Exploiting the Value of Data through Data Value Networks

Judie Attard
University of Bonn, Germany
attard@iai.uni-bonn.de

Fabrizio Orlandi
University of Bonn, Germany
orlandi@iai.uni-bonn.de

Sören Auer
University of Bonn &
Fraunhofer IAIS, Germany
auer@cs.uni-bonn.de

ABSTRACT

Open data is increasingly permeating into all dimensions of our society and has become an indispensable commodity that serves as a basis for many products and services. Governments are generating a huge amount of data spanning different dimensions. This dataification shows the paramount need to identify the means and methods in which the value of data and knowledge can be exploited. While not restricted to the government domain, this dataification is certainly relevant in a government context, particularly due to the large volume of data generated by public institutions. In this paper we identify the various activities and roles within a data value chain, and hence proceed to provide our own definition of a *Data Value Network*. We specifically cater for *non-tangible data products* and characterise three dimensions which play a vital role within the Data Value Network. We also propose a *Demand and Supply Distribution Model* with the aim of providing insight on how an entity can participate in the global data market by producing a data product, as well as a concrete implementation through the *Demand and Supply as a Service*. Through our contributions we therefore project our vision of enhancing the process of open (government) data exploitation and innovation, with the aim of achieving the highest possible impact.

Keywords

value creation, data value network, data value chain, data demand, data supply, exploitation, innovation, impacts, open data

1. INTRODUCTION

Our society is increasingly relying on data products that are the basis for many products and services. Data is becoming more and more relevant as a commodity, even in more traditional sectors such as health, transport, and retail. Examples include government data portals[1], reviews, feedback, and product suggestion on e-commerce websites, weather emergencies forecast[2], patient monitoring[3], citizen participation and decision-making[4], etc. *Data Value Chains* were therefore introduced in order to reflect this dataification [5]. As opposed to traditional value chains for tangible products, data value chains are based on the rationale of extracting the highest possible value from data by modifying, processing and re-using it. The exploitation of this data with added value has the potential to feed a chain of innovative information products and services, making the data value chain the centre of the knowledge economy.

The value chain model describes value-adding activities that connect an industry's supply side, such as raw materials and production processes, to its demand side, such as sales and marketing. The value chain model has been used to analyse and assess the linked activities carried out within traditional industries in order to identify where, within these activities, value is created. This was done with the aim of identifying what activities are the source of competitive advantage within these industries. As successful as the value chain concept was to achieve this aim, during these last years products and services are becoming increasingly digital, and exist in a more non-tangible dimension [16]. In addition, the traditional value chain model does not consider when information is used as a source of value in itself [20]. Thus, the original value chain concept is becoming an inappropriate method with which to identify value sources in today's industries that produce non-tangible products [16].

With the aim of focusing on technologies that enhance data exploitation, in this paper we analyse various implementations of data value chains and build upon previous definitions with the aim of providing the best characterisation. We are hence trying to answer to question: *What is the best representation of the existing processes of creating value upon a data product?* Using a methodology similar to defining a life-cycle, we improve upon existing representations and hence define the *Data Value Network*. Our definition caters specifically for non-tangible data products, thus we focus on the aspects specific to data that differ from the definitions of value chains in literature. Based upon our definition, we focus on the creation and consumption of open data and propose a *Demand and Supply Distribution Model*. We provide this model as an attempt to provide the required knowledge for an entity to start a Data Value Network on a particular data product. As a concrete implementation, we also provide a prototype of the proposed model through a *Demand and Supply as a Service* application. Through the Data Value Network and the Demand and Supply Distribution Model we propose, we aim to highlight the economic potential of a data product. Such an *Economic Data Ecosystem* influences the technical, societal/political, and economic dimensions. We envision this data ecosystem to result in the development of the Web of Data into not just being a technological solution, but also becoming

[1] https://open-data.europa.eu/en/data/
[2] http://centrodeoperacoes.rio/
[3] http://www.immunizeindia.org/
[4] https://www.fixmystreet.com/

a societal change driven by data-based innovation.

While the content of this paper is not restricted to the government domain, it is very important to note that the contributions in this paper are also very relevant in this context. Governments are one of the largest producers and collectors of data in many different domains [1, 7]. This means that governmental entities not only hold a huge value potential in their hands, but if this data is made openly available to the public, a much larger number of stakeholders from the general society can exploit this data. For this reason, in this paper we do not restrict our contributions to solely address governmental entities, but we address all stakeholders, including governments, who would like to exploit data and create value.

We continue this paper by giving an overview of existing literature in Section 2, then providing our motivation and the definition of the Data Value Network in Sections 3 and 4 respectively. We describe the different roles that actors can undertake to participate in the Data Value Network in Section 5, whilst in Section 6 we describe the dimensions that affect, or are affected by, the Data Value Network. In Section 7 we propose the Demand and Supply Distribution Model, and provide a proof of concept through Demand and Supply as a Service. We finally provide our concluding remarks in Section 8.

2. BACKGROUND AND RELATED WORK

The process of *Value Creation* is a somewhat subjective concept that depends on a consumer's perception on the usefulness of the product, the amount they are willing to pay for it, as well as the actual amount spent in such a transaction [14]. Many works in literature focus on various aspects in the context of value creation, such as how to best exploit value creation to achieve economic benefit [2], discussions on value creation within specific domains, such as mobile commerce [3], open data [22], and e-government [24], or how to achieve competitive advantage [9, 18, 19].

The term *Value Chain* was first introduced by Porter [18] in 1985. Porter defines a value chain to be the strategically relevant interdependent activities undertaken by a firm in order to achieve its goal. The activities are physically and technologically distinct activities that are the building blocks by which a firm creates a product valuable to its buyers. Similarly, Kaplinsky [9] defines a value chain to describe the full range of activities that are required to bring a product or service from conception, through the different phases of production, delivery to final consumers, and final disposal after use. The value chain can be considered as a tool that enables the analysis of the interactions between the different activities in order to identify the sources for competitive advantage, or, in other words, how and where the value is created. The activities within a value chain can be classified into five categories, namely; Inbound Logistics, Operations, Outbound Logistics, Marketing and Sales, and Service [18].

