linked Data enables the same consumer to access the specific subset of data he/she needs, and manipulate it easily. Additionally, each data resource is *dereferenceable*, i.e. the resource

URI can be resolved into a web document on the Web of Data. The datasets linked to the data in question give the consumer further context and insight. Through the use of metadata, a consumer can also check the provenance of the data, and ensure that it is a reliable source. Timeliness and versioning information can be obtained in the same manner.

In this section we are analysing how value creation is impacted through the use of Linked Data. This results in a recursion for the data **interlinking** process, as this value creation process requires the linking of data within itself. **Link discovery**, **data interlinking**, and **data integration**, will simply all be enhanced and facilitated through the pre-existence of links between, or within, datasets.

Data exploitation is possibly the value creation process that has the highest impact from the use of Linked Data. Through links between the datasets, and within datasets themselves, data consumers can obtain further insight and context on the data they are interested in. This increased information directly affects the data **interpretation** process, as the data consumer can interpret the data in a more informed manner, and generate knowledge from the existing information. Existing links similarly impact the data **analysis** value creation technique through simplifying the integration process where a number of datasets (or their subsets) are required. The creation of **visualisations** is also enhanced through the integration process. Visualising a dataset against a related dataset has the potential of providing the consumer with a new and different understanding of the data.

Linked Data enables (semi) automated **cleaning** and **repairing** of datasets through the use of **reasoners**. In this way, the violation of logical constraints is easily identified through the dataset's underlying schema.

The above benefits of using data for value creation are only a few, yet they collectively encourage and enhance the exploitation of open (government) data. Of course, this does not mean the implementation of a Linked Data approach does not have its challenges. Various efforts in literature, such as [24], provide discussions on the topic.

### 5.2. Use Case of Linked Open Government Data

[publicspending.net](http://publicspending.net) is a data portal created with the scope of demonstrating the power of economic Linked Open Data in analysing the situation with regards to market, competition conditions, and public policy, on a global scale. The creators of this portal consume and create value upon public spending data

of seven governments around the world. Results of the analysis led on the data are then published on the portal as tables, graphs, and statistics. The stakeholders here participate through all four value-creating roles described in Section 4.1 and execute value creation processes accordingly. Firstly, the public spending data is produced by the various governments (Data Producers). The data is then subject to pre-processing and data-preparation. Through the role of a Data Enhancer, the stakeholders here homogenise and link the data through the Public Spending Ontology<sup>15</sup> and other widely used vocabularies such as Dublin Core and FOAF. The resulting data in RDF is then published (Data Publisher and Data Facilitator Role) on the portal and is available both as bulk datasets and through a SPARQL endpoint. The Service Creator Role is then fulfilled through the application built on top of the data. The stakeholders use the internal data, along with other cross-referenced and external data, to provide a portal acting as an information point. This initiative results in impacts on four dimensions. Economic and technical value is created by providing a platform that gives access to Linked Open Data. This enables other stakeholders to reuse this data to create new products, add further value to it, and potentially market it. Social/Cultural Value is created by enabling stakeholders to explore and scrutinise spending data, giving them a good insight as to what is being spent, where, and by whom. Finally, political value is created since the governments involved are more transparent. This enhances accountability and prevents corruption. In turn, citizens' trust in the government increases.

## 6. Value Creation Assessment Framework

In order to assess the success of open government data initiatives, there exist a large number of assessment frameworks that aim to evaluate the effectiveness of an initiative in achieving its goals and objectives. Yet, rather than assessing the resulting impacts of such an initiative, real-life assessments, as documented in literature (See Section 2), mostly involve checking whether open government data initiatives are obeying existing policies and regulations [26]. Since the latter are not necessarily up to date with current technologies and approaches, this assessment is not really representative of the success of an initiative.

Consider the example of a government publishing the data in PDF. While the entity would be obeying existing laws requiring opening up such data, the use of PDF makes it pretty inconvenient for reuse and redistribution. In this case, one could argue that the open government initiative is not really a success. For this reason, a number of assessment frameworks analyse open government data initiatives based on different criteria [27][28]. The latter include nature of the data, citizen participation, and data openness. While there is still the problem that there is no agreed-upon assessment framework to evaluate open government initiatives, there is also limited literature (such as [29]) that focuses on **the impact of value creation**. Considering many resulting benefits of open government data depend on the creation of value (through the execution of one or more value creation techniques), we deem it essential to assess open government data initiatives on their potential for enabling value creation.

