D4.3 Data Value Chain Database v1

Coordinator: Fabrizio Orlandi (IAIS)

With contributions from: Judie Attard (IAIS), Alan Ponce Rodriguez (SOTON), Luis Daniel Ibáñez González (SOTON)

Reviewer: SOTON

Deliverable nature	Other (O)
Dissemination level (Confidentiality)	Public (PU)
Contractual delivery date	31st January 2016
Actual delivery date	
Keywords	Data value chain database

Table of contents

1

Executive summary

The data value chain database (v1) has been developed in order to ensure that the ODINE project maximizes its reach and keeps tracks of all the possible open data datasets and shareholders. Moreover, this work contributes to the analysis of the current open data ecosystem by providing an essential tool for monitoring open data reuse and evolution. This deliverable describes the current version (v1) of the database as well as the upcoming activities towards the creation of the final ODINE data value chain database (v2).

There is a great potential for the data value chain database to create a positive impact on the open data ecosystem. Important future steps will be about validating and testing the database with real use cases and contacting partners, projects, SMEs and acceleration programmes to further enrich the database.

2

Introduction

The value chain is an important concept that involves identifying the various activities and roles in manufacturing a product. The value chain is also used as a tool that enables the analysis of interactions between the different activities in order to identify the sources for competitive advantage. However, this concept is not representative of the activities and roles in a value chain where the product is a non-tangible data product. In our information society, data increasingly becomes a commodity and the basis for many products and services. In this context, we aim to contribute to the definition of data value chains and identify the activities and roles in a data value chain specifically required when the product is a data product. We 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. Through the data value chains database we project our vision of generating a new economic data ecosystem that has the Web of Data at its core.

In this deliverable, we first provide an overview of the related work on data value chains and introduce some essential definitions. Then, we describe our ODINE Data Value Chain Database available online as a cloud service prototype called "Demand and Supply as a Service". The service follows the concept of a demand and supply distribution model, and it aids entities in participating in the data value "network", hence resulting in a data product.

3 Related Work

In our information society, data becomes increasingly a commodity and the basis for many products and services. Examples are Open Data, Linked Data or Big Data applications, such as *a)* inter-modal transportation information services, integrating open data from a variety of sources, *b)* supplier and product data integration (distributed master data management), *c)* using and integrating RDFa from a variety of sources on the Web, or *d)* large-scale customer or sensor data analytics involving several organisations. In recent years, in order to reflect this "dataification" [Cukier and Mayer-Schoenberger, 2013] and our datacentric society, the concept of data value chains was introduced, building upon the concept of traditional value chains for tangible products.

The rationale of a data value chain is to extract the highest possible value from data by processing and adapting it. While some entities have the primary goal of producing data as a product, other entities produce valuable data as a by-product of their day-to-day activities, which serve a different purpose. Value can be added to the generated raw data to make it re-usable, hence resulting in a new data product. 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. Any traditional sector, such as health, transport, or retail, can thus benefit from new-found opportunities based on digital developments.

The **value chain**, both as a concept and as a tool, has been around for about three decades [Porter, 1985]. 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 to identify what activities are the source of competitive

advantage within these industries. As successful as the value chain concept was to achieve this aim, during the last years products and services are becoming increasingly digital, and exist in a more non-tangible dimension [Peppard and Rylander, 2006]. In addition, the traditional value chain model does not consider when information is used as a source of value in itself [Rayport and Sviokla, 1995]. Thus, the original concept of value chain is becoming an inappropriate method with which to identify value sources in today's industries that produce non-tangible products [Peppard and Rylander, 2006]. Newer definitions of the concept, such as in [Crié and Micheaux, 2006], [Latif et al., 2009], [Lee and Yang, 2000], [Miller and Mork, 2013], [Peppard and Rylander, 2006], cater for the digital dimensions.

