Capability Matrix for Open Data

Fatemeh Zeleti, Adegboyega Ojo

Capability Matrix for Open Data

Fatemeh Ahmadi Zeleti[1] and Adegboyega Ojo[1]

[1]INSIGHT Centre for Data Analytics @NUI Galway,
Digital Enterprise Research Institute, IDA Business Park, Lower Dangan,
Galway, Ireland
{Fatemeh.Ahmadizeleti, Adegboyega.Ojo}@insight-centre.org

Abstract. With the rapid growth of data on the web, a significant number of businesses have embraced the idea of open data to enable innovation, generate revenue and develop new data products and services. However, many businesses are still reluctant, to due to the paucity of information on requisite capabilities to successfully publish and reuse open data to support their business goals. However, scholarly efforts articulating the nature and types of capabilities required for open data processes and innovation are very limited. This could be partly attributed to the fact that most of the discussions on open data capabilities are predominantly in the practice community. We bridge this knowledge gap by mapping Open Data (OD) capabilities described in extant literature, organizing these capabilities into major capability areas and deconstructing these areas to build a capability matrix. The resulting capability matrix provides OD program managers in particular with a framework (or tool) to develop their detailed OD capability requirements. We also believe that the developed framework provides researchers with a foundation for further analysis of OD capabilities.

Keywords: Open data, value chain, open data value chain, Information Technology capabilities, open data capabilities, data-driven virtual organizations

1 Introduction

During the last decade, businesses across the globe have struggled to comprehend and adapt to the changes brought on by the ubiquitous growth of Information Technology and the Internet [1], [2]. One of the changes is the emergence of OD, which resulted from opening up and sharing of non-sensitive information in machine-readable format with businesses and the general public [3],[4]. Drivers for opening up data include ensuring accountability, delivering quality services, reducing operating costs, and stimulating innovations [5], [6] [7], [8]. Data underpins businesses and the economy, and thus key in providing new insights into consumer needs and enabling new products and services to be developed [9].

Recently, attention of major stakeholders in the OD community, including policymakers have shifted to the economic value of the rapidly growing OD asset. For instance, the European Commission estimates that the direct economic gains from

opening up Public Sector Information (PSI) or government data could amount to €40 billion a year. Similar believes across the world has spurred a growing number of OD small and medium enterprises seeking to tap into the shared resources or commons. As new entrants flood the marketplace, OD businesses are seeking to uniquely position themselves through specialization in the landscape to create and capture value for their stakeholders [10].

To creating and capturing value for OD stakeholders, OD business managers are required to employ emerging set of capabilities to catalyze positive change in the business [11]. Capabilities enable a business to achieve specific purpose or output [12].

However, there are no scholarly studies available that gather OD capabilities, map capabilities to OD value chain and provide sophisticated tool for businesses to describe, design, challenge, invent and pivot OD business capabilities. In addition, understanding, using, and representing a set of OD capabilities for OD businesses is challenging for many management teams. This research is the first effort to address these problems by providing OD businesses with an OD Capability Matrix as a tool to specify and develop OD capabilities for specific (virtual) organization needs.

2   Concepts

The concepts described in this research paper are OD, capability types and value chain frameworks. Figure 1, shows the relations between these concepts. Capabilities are required to exploit OD. Capabilities are required to build and implement value chain. With an adequate set of capabilities used to build value chain, value chain can then generate OD value. Each of these is elaborated below.
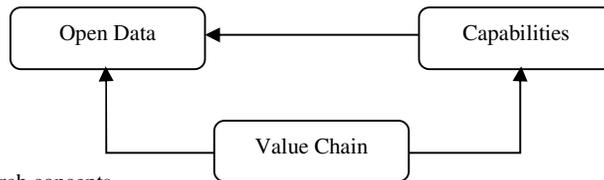
Fig. 1. Research concepts

2.1   Open data

Nowadays, surprising amount of data is generated and stored than at any other time in history [11], [13]. As a practice of good governance, governments globally started to open up their public information in various domains, such as transportation, education, mobility and meteorology [14], [11]. When data is freely accessible and re-usable by public, it could have a larger impact on citizens' ability to hold governments accountable and stimulate innovation [11].