In Figure 2 we provide an overview of commonly evaluated aspects (in blue) of an open government data initiative extracted from our primary studies. These mostly concern implementation aspects, such as the format of the data, and how the initiative respects the requirements set from existing laws and policies.

The bottom part of the figure portrays the missing aspects (in red), i.e. those that are not considered when evaluating the success of an open government data initiative. We propose these aspects as part of a *Value Creation Assessment Framework*. The aim of this framework is to provide a guideline as to what aspects of an open government data initiative should be assessed to determine the potential of an open government data initiative to enable value creation, and thus exploit open government data to its highest potential. Here we briefly describe the aim of each aspect.

**Data Domain:** The domain of the data is important with regards to its potential impact. Some datasets have a higher reuse and value-creation potential.

**Data Usability:** This aspect regards the ease with which the published data can be reused.

**Data Ambiguity:** Data ambiguity is reduced when a representationally rich format (e.g. RDF) is used.

**Data Accuracy:** The extent to which data accurately represents the respective information.

**Data Completeness:** Data is complete when all required information is available, for the representation of the data in question.

**Usability of Format:** Formats such as CSV and RDF are much more usable than PDF. This is because they allow easier reuse of the represented data.

<sup>15</sup>

<https://docs.google.com/document/d/16fxFgtjRZC5AU00RiR0jdzbRFU73cBcOGI8ZZECwI6U/edit?pli=1> (Accessed 31/08/2015)

**Licence of Data:** Other than allowing for reasonable privacy, security, and privilege restrictions, data has the highest value creation potential if it is not subject to any limitations on its use due to copyright, patent, trademark or other regulations.

**Data Accessibility:** This aspect has two dimensions: the ease with which data is understood, and the ease with which data is discovered in a portal or catalogue.

**Data Discoverability:** This aspect depends on the metadata annotating the data in question, and enables stakeholders to more easily find data that is relevant to their needs. Data Discoverability is also affected by the search functions provided by a government portal or catalogue.

**Use of Standards:** Using agreed-upon standards throughout the life-cycle of government data encourages data reuse and integration.

**Variety of Access Options:** Providing various access options to the available data, such as APIs and SPARQL endpoints, encourages stakeholders to create value upon the data as they are able to access the data in their preferred manner.

**Data Timeliness:** Certain data might only be valuable if it is made openly available shortly after its creation.

**Data Diversity:** In the Linking value-creation process, the use of diverse datasets has the potential of releasing new insights or unforeseen results.

**Background Context:** The linking of datasets provides further context to the data in question, enabling stakeholders to have a deeper understanding.

**Consumer Participation:** The participation of

stakeholders in consuming the data is essential for value-creation. There is no use in having data made openly available if it is not exploited.

**Innovation:** Creating new products (data or otherwise) based on open government data is a direct impact of value-creation. Innovations include services and applications.

**Generation of New Data:** The value-creation techniques in the Data Exploitation Process can result in the generation of new data, such as analysis results or visualisations, that provide new interpretations or insight on the existing government data.

**Rate of Reuse:** The rate of reuse of open government data is directly indicative of the value-creation potential in the assessed initiative.

**Data Quality:** Data quality can be defined to be the *fitness for use* of the data in question. This depends on a number of other aspects, such as accuracy, ambiguity, usability, etc.

Since one of the major aims of open government initiatives is the release of social and commercial value, we deem that the proposed aspects are vital to determine the success of an initiative. Hence, these *value creation impact aspects* are used to assess the potential value that can be created through the use of the data product created as a result of each step within data value chain.