We build upon previous definitions and characterise the Data Value Chain (or, more appropriately, the "Data Value Network") according to three dimensions; *technical*, *economic*, and *societal*, where we identify the different roles and activities within the value chain. 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.

4   The Data Value Network

Compared to the literature we discussed in the previous section, here we target the domain of *data* value chains. We identified the lack of literature that discusses the creation of value on a data product, as well as the actual processes used to create this value. 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*. We included common activities executed on data products, where the final aim is usually the consumption of the data product. Due to the differing order of executing the relevant activities, a network was deemed to be the best way to represent the interactive nature of adding value to data products. This is because of the non-tangible nature of data, as well as for the fact that a number of entities can participate in the network simultaneously, as opposed to following the "chain" structure. For example, the Data Curation activity can be executed many times, by different entities, and in various orders (before or after other activities). On the other hand, for example the Data Cleaning action within the Data Curation activity, usually follows the same structure of execution for identifying and correcting errors in the data. For this reason within the generic Data Value Network we can have Data Value Chains for specific actions.

We define a **Data Value Network** to be:

A set of linked activities having the aim of adding value to data in order to exploit it as a product where different **actors** can participate by executing one or more **activities**, and each activity can consist of a number of **actions**. In turn, each action can be broken down in one or more **data value chains**. For example, the Data Cleaning activity in the Data Curation action needs to follow a specific sequence for identifying issues in the data and cleaning them.
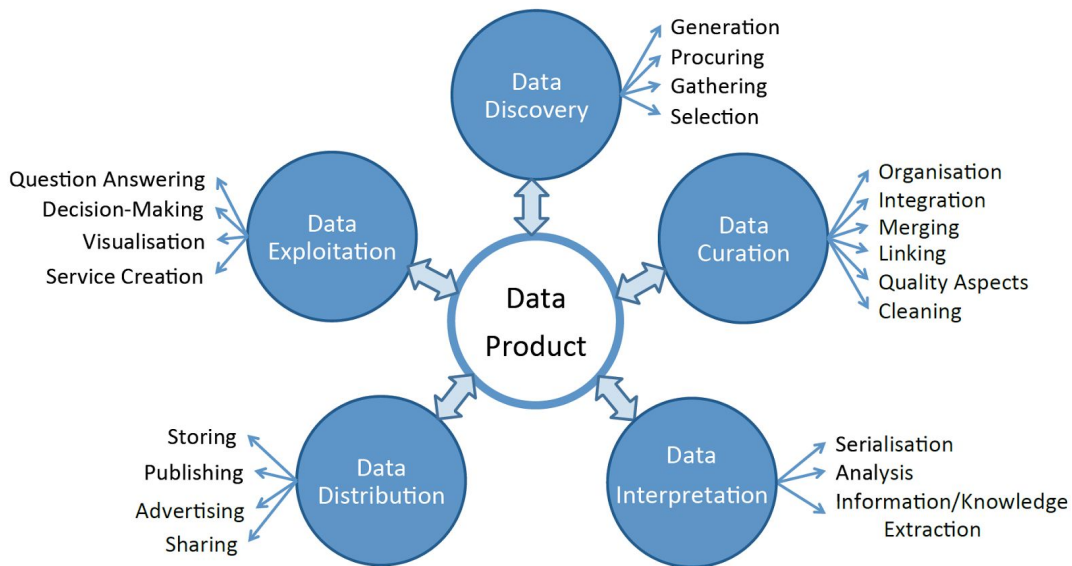
FIGURE 1. THE DATA VALUE NETWORK – ACTIVITIES AND ACTIONS

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 [Porter, 1985], and form a network around the data product, as they can be executed in any order by different actors. This means that 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 the data product by one or more actors. For example, two actors might participate in interpreting the data, while five actors might exploit it. Each activity is also made up of a number of actions. While not exhaustive, the listed actions are the most common and generic processes that can be executed on a data product. Actors can also collaborate within a single activity to co-produce value by participating in different actions.