Building upon Porter's definition, Lee and Yang [12] define the *Knowledge Value Chain*. Their definition differs to Porter's in that the end product is not tangible, and they define a value chain for knowledge, including the knowledge infrastructure, the process of knowledge management, and the interaction between the required components that result in knowledge performance. Knowledge, a step further than information, is data organised in meaningful patterns. The process of reading, understanding, interpreting, and applying information to a specific purpose, transforms information into knowledge. This means that for an entity that is unable to understand knowledge, the knowledge is in fact still only information. This is the *data literacy* problem, where any effort invested in knowledge generation is lost if the target consumer is unable to actually understand the provided knowledge [23]. Similar to Porter, Lee and Yang classify the activities within the knowledge value chain in five categories, namely knowledge acquisition, knowledge innovation, knowledge protection, knowledge integration, and knowledge dissemination. While essentially similar to Porter's definition, these categories are different in that they classify activities specifically involving a non-tangible information product. In [4], Crié and Micheaux provide us with a more generic value chain than Lee and Yang, including raw data in their definition. Within their paper, the authors aim to highlight any issues within the value chain, to provide an overview of the current progress, and also to encourage entities to view the benefits of participating within the data value chain.

Peppard and Rylander [16] also discuss a value chain that is more suited where the product in question is non-tangible, however they focus on service-provision rather than data. The authors introduce the concept of *Network Value*, where value is created by a combination of actors within the network. In contrast to the earlier definition of a value chain, network value does not necessarily follow a linear model, and accounts for the various interconnected actors that work together to *co-produce* value. While these actors or entities should be able to function independently, they operate together in a framework of common principles. This means that an action by a single entity can influence other entities within the network, or otherwise require further actions from them in order to achieve the final product. Morgan et al. [14] provide a similar discussion on the co-production of value through open-source software.

In line with more recent popular themes, Miller and Mork [13] and Latif et al. [11] focus on big data and Linked Data respectively. Miller and Mork discuss the data value chain concerning all required actions in aggregating heterogeneous data in an organised manner and creating value (information/knowledge) that can influence decision making. The authors divide their data value chain in three main categories, namely data discovery, integration, and exploitation. In contrast, Latif et al. propose the *Linked Data Value Chain*. Motivated by the still limited commercial adoption of the Semantic Web, the authors aim to drive the Semantic Web and the use of Linked Data closer to commercial entities. The authors discuss the entities participating in the linked data value chain, their assigned Linked Data roles, as well as the types of data processed within the chain.

In the following section we describe how the value chains in the above literature fail to adequately portray value creation on a data product in our new data-centric society. We hence define the Data Value Network, and propose it as a solution.

3. WHY DO WE NEED TO DEFINE A NEW VALUE CHAIN?

The value chain was initially used to identify what activities led by a manufacturer gave additional value to the product being created, therefore resulting in a *competitive advantage* [18]. Nowadays, in a digital data-centric world, the cost of processing data has drastically decreased, and the access to data from multiple sources such as networks, sensors, and the Internet, has skyrocketed the availability of data. Coupled with the dramatic decrease in the cost of data storage, this is enabling huge datasets to be generated or captured, stored, and processed. Alas, value chains such as the one defined by Porter [18] do not cater for data-specific activities or for activities where information is used as a source of value [20]. Lee and Yang's definition [12], whilst it is focused on a data product, only considers value creation to be based on the creation of knowledge. Crié and Micheaux [4] on the other hand do consider raw data, but still focus on a 'chain' structure which does not reflect

the flexible value creation that is really possible on data products. Peppard and Rylander [16], whilst they improve on the former by defining Network Value, do not focus on value creation on a data product. Latif et al. [11] stick to a strict chain structure and only consider value creation to be the evolution of raw data to Linked Data and human-readable data. Miller and Mork [13] provide the most in-depth definition of value creation upon a data product, however they still stick to a chain structure.

Through these definitions of (data) value chains, we identified the lack of literature that discusses the actual processes used to create this value. We also distinguish the need to represent the fluidity of value creation which is not easily characterised using a linear structure. Moreover, rather than limiting our discussion to the economic impact, we identify three impact dimensions that influence, or are influenced by, value creation.

4. THE DATA VALUE NETWORK

After considering existing value chain definitions, and identifying different real-life data value chains and the contained activities and roles, we define a **Data Value Network**. Similarly to a life-cycle, the Data Value Network maps the ongoing processes through which value is created upon a data product. We included common activities executed on data products, where the final aim is usually the consumption of the data product, and characterise them according to three dimensions. Due to the differing order of executing the relevant activities, a star network with the data product as its central node was deemed to be the best way to represent the interactive and fluid nature of adding value to data products. We define a *Data Value Network* to be:

> *A set of independent activities having the aim of creating value upon data in order to exploit it as a product*

where different **actors** (e.g. data producers, data consumers) can participate by executing one or more **activities** (e.g. Data Discovery, Data Exploitation), and each activity can consist of a number of **actions** or value creation techniques, (e.g. Gathering, Visualisation, Service Creation). In turn, each action can consist of one or more **data value chains**, since they might need a series of processes to be executed in order (e.g. visualisation requires identifying the data to visualise, then deciding on a visualisation method, then rendering the visualisation).

The Data Value Network thus has the following features which characterise it and distinguish it from the value chain definitions covered in Section 2:

- **Non-Tangible Data Product** - The aim of the data value chain is to process data in order to make it usable for a generic or specific use, enabling the data itself to be considered as a product. There are numerous differences between a tangible product and a data product within a value chain. The most obvious being the *non-perishable nature* of data [20]. While it can become outdated, its exploitation or use will not consume it. Data can be re-used over and over, even for different purposes than the one it was originally planned for. This changes the competitive dynamics of the entities involved in the data value chain.

- **Non-Sequential** - The data value chain does not necessarily follow a sequential structure, rather, any activity can follow, or precede, any other activity. Activities can be executed in tandem, and other activities can be skipped or repeated. Furthermore, all activities result in a data product except for *Data Exploitation*. In this case, the outcome of the activity might or might not result in a data product, depending on how the product itself is consumed.

- **Multiple Actors** - One or more actors can participate to co-produce value within an activity. While in [18], the author describes a value chain where all the roles are executed by one entity, in a data value chain we need to cater for different actors, who can also *participate in one or more actions*. The actors can also collaborate in order to *co-produce* value.

- **Nested Value Chains** - Each activity can be broken down into further, more specialised actions, each of which can be a data value chain within itself and could possible cater for different scopes. For example the *Data Discovery* activity can include both the *procurement* of data from a different entity, or the *generation* of new data specifically for the required purpose. The latter two actions will then include different processes and approaches, thus having different data value chains.