The more technical view of OD is that it is machine-readable [8], [3]. OD is published in common standards, accessible through non-proprietary software, and

subject to open licenses [15]. However, there are also limits to what can be released. For example, personal information and information related to national security [5].

OD can help uncover consumer preferences, allowing businesses to improve new products [8], increase revenue, and expand the supply and value chain [16]. However, to benefit and capture value from OD and build or expand the business value chain, businesses are required to identify and develop set of capabilities. Section below presents definition of capability and describes capability types.

2.2 Capability and Capability Types

According to Ulrich and Rosen [12], a "business capability" is a particular ability or capacity that a business may possess to achieve a specific purpose or outcome. In business study and community, this definition of capability has been in common use for the past several years.

Researchers argue that different types of capabilities are required in order to provide businesses a competitive advantage [17], [18]. However, in this research work, we focus on value capabilities of OD businesses. In table 1, business capability types and capabilities associated are shown.

Table 1. General business capabilities

Value capability	Innovative/ Dynamic capability	Competitive capability
-Individual/competences -Business process -Organization -IT infrastructure -Management/ governance -Technological	-Process innovation -Knowledge mgt. -Manufacturing performance -Supply chain integration	-IT (Strategic choices) -Manufacturing strategy -Business operational (Localization/ Internationalization)

Value capability: Value capability is specific set of capabilities required for adding value to the product or output. This capability consists of the following elements:

Individual/competences: Jaques and Stamp [18] define it as the extent and complexity of the context within which an individual can operate.

Business process: is a collection of related, structured activities or tasks that produce a specific service or product for a particular customer or customers [19], [20].

Organization: refers to the way systems and people in the organization work together to get things done [21].

IT infrastructure: is the technological foundation of computer, communications, data and basic systems. It includes hardware, software, network resources and services [17], [22], [23], [24], [25].

Management/Governance: is a process of controlling things or people and action of governing an organization [21].

*Technological: t*echnology is a knowledge embedded in products and processes on doing practical things, especially producing things [26], [23], [27].

Each of the capabilities described above are valuable and necessary for a business and/or organization. These capabilities come together to form more holistic approach to build and manage value chain. Section below presents three value chain frameworks.

2.3   Value Chain Frameworks

Businesses must pay attention to how to use OD to create value for their business [28]. In OD context, value is not only money but, value can also be economic, social, transparency, democratic and etc. Thus, it is essential for a business to establish value chain that suits the purpose of the business [28]. From the literature, we extracted the three well grounded value chain frameworks. *Porter's Value Chain,* which is based on the concept of physical value chain of the firm [29], [30]; *Rayport and Sviokla's Value Chain,* which is based on the concept of virtual value chain of the firm [28] and the *Public Open Government Data Value Chain,* which is based on the concept of PSI value chain [31]. For analysis of OD capabilities in section 4, we use PSI value chain as it visibly put in place both Physical and Virtual value chains. Besides, PSI value chain also suits better to the context of OD.

Open Government Data Value Chain: Understanding value chain of open government data is necessary to establish knowledge on who are the actors and what are the main roles of actors in various data related activities.

The following PSI value chain is in the context of the European Commission PSI Directive. Value chain identified four main phases:
Data Generation, Data Collection, Aggregation and Processing, Data Distribution and Delivery and Final Data Use [31], [32].

3   Eliciting Open Data Value Capability Areas

List of OD value capabilities have been identified and extracted from OD literature. We studied each capability and were able to extract and use top-level capability as capability areas for classifying low-level capabilities. Table 2 shows OD capability areas and capabilities associated with each area.