Let us consider the example of data produced in the day-to-day activities of a governmental entity. The Government executes the Data Discovery activity by generating the data. The Government then decides to publish the collected environmental data onto a data portal. So, the Government Cleans and Organises (Data Curation Activity) the data, and Publishes (Data Distribution) the data on the portal. An environmental NGO wants to use this data to see how to target their future efforts towards improving the environment. For this reason they use the data on the portal by Analysing the data and Extracting Information (Data Interpretation Activity). They then Visualise the results, and base their Decision Making process (Data Exploitation Activity) on the results. In turn, the NGO also Publishes and Shares (Data Distribution) the results on their website and social media.

For more information about the Data Value Network please refer to our recent publication [Attard et al., HICSS 2016].

5   Data value chain database v1

Demand and Supply Knowledge Base

Due to the potential of data to be used over and over (until it remains relevant), the economic impact of adding value to it is different than with physical products. First and foremost this is evident in the reuse of data in another context, or domain, that it was originally envisaged for. For example, e-commerce businesses, use historical purchase data to identify patterns and suggest items to users. Moreover, the data can be processed repeatedly 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, the license, 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 than the one originally envisaged, as is particularly evident in the sixth entry in *Table 1.*

#	Publisher	Domain	Nature of Content	Access Method/Data Format	Licence	Consumer	Aim	Description
1	Accuweather	Weather Data	Images, Statistical, Geographical, Sensor	API	Proprietary	Accuweather	Forensic Services	The weathe... entity in or... weather-in... (http://ww... es-in-canad... (https://en...
2	MusicBrainz	Music Data	Records	API	Open	BBC	News Enhancement	Music data... music site (... discographi...
3	Europeana	Cultural Data	Multimedia, Records	API, SPARQL Endpoint, Datadump	Open	Historiana.eu	Educational Portal	Data from t... Historiana....
4	Her Majesty's Treasury	Government Expenditure Data	Statistical	Sparql Endpoint, CSV	Open	wheredoesmymoneygo.org	Informative Portal	Where Doe... citizen enga... information... (http://data...
5	Safecast	Environmental Data	Sensor	API, CSV	Open	Fukushima Government	Radiation Awareness	The Fukush... from Safeca... levels in dif... (http://fuku...
6	Vehicle	Vehicular Data	Sensor	n/a	Proprietary	Progressive	Insurance Services	Progressive... identify a p... according t...
7	Office for National Statistics	Crime	Statistical	CSV, Spreadsheets	Open	Walkonomics	Environmental Safety	Walkonomi... National St... person can... various cat... pavement c...