- **Recurring Value Chain** - Rather than the value chain ending with the consumption of the product, the data value chain can recur as long as the data in question is still relevant. Thus, we can consider it to be similar to a life-cycle, where, the output of an activity can be the input to any other activity.

- **Independent Activities**: Unlike the value chain defined by Porter [18] the value creating processes are not interconnected, and can exist *independently*.

We portray our definition of the Data Value Network in Figure 1, where the activities and related roles all belong to a **data-centric** domain. Here the data product is central to the Data Value Network, as data, in whatever state it is in, can be considered to be a product and consumed. There are five activities, namely *Data Discovery*, *Data Curation*, *Data Interpretation*, *Data Distribution*, and *Data Exploitation*. These activities are the data centric counterparts to the activities defined by Porter [18] (cf. Section 2), and form a network around the data product, as they can be executed in any order by different actors. Similar to the Network Value concept introduced in [16], the Data Value Network does not necessarily follow a linear model, and some of the activities can be repeated or skipped, and can precede or follow any other activity. This network model also emphasises that there is no “end" to the Data Value Network. Rather, it can recur as long as the data product is still relevant. The five activities in the network can be executed on a data product by one or more actors. For example, two actors might participate in interpreting the data (co-production of value), while five actors might exploit it in different ways. Each activity is also made up of a number of actions. While not exhaustive the listed actions are, according to existing literature, the most common and generic processes that can be executed on a data product.

Figure 2 shows how a data product can evolve over time under the execution of different activities led out by different actors. It is important to note that each data product in the diagram forms the core node in a new instance of a Data Value Network, since an activity on a data product could result in a new version of the data product (e.g. by organising a dataset), or even a new data product (e.g. through knowledge extraction or merging). Then, the evolution of a data product will eventually result in a ‘branching’ out of various value-added versions, similar to the branching out of D_2 to two different versions; D_3 and D_4.

In order to be used within a Data Value Network, data does not require any specific characteristics. It can be raw, previously-processed, machine or human readable, and it can also be data re-

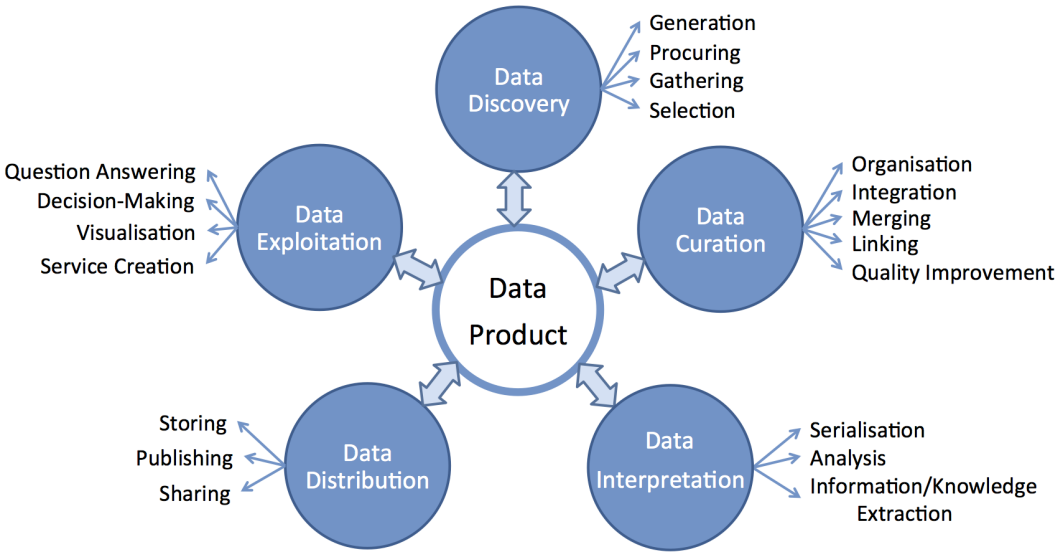

Figure 1: The Data Value Network (Activities and Actions)

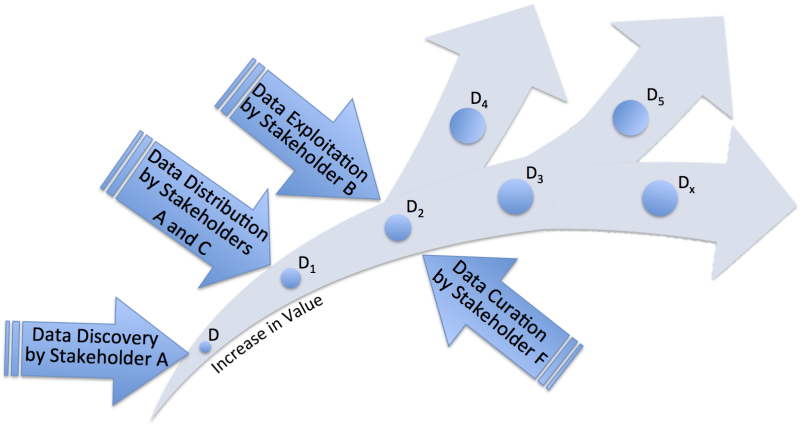

Figure 2: Tree structure of an evolving data product D, with interaction from different actors

garding any domain, such as statistical, financial, geographical, demographic, societal, etc. Unfortunately, while data is abundantly available, there are also a number of obstacles which hinder its exploitation. Data is of a *heterogeneous nature*, with a plethora of formats, models, and schemas. This makes it very challenging to integrate, even if it belongs to the same domain. Furthermore, data is usually available from a *variety of sources* which can also be in isolated data repositories, or otherwise privately-owned spaces. This of course restricts the openness of the data and therefore also its re-use.

Data quality is an aspect of data that has a major effect on the value-creation potential of the data itself. Due to the subjectiveness of data quality [15], it is very challenging to assess the quality of data, which can depend on a myriad of characteristics. Furthermore, there is no single agreed-upon definition of quality [10] due to the cross-disciplinary nature of the concept. However, data quality is commonly perceived to capture *fitness for use* [8], which in turn is a multi-dimensional concept that has both subjective perceptions and objective measurements based on the data in question [17].

5. ACTORS' ROLES IN A DATA VALUE NETWORK

In this section we identify the various roles that an actor can adopt to participate within the Data Value Network. An entity, be it a government entity, a public entity, a citizen, a company, or an enterprise, can participate as an actor within the Data Value Network through one or more roles. This flexibility between the undertaken roles is due to the adaptive nature of the data, where value can be added in a number of different ways, as discussed previously. Through the actors' interaction or collaboration, the outcome of a Data Value Network is a data product of any kind or shape, that can be processed, integrated, maintained, shared and published in order to add value to it.