Table 2. OD value capabilities

OD capability areas and capabilities associated with each area	*References*
Publishing Solution	
Publishing as Linked Data	[15]
Sustainable publishing Solution	[33]
Publishing in different format; machine-readable data	[34], [15], [11], [31]
Publishing on web as API to be queried or data dump to be downloaded as a whole	[15]

Development of software tools to scrape, clean, format, visualize and create API services on the web	[15]
Data Generation	
Efficient design and features to collect massive data	[33]
Technology and Infrastructures	[9]
Reuse of PSI	[35], [9]
Linking information from different sources	[35]
Data Retrieval	
Sophisticated Querying	[34]
Efficient design and features to collect massive data	[33]
Data Processing	
Cleaned data to fill gaps, eliminate invalid records or duplicates, standardize attribute values	[15]
Harmonizing data in terms of format	[34]
Format transformations to allow effective machine reading	[15]
Create mash-up	[15], [33], [11]
Data reform and refine	[15], [33]
Data Analysis, Visualization and Visual analytics	[34], [15], [33], [11]
Data Validation	[34], [15], [33], [11], [31]
Data Quality	[34], [15]
Cataloguing data	[15]
Usage of platforms capable to convert datasets in data streams	[34]
Data geo-referencing	[33]
Provision of computing capacity	[33], [13]
Standardizing Linked Data to allow joining to other datasets	[15]
Data Release	
Proactively release data	[34]
Data structuring	[15], [33]
Data classification	[15], [33]
Support data with metadata	[34]
Data update and maintenance	[34], [15], [33]
Data Storage and Computing Facilities	
Data storage	[33]
Computing capacity	[33]
Providing Access to Data and APIs	
Guarantee on data availability	[34], [15]
Commoditization and democratization of data	[33]
Data distribution channel quality	[36], [15], [33]
Data exposure via GUI	[33]
Data exposure via APIs	[34], [33]
Freeing data	[15]
API development	[34]
Using APIs	[34]
Testing and Bug fixing	[34]
Data change feed	[34]

Data Usage	
Help and guideline on accessing, using and adding data, information or knowledge to the original data source	[34], [15], [11]
Available data on the Web to the public and in formats that citizens can reuse	[35]
Support data intermediaries	[11]
A general search engine helping to locate data	[15]
Dedicated service searching purely for datasets and providing useful categorization and tagging	[15]
Management/Governance	
Collaboration	[15]
Training programmes	[15]
Technical expertise of data holders and publishers	[15], [33]
Quick response accepting additional data for advanced features	[33]

Publishing solution: publishing OD requires methods and mechanisms. For example, publishing as Linked Data is one publishing solution.

Data generation: a business and an organization need to generate data out of data processed for further processing. Data generation design, technology and infrastructure are required to create and collect massive data from various sources.

Data retrieval: this includes extracting the wanted data from data storage. This process requires sophisticated querying and appropriate design for retrieval.

Data processing: this includes manipulating and aggregating data to produce meaningful data and information that can be used by stakeholders. Data processing includes cleaning data, harmonizing, format transformation, mash-up, data reform and refine, analysis, visualization, validation, cataloguing and etc.

Data release: data processed and aggregated by an organization or a business must be released to enable data re-use. Before release, data need to be structured, classified and supported with metadata. Data released require data update and maintenance to ensure continuous data re-use.

Data storage and computing facilities: data is backed up in data storage, which is a device that retains data used for computing. It is essential for an organization to appropriately estimate data storage and computing capacity to ensure data quality.

Providing access to data and APIs: data processed should be freely available, accessible and exposed via GUI and APIs through qualified channel of distribution to be used by public. APIs can be developed by responsible entities and developers for data re-use.

Data users: data users are all stakeholders involved in the context of OD. This is to enable and support users mostly citizens with less technical expertise to effectively use OD.

Management/Governance: this includes all parties involved in an organization or a business with main responsibilities to control the process of generating, processing and make data available to public. In addition, training programs, collaboration and quick response to changes can effectively enhance the process of opening up of data. A major requirement for the capability is decision making.

4   Analysis of Open Data Capabilities

In section 4.1, we align OD capability areas to the four stages of PSI value chain. In section 4.2, we provide a deconstruction of OD capability areas.

4.1 Aligning capability areas to PSI value chain phases

As OD has evolved, businesses have responded by quick adoption of OD. OD changed the way business performs and has changed the business for the better. However, this requires defining and implementing new set of capabilities, processes, resources and management across the business.