## 7. Concluding remarks

The main challenge in public value is that open data has no value in itself, yet it becomes valuable when it is used. In our information society, value creation processes have the

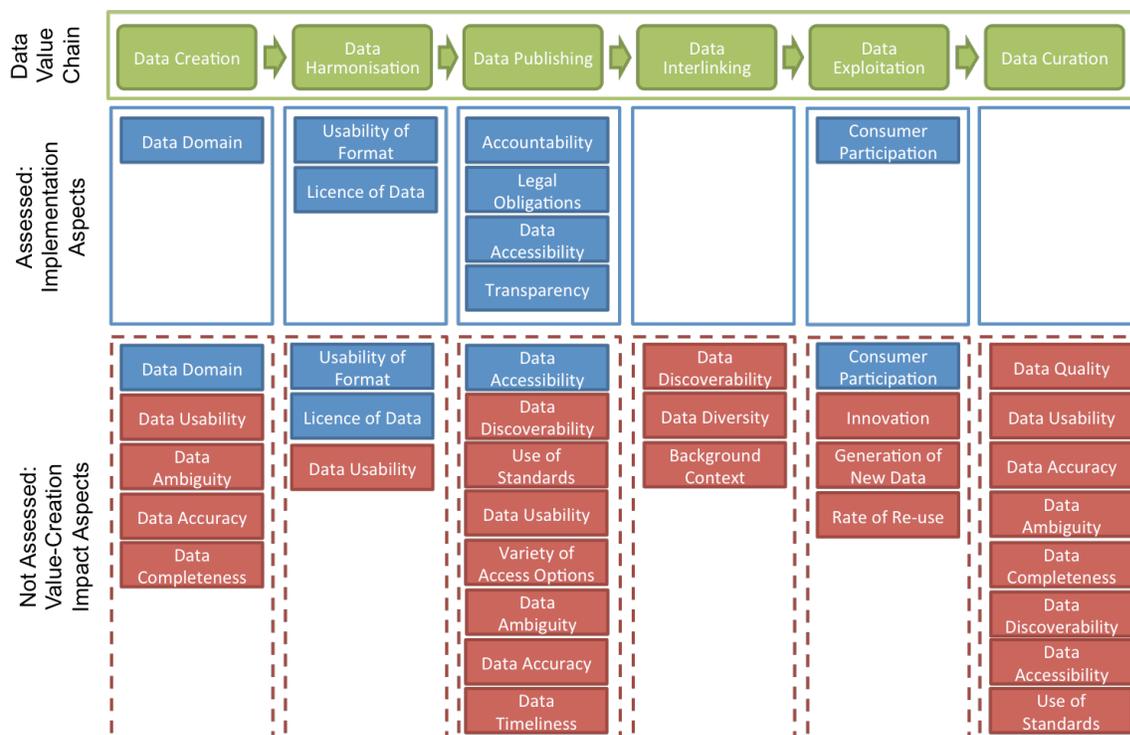


Figure 2: Aspects assessed in existing frameworks (blue), and proposed aspects for Value Creation Assessment Framework (Red)

potential of extracting the maximum value from data by building on its intelligent use. All stakeholders of value creation can participate through different roles, yet they have one common goal; that of creating a data product. Different dimensions impact the creation of such a product, namely technical, policy/legal, economic/financial, organisational, and cultural. Some of these dimensions are in turn also impacted by value creation. The use of Linked Data in creating value enhances the process, and also aids us to gradually proceed through various degrees of data products: starting with data, to information, and ultimately to knowledge. In order to truly assess the value creation process of an open government initiative, we propose an assessment framework that focuses on the potential impact achievable from a data product generated through a value creating process. As future work we intend to explore metrics that can be used to assess the suggested aspects within the framework. Step by step the vision of having open government data exploited to its full potential can be acquired.