Data Producer: The data producer has the role of obtaining or creating the data. This can be achieved through a number of different actions. The role of a data producer can be considered as one of the most important roles within the Data Value Network, as any activity or action in the network depends on the available data. If the data producer does not obtain relevant data for the use case at hand, then the Data Value Network will not reach its target to obtain the intended value out of this data.

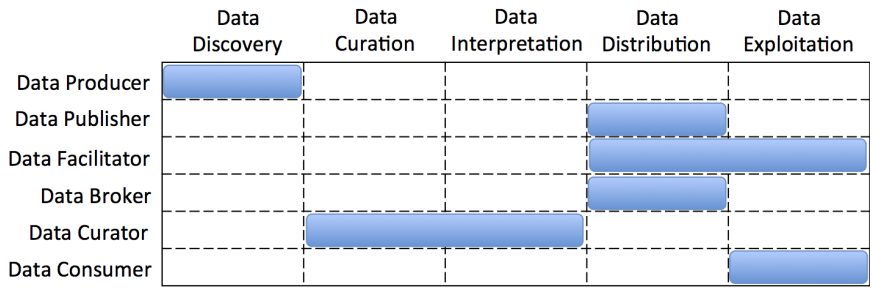

Figure 3: The Activities in which an Actor can participate through each Role

Data Publisher: The data publisher role is also essential in the value network. This role involves the distribution of the data product. The distribution process enables other stakeholders to discover potentially useful data products.

Data Facilitator: This role involves entities that, in some way or another, aid the other stakeholders in using, re-using, or exploiting, data products. This can be done through the provision of software, services, or other technologies. For example, the creator of a government data portal is facilitating the use and re-use of government data from other stakeholders by organising heterogeneous government data in a single location.

Data Broker: A data broker has the role of acting as a 'matchmaker' between the requirements of data consumers, and the data products of data producers/publishers, hence encouraging data reuse throughout the Data Value Network. Since the data is here considered to be a product, the data broker has the role of linking the roles of the data producers/publishers and the data consumers, enabling the balancing of the supply and demand. Whilst similar to the role of a data publisher, a data broker does not only passively share the data, but actively provides the data to a relevant consumer.

Data Curator: The role of a data curator is to modify or enhance the data in a manner that it is more usable for the target aim. A data curator can influence the outcome of the Data Value Network by adapting the data so that its highest value potential can be exploited.

Data Consumer: The data consumer role can be considered the final role in the Data Value Network, however, this is not always the case. For example, when a consumer gives feedback, the feedback can in turn be used as part of the data product. In the case of crowdsourcing, the data consumer also has the role of a curator, blurring the lines between both roles.

6. IMPACTS OF THE DATA VALUE NETWORK

We here characterise the Data Value Network according to three impact dimensions, namely **technical**, **economic**, and **societal/political**. As opposed to a number of related works in literature that focus solely on the economic dimension, we here provide a more holistic picture of Data Value Networks as a basis for the information-based society. We inspire our definition of these dimensions from the *Triple Base Line*; a framework aiming to measure the degree of sustainable growth through the incorporation of three dimensions of performance, namely social, environmental, and financial [6]. Considering the non-tantangibile nature of a data product, we hence consider the technical dimension, rather the environmental. The technical, economic, and societal/political dimensions directly influence, or are directly influenced, by creating a Data Value Network infrastructure that connects the various roles within any data-based domain. These dimensions also affect the direct and indirect benefits that can be extracted from re-using open data; namely market size, number of jobs created, cost savings, and efficiency gains (including saving lives, saving time, and environmental benefits) [25].

6.1 Technical Dimension

This dimension directly concerns the Data Value Network and its use as a tool to identify value creation in the activities or processes forming the network. Data activities in a Data Value Network all have the purpose of adding value to data. We can consider 'adding value' to be equivalent to 'making the data more usable, or more fit for use'. So, for example, while data in PDF format is easily human-readable, it's conversion to RDF would make it more usable where the use case requires data to be machine readable. The opposite can also stand true. Therefore, we here discuss, in different levels of abstraction, the activities and actions that add value to data.

Data Discovery: Data discovery is the process of obtaining data. As shown in Figure 1, this activity can also follow another activity. For example, in order to add value to a dataset about weather, the data discovery activity is executed with the aim of finding geolocation data to link it with the weather data. Sources of data can be as varied as sensor data, the Internet, private companies, governmental entities, and social media, amongst others. Various actions can be implemented to discover data, including:

- *Data Generation*: where data is created for the specific use case on hand;
- *Data Procuring*: where a negotiation is performed with another entity in order to obtain their data;
- *Data Gathering*: where data is aggregated from different entities/locations;
- *Data Selection*: where a subset of available data is selected.

Data Curation: This is a very generic activity that can encompass a large number of different actions, all of which modify the data in some way or another. This activity can recur numerous times, until the required data product is obtained. Data can be modified through:

- *Data Organisation*: where data is structured in such a way that it is more understandable, or that the data follows some pattern, for example government budget data is organised by year;
- *Data Integration*: where new data is added to available data, for example the integration of weather conditions data to accident information can be done by an insurance company to check the legitimacy of a claim, or adding user feedback to product data;

- *Data Merging*: where different datasets are merged in order to obtain further information, for example population data merged with geographical data in order to obtain population density;
- *Data Linking*: where data is linked to other existing data in order to provide context, for example geographic data linked to textual descriptions about the locations in question;
- *Quality Improvement*: where the quality of data is assessed, and if necessary improved, cleaned, or repaired. For example ensuring the data is timely and trustworthy, removing duplicate data, ensuring the data is consistent and complete, and adding provenance data.

Data Interpretation: This activity involves presenting the data in a different manner, in order for it to be more understandable. First and foremost, this can include extracting information or knowledge from raw data, but it can also include data serialisation (lifting or lowering data into different formats e.g. CSV to RDF or JSON to PDF), and also implementing analysis techniques in order to identify patterns. The latter patterns can then aid actors in the Data Value Network in actions such as decision-making.

Data Distribution: Data distribution is the activity involving making the data available as a product. This means that other entities can search for and discover this data. Yet again, this activity can be executed in different ways, including storing and publishing data, advertising it, and sharing it.