However, business managers are sometimes accused of not really understanding how value can be created and captured from OD. Value chain facilitates managers to know what stages they should follow to create and capture value however, knowledge of what capabilities are necessary to achieve each stage is essential. Previously in table 2, we identified OD capability areas and in figure 2, we align these capabilities to different phases of PSI value chain. This helps managers to establish knowledge on the capabilities require for different phases of value chain.

In figure 2, we showed this alignment.
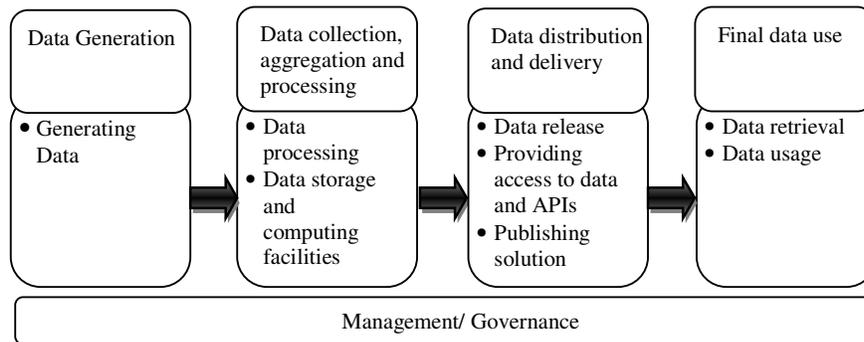
Fig. 2. Alignment of capability areas to PSI value chain

Phase 1: Generating data – according to OECD working paper [31], data generation phase covers all capabilities required for generating data by public sector entities. In our paper, we identified an OD capability area required to carry out this phase of value chain.

Phase 2: Data collection, aggregation and processing - Raw data may not have enough quality and meaning to be used by public, therefore; as it was reported in OECD working paper [31]; data need to be aggregated, linked, and or manipulated in order to add value before being open and freely distributed to public. Our study identified combination of OD capability areas required for data processing. Before data is open to public, it needs to be cleaned, analyzed, standardized, harmonized, transformed in the right format, catalogued and supported with metadata. Quality of

data must be obtained and processed lawfully before publishing. Data storage and computing facilities are necessary to be pooled together for efficiency of data processing and aggregating.

Phase 3: Data distribution and delivery - according to OECD working paper [31], data processed need to be distributed to public to enable access and re-use. In our paper, we identified combination of OD capability areas. Public sector entities and businesses are obliged to define precise publishing solutions, providing access to data and APIs and ultimately releasing data.

Phase 4: Final data use - data previously distributed to public need to be re-used by different users to sustain public value creation [31]. Our paper identified the data retrieval and data usage capabilities are required.

Management/ governance is required throughout all phases of value chain to control people and tasks and to ensure data are processed and released according to the goal of the business.

4.2 Deconstructing OD capabilities

The deconstruction of OD capabilities is a matrix of strategic management and value chain analysis for OD businesses, which aims at capturing all capabilities an OD business require – from generating data to final use and re-use of data - for creating and capturing value from OD.

The matrix is a strategic tool for OD businesses of any scale to exercise and exploit for their business. Capability matrix allows OD business managers to identify what capabilities are valuable to the business they are running. As different businesses have different business model, specific OD capabilities for OD businesses need to be specified.

OD capability matrix (Table 3) is based on the general business value capabilities (table 1), OD capability areas (table 2), and the PSI value chain phases.

Table 3. OD capabilities matrix

General Business Capabilities / OD Capability Areas	Data generation	Data collection, aggregation and processing		Data distribution and delivery			Final data use	
	Generating data	Data processing	Data storage and computing facilities	Data release	Providing access to data and APIs	Publishing solution	Data retrieval	Data usage
Individual/competences								
Business process								
Organization								
IT infrastructure								
Technological								
Management/ governance								

There is only one approach to utilize capability matrix in an OD business. Business managers need to identify specific individual, process, organization, IT infrastructure, technological, and management capabilities for all value chain phases. They need to start from the first phase of value chain and define these capabilities for each capability area separately. For example, the first value chain phase is *Data Generation* and this includes set of capabilities required for generating data. Managers should identify what are the individual, process, organization, IT infrastructure, technological, and management capabilities they need for generating data. OD business management/governance is necessary throughout the value chain.