TABLE 1: DEMAND AND SUPPLY KNOWLEDGE BASE EXAMPLE

Demand and Supply Distribution Model

We here propose the Demand and Supply Distribution Model (Figure 2) as the potential approach towards entering the global data market by producing a data product, and generating an economic data ecosystem. This model is based upon the knowledge base described in Section 5.1 where the knowledge base acts as a dynamic leveller between 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. This model we propose can be a solution to these problems, where information about entities' Data Value Networks are indexed in a knowledge base such as the one we provide, making them available for search and discovery. Hence, an entity participating in a Data Value Network and having a data product can 'advertise' this product on the Demand and Supply Distribution Model. The entity should provide a description of the data, and its original use. On the other hand, a consumer can provide any success stories about using this data, or otherwise request other data which is not yet available. 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 undertakes the role of a "data broker".
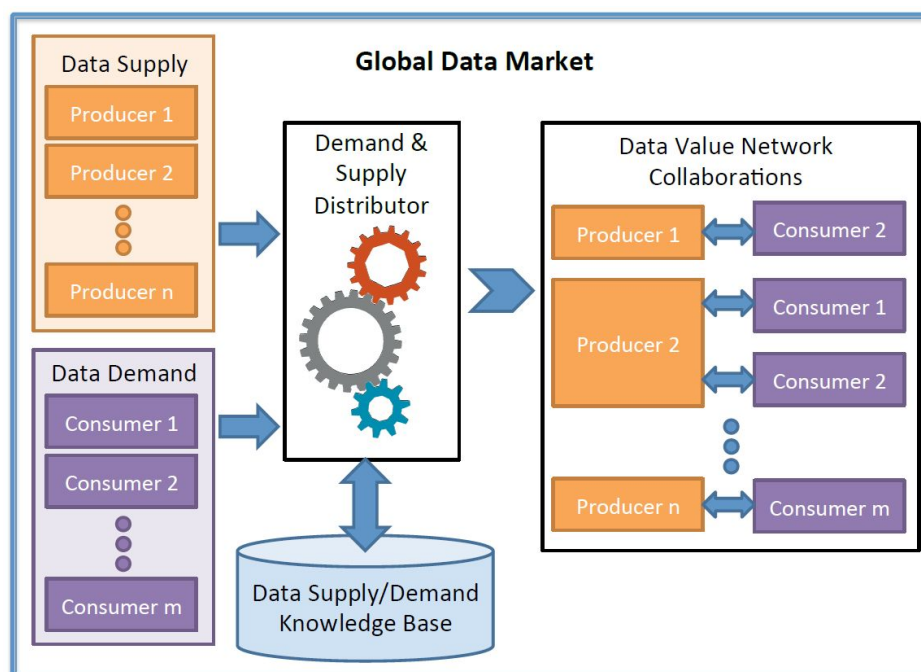
FIGURE 2: DEMAND AND SUPPLY DISTRIBUTION MODEL

By following the Demand and Supply Distribution Model, three Data Value Chain dimensions, namely the Economic, Technical, and Societal dimensions, will be affected in various ways. The *economic* dimension will be affected through the specialisation of the entities who participate in the Data Value Network. By identifying the right niche to target through the model, entities can

specialise in providing an innovative data product that would give them the best competitive advantage when participating in the global data market. For example, seeing that there is a new demand for weather data of a specific nature, a stakeholder can be the first to provide such data. This has the potential of providing a sustainable income growth. Through the entities' specialisation, the *technical* dimension is affected by the development and execution of different Data Value Networks. The difference in the Data Value Network activities between entities can set them apart in the competition, so the intelligent execution of the network is essential. Finally, the *societal* dimension is affected through the global market. The availability of data as a product means that more and more data consumers will innovate and create services upon this data.

Demand and Supply as a Service

As a first version of the ODINE data value chain database, we created a cloud service in the form of a portal. Based on Table 1, we provide an entry point to the ODINE value chain and the Economic Data Ecosystem. The service is currently available online at ***http://purl.org/net/dsaas***. The portal caters for two discrete roles, reflecting the Demand and Supply Distribution Model, namely data producers (Supply) and data consumers (Demand).

The Demand and Supply as a Service provides two different ways for consuming data:

1. A faceted browser enables data consumers (humans) to browse the Data Supply/Demand Knowledge Base (as shown in Figure 2) of existing data that they can consume. Figure 3 shows a screenshot of this feature. Consumers can filter the available data by *licence* (open or proprietary - which would mean if the data is free for consumption or require some kind of registration or fee), *data provider/publisher*, *domain* of the published data, *API* or *consumption formats* (e.g. RESTful APIs, SPARQL Endpoints), and available data *content types* (e.g. images, sensor data, records). Moreover, in case a consumer needs specific data that is not available yet, we provide an online form that consumers can fill, providing details about the data that is required (Figure 4). This includes the *domain* of the data required for consumption, the *nature* the content required (e.g. images), and the *intended use* for the required data.