Data Exploitation: This activity can be considered as the final goal of the data value chain. It involves consuming the data as a product. Visualisation can be considered as an example of *passive* exploitation, where an actor consumes the data as information or knowledge. An example of a more *active* consumption of the data can be the use of data to influence decision-making, for example, a government might consider citizens' feedback before taking a decision. Question answering and service creation are other examples of active consumption of data. In the former data is collected and analysed in order to solve a specific question, whilst service creation is the provision of a service through the use of existing data, for example a mobile public transport timetable application.

6.2 Economic Dimension

This relatively new approach of trying to exploit data for its maximum value can directly affect the economy through income growth, particularly so due to the growing participation of data in the global economy. Any actor, participating through any role within the Data Value Network, can enter the data market in a manner that would provide for income growth. Using an Open Data Maturity Model, the authors of [25] have estimated the (total) market size of open data in the European Union to be between 193 Bn and 209 Bn Euro for 2016. A key capability to participate in the data market is the capacity to innovate, and to ensure continuous improvement in product and process development [9], as well as identifying the correct competitive scope [18]. The Data Value Network also increases competition. Having a global market where a number of actors provide a similar data product, consumers can identify the best product for their use case. Producers can compete with each other by attempting to provide the best data product for their target consumers and build up a reputation. This reputation, if good, would also help actors in obtaining new consumers. This competition can be considered as a small economic ecosystem that depends on the Data Value Network as is undertaken by each actor. Through *Value Chain Analysis*; the use of the value chain as a tool to identify sources of competitive advantage, the relevant entities can decompose the entire Data Value Network (or chain) into the economic structure of the various activities within the network. This aids actors to efficiently allocate resources and efforts along the value network. This is especially relevant within a Data Value Network, as a number of actors participate, rather than a single actor executing all activities. Therefore, actors can focus on their core competences and *specialise* in the activities with the highest value adding capabilities [9].

Due to the potential of data to be used over and over (until it remains relevant), the economic impact of adding value to it and using it as a product is different when compared to the more traditional product manufacturing. First and foremost this is evident in the re-use of data in another context, or domain, that it was originally envisaged for. For example, e-commerce businesses use historic purchase data to identify patterns and suggest items to users. Moreover, the data can be processed repetitively in order to make it more usable for a specific use case, for example, by changing its format, removing irrelevant data, or linking it with other data. Data can also be interpreted and made human-readable by extracting knowledge from it. For example, in the case of government data, this data processing would enable all citizens to exploit the data, and potentially even give their feedback. In turn, this feedback could be added value that the governmental entity can exploit.

In Table 1 we show an excerpt of a knowledge base where we indicate various datasets and related relevant information, such as the domain of the data product, and the way it was consumed. The purpose of this knowledge base is to portray the development and exploitation of a number of data products with an economic motivation. As can be seen, there is a large variety in the domains of the data product. Any type of data can be used within a Data Value Network, besides also being re-used in use cases other then the one originally envisaged, as is particularly evident in the fifth entry in the table.

6.3 Societal/Political Dimension

The establishment of Data Value Networks and the better exploitation of data to create value can have major implications on societal and political aspects. Apart from potentially improving the economy, as discussed in Section 6.2, it can also improve the standard of living. The simple collection and analysis of data can be used to identify and target problems in countries' standard of living. For example, the World Health Organisation (WHO) used data[16] about citizens' lifestyle and quality of life to create a global action plan to prevent and control non-communicable diseases[17]. This is just one effort from many that the WHO undertakes after analysing such data, resulting in a concrete action plan to eradicate substantial problems.

Another important impact on society is the increase in *citizen*

[5] https://musicbrainz.org/
[6] http://www.bbc.com/
[7] http://www.bbc.co.uk/music
[8] http://www.europeana.eu/
[9] https://www.gov.uk/government/organisations/hm-treasury
[10] http://data.gov.uk/dataset/coins
[11] http://blog.safecast.org/
[12] http://fukushima-radioactivity.jp/world-mapsearch.php
[13] https://www.progressive.com/auto/snapshot/
[14] http://www.ons.gov.uk/ons/index.html
[15] http://www.walkonomics.com/
[16] http://apps.who.int/gho/data/?theme=home
[17] http://www.who.int/nmh/publications/ncd-action-plan/en/

	Publisher	Domain	Access Method/ Data Format	Consumer	Aim	Short Description
1	MusicBrainz[5]	Music Data	API	BBC[6]	News Enhancement	Music data from MusicBrainz is linked to news data on the BBC music site[7] in order to provide discographies and track listings across about 700 artist pages.
2	Europeana[8]	Cultural Data	API, SPARQL Endpoint, Datadump	Historiana.eu	Educational Portal	Data from the Europeana database is aggregated in the Historiana.eu portal in order to act as an educational website.
3	Her Majesty's Treasury[9]	Government Expenditure Data	Sparql Endpoint, CSV	wheredoesmymoneygo.org	Informative Portal	Where Does My Money Go? aims to promote transparency and citizen engagement through the analysis and visualisation of information about UK public spending, extracted from the COINS[10] dataset.
4	Safecast[11]	Environmental Data	API, CSV	Fukushima Government	Radiation Awareness	The Fukushima Government used radiation measurements data from Safecast in order to populate maps, showing the radiation levels in different locations[12].
5	Vehicle	Vehicular Data	n/a	Progressive[13]	Insurance Services	Progressive aggregates sensor data from a vehicle in order to identify a person's driving style, and then adapts insurance policies according to how safe the person drives.
6	Office for National Statistics[14]	Crime	CSV, Spreadsheets	Walkonomics[15]	Environmental Safety	Walkonomics use crime statistics provided by the Office for National Statistics in order to develop an app. Using this app, a person can check the, "walkability" of a given street based upon various categories, such as fear of crime, road safety, and pavement quality.

Table 1: Demand and Supply Knowledge Base Excerpt

social control. Through the distribution of data, especially governmental data, citizens are enabled to generate knowledge and actively participate in governance processes, such as decision-taking and policy-making, rather than voting in an election every few years. For example, if budget data is published as open data, citizens can scrutinise the data and give feedback, which the government can then use in the decision-making processes. In this manner, citizens also inform themselves, and as a result are able to make better decisions [21]. This transparency also proves the legitimacy of government actions, and potentially corruption events can also be detected.