5   Discussion

Both government and business data are rich sources of potentially valuable but currently relatively untapped resource as the data is frequently locked up by the data holders. According to existing literature, considerable OD efforts are largely technology driven. OD publishing process shows that OD publishing requires capabilities for collecting and generating of data, processing, securing privacy and the development of standards for publishing and use of data. Transformation requires the development of these capabilities and development of capabilities require business management team to understand them. Our work has revealed ten OD capabilities areas which have been structured based on the OD value chain. These capability areas contain a number of capabilities which we deconstruct based on the six classical and domain-agnostic capability types already identified in literature.

  As a framework, the question about the validity of the framework naturally arises. Given that it was constructed by analyzing capabilities described in existing literature and initiatives, the coverage (or content validity) of the model linked to how exhaustive our review was or to the extent to which, the obtained capabilities cover the major OD value chain stages. We have shown in Section 4 that the ten capability areas adequately cover all the OD value chain stages.

  Since there are no OD capability framework in literature, comparison of the developed framework with similar frameworks is not possible. However, detailed specification of the framework elements for pragmatic or tooling purpose is the next logical step.

  In addition to serving as OD capability planning tool, the framework could also serve as tool for benchmarking or measuring OD capabilities in organizations.

6   Conclusion

A number of list of political, social and economic benefits have been associated with the use and impact of OD. The economic aspect of OD has naturally generated a lot of interest resulting in a number of OD business models. In this paper, we developed an OD capability matrix as a tool to support design of OD business capabilities. In addition, the OD Capability Matrix could help the implementation of OD business

models. Matrix can assist OD business managers to understand and describe how capabilities should be utilize and extended throughout the OD value chain of the business. As there is no OD capability framework in literature, our OD Capability Matrix provides a significant starting point for OD businesses to plan and develop the requisite capabilities to support their business models. With time, concrete experience from practice will be useful in refining the capability framework.

Regarding future work, our main interest is to develop the framework into a concrete tool (similar to the Business Model Canvas) to support OD practitioners. In this regard, business modelers and managers are encouraged to utilize the Matrix.

From the research perspective, we intend to refine the capability matrix to reflect maturity levels of OD capabilities. To represent the matrix as a modeling technique, a potential future work would be to offer more specific guidelines for designing the matrix. Moreover, we are considering the possibility of generation of OD capability patterns from the OD capability matrix. Another potential future work would be to study how capability driven development approach can support evolving OD businesses and facilitates adjustment of capabilities according to changing context.

References

[1]   F. G. Goethals, "The Unified Business Model Framework," Lille, 2009.

[2]   R. Casadesus-Masanell and J. E. Ricart, "How to Design A Winning Business Model," *Harward Bus. Rev.*, no. February, 2011.

[3]   T. Davies, F. Perini, and J. M. Alonso, "Researching the emerging impacts of open data ODDC conceptual framework," 2013.

[4]   G.-J. Ren and S. Glissmann, "Identifying Information Assets for Open Data: The Role of Business Architecture and Information Quality," *IEEE 14th Int. Conf. Commer. Enterp. Comput.*, pp. 94–100, Sep. 2012.

[5]   G. Vickery, "Review of Recent Studies on PSI Re-Use and Related Market Development," Paris, 2011.

[6]   A. Zuiderwijk and S. Van Den Braak, "Workshop : Linking open data - Challenges and Solutions," in *13th Annual International Conference on Digital Government Research Workshop*, 2012, pp. 304–305.

[7]   A. Zuiderwijk and M. Janssen, "Open data policies, their implementation and impact: A framework for comparison," *Gov. Inf. Q.*, Dec. 2013.

[8]   J. Manyika, M. Chui, P. Groves, D. Farrell, S. Van Kuiken, and E. A. Doshi, "Open data : Unlocking innovation and performance with liquid information," 2013.

[9]   HM Government, "Seizing the data opportunity: A strategy for UK data capability," UK, 2013.

[10]   IBM Business Consulting Services, "Component business models Making specialization real," 2005.