## 8. References

- [1] L. Rojas, G. Bermudez, and J. Lovelle, "Open Data and Big Data: A Perspective from Colombia," in *Knowledge Management in Organizations*, vol. 185, 2014.
- [2] K. Janssen, "The influence of the PSI directive on open government data: An overview of recent developments," *Gov. Inf. Q.*, vol. 28, no. 4, Oct. 2011.
- [3] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, "Benefits, Adoption Barriers and Myths of Open Data and Open Government," *Inf. Syst. Manag.*, vol. 29, no. 4, 2012.
- [4] B. Ubaldi, "Open Government Data." OECD Publishing, 27-May-2013.
- [5] M. E. Porter, "Competitive Advantage: Creating and sustaining superior performance," New York, vol. 15.
- [6] D. Crié and A. Micheaux, "From customer data to value: What is lacking in the information chain?" *J. Database Mark. Cust. Strateg. Manag.*, vol. 13, no. 4, 2006.
- [7] A. Latif, A. Us Saeed, P. Hoefler, A. Stocker, and C. Wagner, "The Linked Data Value Chain: A Lightweight Model for Business Engineers," in *Proceedings of International Conference on Semantic Systems*, 2009.
- [8] C. C. Lee and J. Yang, "Knowledge value chain," *J. Manag. Dev.*, vol. 19, no. 9, 2000.
- [9] "Elements of a data value chain strategy | Digital Agenda for Europe | European Commission." [Online]. <https://ec.europa.eu/digital-agenda/en/news/elements-data-value-chain-strategy>. [Accessed: 22-04-2015].
- [10] Ping Wang and Haiying Hua, "A model of government information value-added exploitation based on cloud computing," in *2011 International Conference on Business Management and Electronic Information*, 2011, vol. 2.
- [11] J. Attard, F. Orlandi, S. Scerri, and S. Auer, "A Systematic Review of Open Government Data Initiatives," *Gov. Inf. Q.*, 2015, doi:10.1016/j.giq.2015.07.006.
- [12] K. J. Reiche and E. Höfig, "Implementation of Metadata Quality Metrics and Application on Public Government Data," in *COMPSAC Workshops*, 2013.
- [13] A. Zuiderwijk and M. Janssen, "Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View," in *Open Government*, vol. 4, Springer New York, 2014.
- [14] P. Conradie and S. Choenni, "Exploring Process Barriers to Release Public Sector Information in Local Government," *6th International Conference on Theory and Practice of Electronic Governance*, 2012.
- [15] J. Benington, "From private choice to public value," *Public value Theory Pract.*, 2011.
- [16] T. Jetzek, M. Avital, and N. Bjørn-Andersen, "Generating Value from Open Government Data," *International Conference on Information Systems*, 2013.
- [17] N. Shadbolt and K. O'Hara, "Linked Data in Government," *Internet Comput.*, vol. 17, no. 4, 2013.
- [18] B. Villazón-Terrazas, L. Vilches, O. Corcho, and A. Gómez-Pérez, "Methodological Guidelines for Publishing Government Linked Data," in *Linking Government Data*, 2011.
- [19] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data - The Story So Far," *Int. J. Semant. Web Inf. Syst.*, 2009.
- [20] T. Heath, "How Will We Interact with the Web of Data?," *IEEE Internet Comput.*, vol. 12, no. 5, Sep. 2008.
- [21] R. Cyganiak, F. Maali, and V. Peristeras, "Self-service Linked Government Data with Dcat and Gridworks," in *Proceedings of the 6th International Conference on Semantic Systems*, 2010.
- [22] E. Kalampokis, E. Tambouris, and K. Tarabanis, "A Classification Scheme for Open Government Data: Towards Linking Decentralised Data," *Int. J. Web Eng. Technol.*, vol. 6, no. 3, Jun. 2011.
- [23] D. DiFranzo, A. Graves, J. S. Erickson, L. Ding, J. Michaelis, T. Lebo, E. Patton, G. T. Williams, X. Li, J. G. Zheng, and others, "The web is my back-end: Creating mashups with linked open government data," *Link. Gov. Data*, 2011.
- [24] N. Shadbolt, K. O'Hara, T. Berners-Lee, N. Gibbins, H. Glaser, W. Hall, and M. c. Schraefel, "Linked open government data: lessons from Data.gov.uk," *IEEE Intelligent Systems*. 01-May-2012.
- [25] T. Heath and C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*. 2011.
- [26] R. Sandoval-Almazan and J. R. Gil-Garcia, "Towards an Evaluation Model for Open Government: A Preliminary Proposal," in *Electronic Government*, vol. 8653, M. Janssen, H. Scholl, M. Wimmer, and F. Bannister, Eds. Springer Berlin Heidelberg, 2014.
- [27] S. Bogdanović-Dinić, N. Veljković, and L. Stoimenov, "How Open Are Public Government Data? An Assessment of Seven Open Data Portals," in *Measuring E-government Efficiency*, vol. 5, Springer New York, 2014.
- [28] R. P. Lourenco, "Open Government Portals Assessment: A Transparency for Accountability Perspective." in *EGOV*, 2013, vol. 8074.
- [29] I. Susha, a. Zuiderwijk, M. Janssen, and a. Gronlund, "Benchmarks for Evaluating the Progress of Open Data Adoption: Usage, Limitations, and Lessons Learned," *Soc. Sci. Comput. Rev.*, 2014.