The use of data to create innovative services can aid relevant entities in consuming such data. For example, the use of public school locations can help a new family identify the best school for their child in their neighbourhood. Another example is the use of public transport timetables in mobile applications, which help commuters plan their trips on the go. The development of such services based on adding value to data and generating knowledge also results in in the creation of jobs.

Data Value Networks can also help to preserve and better showcase the cultural identity and diversity of a region. A perfect example in this regard are digital museum, archive and library aggregators, who collect metadata about millions of cultural heritage artefacts. Here, museums, libraries and archives are data providers to a hierarchic network of geographic or thematic data aggregator nodes. The German Digital library[18], for example, aggregates semantic metadata from more than 2,000 memory institutions and feeds this data into the European cultural heritage portal Europeana[19], as similarly do many other national aggregators. As a result, this vast Data Value Network allows citizens to explore the cultural heritage in completely novel ways, facilitating exploration across institutional, administrative and thematic boundaries.

7. DEMAND AND SUPPLY DISTRIBUTION MODEL

[18] https://www.deutsche-digitale-bibliothek.de/?lang=en

[19] www.europeana.eu/

We here focus on the Data Discovery and Data Distribution activities as core enablers to the Data Value Network to propose a Demand and Supply Distribution Model (Figure 4). These two activities are crucial to the Data Value Network in that without the existence (creation) of data, and its consumption (achievable through data distribution), there can hardly be any value creation upon a data product. In other words, if the data does not exist, or no one knows it exists, no value can be created upon it. The Demand and Supply Distribution Model hence provides an entry point for stakeholders to create value and participate in a Data Value Network by enabling and enhancing the Data Discovery and the Data Distribution activities. This model is based upon the knowledge base described in Section 6.2, where the knowledge base acts as a *dynamic leveler* between data supply and demand.

Entities participating as data producers or publishers in the Data Value Network can be overwhelmed by the amount of competition in the global market. Likewise, data consumers can find it difficult to identify whether the data product they need is already on the market. Moreover, if the data is created with a specific use case in mind, it might be difficult to envision or implement it's use in a different domain. This model we propose can be a solution to these problems, where information about data products resulting from entities' Data Value Networks are indexed in a knowledge base, making them available for easier search and discovery. Using this knowledge base, data consumers can easily identify publishers or producers that are providing the data product that they require. Similarly, data producers can be aware of the data products already on the market, thus having the opportunity to target a niche, if it exists, rather than attempting to compete with established data producers. Basically, this model maps the real-life data demand and supply picture and undertakes the role of a Data Broker, as defined in Section 5. By following this model stakeholders can hence optimise their process of participating in the value creating process upon data products by having a clear picture of the supply and demand, and acting accordingly.

7.1 Demand and Supply as a Service

In order to act as proof of concept, we created a cloud service in the form of a portal. Based on Table 1, we publish information

and knowledge on existing datasets as Linked Data and provide the Demand and Supply as a Service (DSAAS); an entry point to the Data Value Network and hence also to the *Economic Data Ecosystem* (Online: `http://butterbur22.iai.uni-bonn.de/dsaas/`). The portal caters for two discrete roles, reflecting the Demand and Supply Distribution Model, namely data producers (Supply) and data consumers (Demand), and aids the value creation process through enabling and enhancing data discovery and re-use, collaborations, and providing contributions to the data market.

The DSAAS provides two different ways for consuming data; a faceted browser and a RESTful API. The faceted browser enables data consumers (humans) to browse the Data Supply and Demand Knowledge Base (as shown in Figure 4) of existing data that they can consume or even contribute to. The RESTful API, on the other hand, provides automated access to the Knowledge Base. This enables third parties to provide their own applications based on the available data. Following, we explore the various functionalities of the DSAAS:

Browsing existing datasets: Data Consumers, or other entities who want to discover the current state of the data market, can browse existing datasets through a faceted browser. Results can be filtered according to various aspects of the datasets, including the licence, publisher, data format, data content, geographical coverage, etc. This faceted browser hence enables stakeholders to quickly search for datasets that are relevant to their current need. Once a particularly interesting dataset is identified, a user can also view all the details about the dataset, as well as any use cases of other stakeholders who already used the dataset. These use cases are provided with the intention of giving potential consumers further insight on the (re)usability of the dataset in question, especially in contexts not in the domain of the original dataset use.

Adding new data: Apart from browsing existing data, stakeholders can also use the DSAAS to add new datasets. This is very simply done by filling the relevant information in a template. The latter includes information such as the producer of the dataset, the formats available for consumption, the licence under which the dataset can be used, the language of the dataset, the temporal and geographic coverage, etc. Moreover, if the stakeholder knows another stakeholder who used their dataset in a use case, this information can be added as well. The more information that a stakeholder can provide about the dataset in question, the more easily the dataset can be discovered (and re-used) by data consumers.

Browsing the data requests: Similar to the functionality of browsing existing datasets, stakeholders can browse through data requests posted on the DSAAS. These requests indicate a need for specific data. This will allow data producers to identify a niche in the data market that was not previously catered for by any other data producer, and provide them with the opportunity to target it in order to be on the forefront of the competition. Data consumers, on the other hand, can second any data request in order to raise awareness about its relevance to various stakeholders.

Adding a new data request: Adding a data request, or a demand, is simply done through filling a form, where stakeholders can enter information about the data that they require to varying degrees of details.

Getting a recommendation: Whilst still not fully implemented, we are currently exploring how to exploit the information provided by a user for a data request to recommend any similar existing datasets. Initially this will be implemented by comparing the domain, keywords, and descriptions.

In order to best represent the supply and demand of data, we defined the **Demand and Supply Ontology**[20] (DSO) to act as the underlying schema to the DSAAS. We re-use existing concepts from DCAT[21], Dublin Core[22], and FOAF[23] to ensure interoperability and easier interlinking. The DSO (along with the DSAAS) improves upon existing schemas and initiatives such as DCAT, CKAN[24], and Datahub[25] in that it enables us to represent not only the dataset in question (resulting in a catalogue of datasets), but also all the involved actors, as well as their relationships with the data at hand, hence providing some context on the provenance of the data. Moreover, the DSO also enables us to represent the context of the re-use of a dataset. This representation allows us to holistically portray the picture of the supply and demand within a data market. It also

[20] `https://w3id.org/dso`
[21] `https://www.dcat.org/`
[22] `http://dublincore.org/`
[23] `http://xmlns.com/foaf/spec/`
[24] `http://ckan.org/`
[25] `https://datahub.io/`