[11]   T. van den Broek, M. Rijken, and S. van Oort, "Towards Open Development Data," 2012.

[12]   W. Ulrich and M. Rosen, "The Business Capability Map : The ' Rosetta Stone ' of Access to the Experts," *Enterp. Archit.*, vol. 14, no. 2, 2011.

[13]   M. Avital and N. Bjorn-andersen, "The Value of Open Government Data : A Strategic Analysis Framework," in *SIG eGovernment pre-ICIS Workshop*, 2012, no. 2002.

[14]   K. Relations, N. Huijboom, and T. Van Den Broek, "Open data : an international comparison of strategies," 2011.

[15]   N. Julien, "Business Opportunities Arising from Open Data Policies," Imperial College London, 2012.

[16]   Capgemini Consulting, "The Open Data Economy Unlocking Economic Value by Opening Government and Public Data," 2013.

[17]   G. D. Bhatt and V. Grover, "Types of Information Technology Capabilities and Their Role in Competitive Advantage : An Empirical Study," *J. Manag. Inf. Syst.*, vol. 22, no. 2, pp. 253–277, 2005.

[18]   E. Jaques and G. Stamp, "Level and Type of Capability in Relation to Executive Organization," Brunel, 1995.

[19]   S. Steiner, B. Abraham, and J. Mackay, "Understanding Process Capability Indices," Waterloo, Ontario, 1997.

[20]   Symphony Technologies Pvt Ltd, "Measuring Your Process Capability," India.

[21]   V. Ambrosini and C. Bowman, "What are dynamic capabilities and are they a useful construct in strategic management?," *Int. J. Manag. Rev.*, vol. 11, no. 1, pp. 29–49, Mar. 2009.

[22]   S. Mithas, N. Ramasubbu, M. S. Krishnan, and V. Sambamurthy, "Information Technology Infrastructure Capability and Firm Performance : An Empirical Analysis," Michigan, 2009.

[23]   H. Gheysari, A. Rasli, P. Roghanian, and H. Jebur, "The Role of Information Technology Infrastructure Capability (ITIC) in Management," vol. 2, no. 2, pp. 36–40, 2012.

[24]   A. S. Bharadwaj, "A Resource-Based Perspective on Information Technology Capability and Firm Performance: an Imprical Investigation," *MIS Q.*, vol. 24, no. 1, pp. 169–196, 2000.

[25]   W. Xia and W. R. King, "Determinants of Organizational IT Infrastructure Capabilities : An Empirical Study," 2002.

[26]   E. Arnold and B. Thuriaux, "Developing Firms' Technological Capabilities," 1997.

[27]   (BNCSTD) Brazilian National Council for Scientific and Technological Development, "The Relationship between Technological Capability and Firm Performance in an Emerging Economy," Rio, 2011.

[28]   J. F. Rayport and J. J.Sviokla, "Exploiting the Virtual Value Chain," *Harward Bus. Rev.*, 1995.

[29]   G. D. Bhatt and A. F. Emdad, "An analysis of the virtual value chain in electronic commerce," *J. Enterp. Inf. Manag.*, vol. 14, no. 1, pp. 78–84, 2001.

[30]   M. E. Porter, "Competitive Advantage: Creating and sustaining superior performance," New York, 1985.

[31]   B. Ubaldi, "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives," 22, 2013.

[32]   M. de Vries, "14 years of PSI Policy and its Impact," in *Madrid Conference*, 2012, no. February.

[33]   E. Ferro and M. Osella, "Eight Business Model Archetypes for PSI Re-Use," London, 2013.

[34]   J. Musings, "Open Data Business Models," 2012. [Online]. Available: http://www.jenitennison.com/blog/node/172. [Accessed: 06-Feb-2014].

[35]   L. A. R. Rojas, J. M. C. Lovelle, G. M. Tarazona, and C. E. Montenegro, "Open Data as a key factor for developing expert systems: a perspective from Spain," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 2, no. 2, p. 51, 2013.

[36]   A. Osterwalder, "The Business Model Ontology: A Proposition in a Design Science Approach," UNIVERSITE DE LAUSANNE ECOLE, 2004.