2. The second way of consuming data is through a *RESTful API*. This API enables automated access to the Knowledge Base. This enables third parties to provide their own applications based on the available data. In a similar way to the consumers' faceted browsing experience, data producers can explore any data requests by the consumers. As described above, this helps them identify a niche in the market that was not targeted by any other producer previously. Therefore, once producers have data available, the Demand and Supply as a Service enables them to fill in a form with relevant details about the data they produced. This includes information about the data producer/publisher, the domain and licence of the data, and methods available for accessing the data. Additionally, a data producer can also add success stories of other entities already re-using the data in question.

FIGURE 3: FACETED SEARCH FOR DATA AVAILABLE FROM PUBLISHERS

FIGURE 4: REQUEST DATA PROVIDERS FOR SPECIFIC DATA THAT IS NOT CURRENTLY AVAILABLE

6   Conclusions and Future Work

Data is a commodity in our information society and the "dataification" of products has led to the need for a change existing value chains. Our aim is thus to project our vision of generating a new Economic Data Ecosystem based on the concept of data value chains. With the Web of Data at its

core, we propose the use of the Data Value Network as a process with which to generate value as a data product. Fashioned as a network, it enables the coproduction of value through the interaction of a number of stakeholders.

Additionally, we define 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 application we developed acts as proof of concept to 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 has the potential of creating a sustainable environment of data re-use, enhancing the value creation cycle within the data value chain.

Future steps include aligning this data model to projects such as the Open Data Monitor[1] and the ODINE Stakeholder database (described in deliverable D5.3). This alignment would allow us to bootstrap the data value chain database with a large amount of records about existing datasets and their characteristics (e.g. domain, licence, publisher, etc.). The collection of data for the database then would not be only done manually but automatically. The collected data would still need to be complemented with information about the kind of usage for each dataset and the demand for datasets. This information has to be curated manually. However, we plan to engage with the ODINE network of SMEs applying to the programme and distribute a questionnaire that would help us collecting information about supply and demand of open datasets.

As our effort is aligned with the ODINE Stakeholder DB, we are also in cooperation with the following projects:

- European Data Market study - http://datalandscape.eu
- Apps4EU - http://www.appsforeurope.eu/
- Finodex - http://www.finodex-project.eu/
- OpenDataMonitor - http://opendatamonitor.eu/
- OpenAIRE2020 - https://www.openaire.eu/
- FIWARE - http://fiware.org
- DIGIWHIST - http://digiwhist.eu/

7   **References**

Attard, J., Orlandi, F., and Auer, S. (2016). Value Creation on Open Government Data. In *Proceedings of the 49th Hawaii International Conferences on System Sciences (HICSS 2016),* Kauai, HI:IEEE.

Crié, D., & Micheaux, A. (2006). From customer data to value: What is lacking in the information chain? Journal of Database Marketing & Customer Strategy Management, 13(4), 282–299.

Cukier, N., & Mayer-Schoenberger, V. (2013). The Rise of Big Data. Foreign Affairs. Retrieved from https://www.foreignaffairs.com/articles/2013-04-03/rise-big-data

---

[1] http://opendatamonitor.eu/

Latif, A., Us Saeed, A., Hoefler, P., Stocker, A., & Wagner, C. (2009). The Linked Data Value Chain: A Lightweight Model for Business Engineers. In *Proceedings of International Conference on Semantic Systems* (pp. 568–576).

Lee, C. C., & Yang, J. (2000). Knowledge value chain. Journal of Management Development, 19(9), 783–794.

Miller, H. G., & Mork, P. (2013). From Data to Decisions: A Value Chain for Big Data. IT Professional, 15(1), 57–59.

Peppard, J., & Rylander, A. (2006). From Value Chain to Value Network: European Management Journal, 24(2-3), 128–141.

Porter, M. E., & Millar, V. E. (1985). How information gives you competitive advantage. Harvard Business Review, 63(4), 149–162.

Rayport, J. F., & Sviokla, J. J. (1995). Exploiting the Virtual Value Chain. Harvard Business Review, 73, 75.