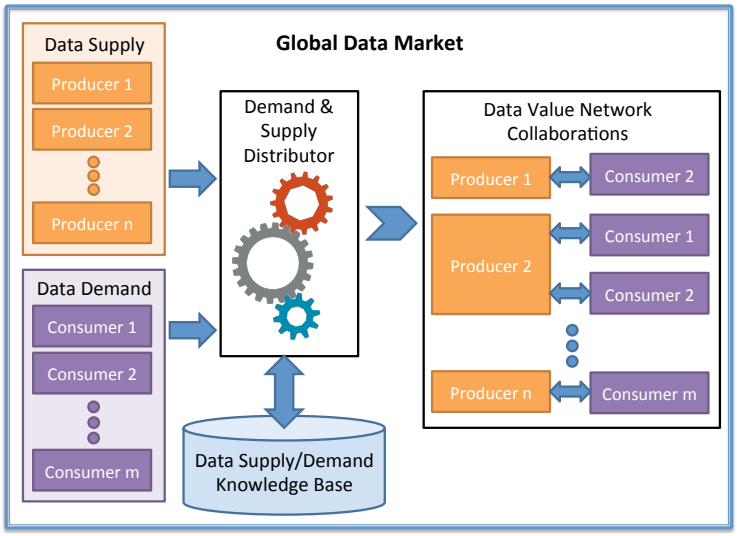

Figure 4: Demand and Supply Distribution Model

enables us to store and publish data using Linked Data principles. The core concepts of the DSO (shown in Figure 5) are:

Dataset: This concept represents information about existing datasets that can be consumed. This concept has properties that describe the dataset at hand, including the licence, theme, language, temporal coverage, description, date issued, data modified, etc.

Distribution: This concept reflects information about the access method of a Dataset. Hence, a dataset can have one or more Distributions. A Distribution has properties that describe the data format, the access and download URLs, the date modified, etc.

UseCase: A UseCase is an example of a 'success story' of the use of a Dataset. Properties of a UseCase include a description of the use case, the distribution and the dataset it uses, the stakeholders' motivation, etc.

DataRequest: This concept encompasses similar details to a Dataset concept, except that this concept reflects the *need* for a dataset.

Publisher: A Publisher is the entity or stakeholder responsible for creating a Dataset.

Consumer: A Consumer is responsible for generating a UseCase through consuming an existing dataset.

Distributor: As opposed to a Publisher, a Distributor is the entity responsible for creating a distribution of a Dataset. For example, whilst the original dataset was created by a governmental entity on paper, a Distributor would be the one responsible for creating an RDF version of the same data.

Agent: Apart from being a super-class to the Publisher, Consumer, and Distributor concepts, an Agent is an entity who is responsible of submitting a DataRequest, or otherwise seconding an existing one.

In order to establish its validity, the DSAAS is already being used in the ODINE Project[26]. The latter is an open data incubator that provides access to hundreds of companies and SMEs working on open data businesses. Starting in July 2015 (and ending in August 2016), the calls for such entities have already attracted the participation of over 300 companies or SMEs. Use cases and datasets used within the latter SMEs are being fed into the DSAAS knowledge base, hence creating a network of connections and collaborations between the datasets and their producers/consumers. We envisage that once the knowledge base is more substantial, we can also provide additional functions such as a crowdsourced effort for knowledge base curation, importing of existing data catalogs, and the provision of a recommender system built on top of the knowledge base.

With the aim of determining the potential benefits of using the DSAAS, we lead out a preliminary evaluation where a number of data producers and consumers were requested to fill out a survey[27]. At the moment of writing, twelve persons responded the survey[28], of which nine are both data consumers and publishers, whilst two are only data consumers. When asked about the most common challenges in consuming open datasets, the respondents of the survey indicated that the low discoverability of the dataset in question is the challenge they faced most (11), with the challenge of lack of licence information being close second (10). Lack of provenance information and the uncertainty whether a dataset even existed followed (8 each), whilst the lack of use cases of previous use completed the list of challenges (7). These challenges are in line with aims of the DSAAS, in that we provide the essential data to enhance dataset discoverability, as well as provenance, licence, and use case information. For the rest of the questions in the survey, which directly concern the foreseen benefits of using the DSAAS, we use the Likert scale to evaluate the degree to which the evaluators agree with the specified benefits. The results are as follows (SA-Strongly Agree, A-Agree, N-Neither agree not disagree, D-Disagree, SD-Strongly Disagree):

- The DSAAS can help stakeholders to easily identify the demand in the data market (by enabling stakeholders to submit data requests) - SA:1, A:6, N:2, D:2, SD:1
- The DSAAS can help stakeholders to easily identify the supply in the data market (by listing existing datasets) - SA:2, A:7, N:1, D:0, SD:2
- The DSAAS can help stakeholders to identify a niche in the data market, and hence target it specifically (through catering for a data request) - SA:2, A:5, N:3, D:2, SD:0
- Stakeholders are encouraged to re-use datasets if success stories (use cases) are provided - SA:2, A:8, N:1, D:0, SD:1
- The DSAAS encourages stakeholders to collaborate with each other by showing their interests in specific dataset domains - SA:3, A:4, N:4, D:1, SD:0
- The DSAAS would be a good tool to showcase datasets and encourage their consumption - SA:5, A:5, N:1, D:0, SD:1
- By allowing consumers to put a request for a dataset, the DSAAS could possibly make the acquirement process faster - SA:2, A:9, N:1, D:0, SD:0

Whilst still preliminary, through this evaluation we can conclude that overall the survey responders agree with the benefits of using the DSAAS that we portray. Ten out of twelve respondents agree that the DSAAS encourages dataset sharing and consumption, and eleven respondents agree that it can help the data acquirement process. Although there are varying opinions on the benefits, the majority of the respondents always agrees that the tool will improve their participation in the data market. These results, while not conclusive, certainly indicate the potential of our approach. In order to confirm these results conclusively, we plan to gain further insight into the usefulness of the DSAAS through further, more in-depth studies that include an analysis of the use of the DSAAS in a real-life scenario. In such studies we can assess the resulting impacts of using the DSAAS, including any improvement in efficiency and efficacy of the implemented approach using the DSAAS.

8. CONCLUSION

The increasing dataification within our information society, particularly in a government setting, has required the need for the specification and implementation of new value chains in order to enhance the data exploitation process. With the aim of projecting our vision of generating a new *Economic Data Ecosystem* based on data value chains, we propose the *Data Value Network*. The Data Value Network models the co-production of value through the interaction of a number of actors who participate through a number of roles. We also specify the various activities and actions used to modify the data and innovate by adding value to data-centric products. Rather than limiting the impact of the Data Value Network to the economic dimension, we advance on existing literature by characterising the Data Value Network according to three dimensions; *technical*, *economic*, and *societal/political*. We also define

[26] https://opendataincubator.eu/

[27] Complete survey: https://goo.gl/2N3nnZ

[28] Survey responses: https://goo.gl/NzxVI2

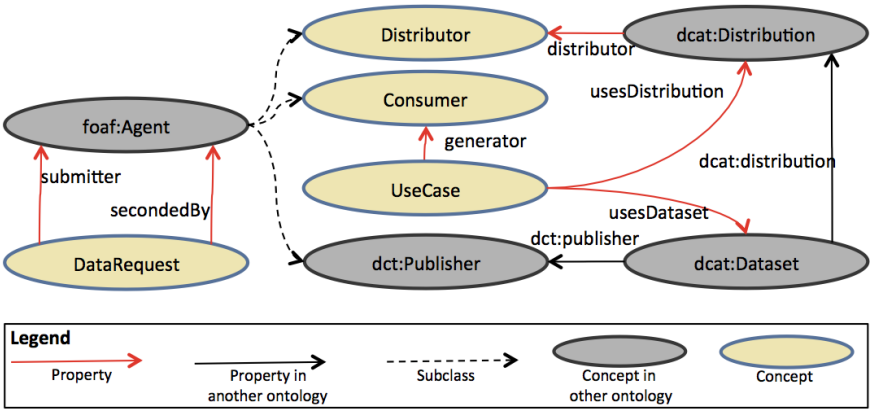

Figure 5: The main concepts in the Demand and Supply Ontology

the *Demand and Supply Distribution Model*, which provides an insight on how an entity can successfully enter the global data market, whilst maintaining a competitive edge. The Demand and Supply as a Service (DSAAS) application acts as concrete implementation of the proposed model. Acting as a dynamic leveller, this service enables stakeholders to more easily advertise existing data products, or otherwise create a request for specific data. This match-making service provides the required knowledge to participate in a Data Value Network, and has the potential of creating a sustainable environment of (government) data re-use, enhancing the value-creation cycle within the Data Value Network.

9. REFERENCES

[1] C. Alexopoulos, A. Zuiderwijk, Y. Charalabidis, E. Loukis, and M. Janssen. Designing a second generation of open data platforms: Integrating open data and social media. In *Electronic Government - 13th IFIP WG 8.5 International Conference, EGOV 2014, Dublin, Ireland, September 1-3, 2014. Proceedings*, pages 230–241, 2014.

[2] R. Amit and C. Zott. Creating Value Through Business Model Innovation. *MIT Sloan Managment Review*, (53310), 2012.

[3] S. J. Barnes. The mobile commerce value chain: analysis and future developments. *International Journal of Information Management*, 22(2):91–108, Apr. 2002.

[4] D. Crié and A. Micheaux. From customer data to value: What is lacking in the information chain? *Journal of Database Marketing & Customer Strategy Management*, 13(4), July 2006.

[5] N. Cukier and V. Mayer-Schoenberger. The Rise of Big Data. *Foreign Affairs*, (May/June):28–41, 2013.

[6] J. Elkington. Towards the sustainable corporation: Win-win-win business strategies for sustainable development. *California Management Review*, 36(2), 1994.

[7] K. Janssen. The influence of the {PSI} directive on open government data: An overview of recent developments. *Government Information Quarterly*, 28(4):446 – 456, 2011.

[8] J. M. Juran. *Juran's Quality Control Handbook*. McGraw-Hill, 4 edition, 1974.

[9] R. Kaplinsky and M. Morris. A Handbook for Value Chain Research. (September), 2002.

[10] J. Kučera, D. Chlapek, and M. Nečaský. Open government data catalogs: Current approaches and quality perspective. In *Technology-Enabled Innovation for Democracy, Government and Governance*, volume 8061 of *Lecture Notes in Computer Science*, pages 152–166. Springer Berlin Heidelberg, 2013.

[11] 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*, pages 568–576, 2009.

[12] C. C. Lee and J. Yang. Knowledge value chain. *Journal of Management Development*, 19(9):783–794, 2000.

[13] H. G. Miller and P. Mork. From Data to Decisions: A Value Chain for Big Data. *IT Professional*, 15(1), Jan. 2013.

[14] L. Morgan, J. Feller, and P. Finnegan. Exploring value networks: theorising the creation and capture of value with open source software. *EJIS*, 22(5), 2013.

[15] K. O'Hara. Enhancing the quality of open data. In L. Floridi and P. Illari, editors, *The Philosophy of Information Quality*, volume 358 of *Synthese Library*. Springer International Publishing, 2014.

[16] J. Peppard and A. Rylander. From Value Chain to Value Network:. *European Management Journal*, 24(2-3), Apr. 2006.

[17] L. L. Pipino, Y. W. Lee, and R. Y. Wang. Data quality assessment. *Commun. ACM*, 45(4):211–218, Apr. 2002.

[18] M. E. Porter. *Competitive Advantage: Creating and sustaining superior performance*. NY: Free Press, 1985.

[19] M. E. Porter and V. E. Millar. How information gives you competitive advantage. *Harvard Business Review*, 63(4):149–162, 1985.

[20] J. F. Rayport and J. J. Sviokla. Exploiting the Virtual Value Chain. *Harvard Business Review*, 73:75, 1995.

[21] L. A. R. Rojas, G. M. T. Bermúdez, and J. M. C. Lovelle. Open data and big data: A perspective from colombia. In *KMO'14*, pages 35–41, 2014.

[22] D. Schlagwein and D. Schoder. The Management of Open Value Creation. In *HICSS-44*. {IEEE} Computer Society, 2011.

[23] S. Shah, A. Horne, and J. Capella. Good Data Won't Guarantee Good Decisions - Harvard Business Review. *Harvard Business Review [Internet]*, (April), 2012.

[24] A. Wassenaar. E-Governmental Value Chain Models. In *11th International Workshop on Database and Expert Systems Applications (DEXA'00), 6-8 September 2000, Greenwich, London, {UK}*. {IEEE} Computer Society, 2000.

[25] C. Wendy, C. Wae San, F. Sander, and v. S. Eva. Creating

value through open data, 2015.