

Open Data Capability Architecture - An Interpretive Structural Modeling Approach

Fatemeh Ahmadi Zeleti¹, Adegboyega Ojo^{1,2}

¹Insight Centre for Data Analytics
NUI Galway, Republic of Ireland

²Department of Applied Informatics in Management,
Faculty of Management and Economics,
Gdańsk University of Technology, Poland

{fatemeh.ahmadizeleti, adegboyega.ojo}@insight-centre.org

Abstract

Despite of increasing availability of open data as a vital organizational resource, large numbers of start-ups and organizations fail when it comes to utilizing open data effectively. This shortcoming is attributable to the poor understanding of what types of capabilities are required to successfully conduct data related activities. At the same time, research on open data capabilities and how they relate to one another remains sparse. Guided by extant literature, interviews of these organizations, and drawn from Interpretive Structural Modeling (ISM) approach which are pair comparison methods to evolve hierarchical relationships among a set of elements to convert unclear and unstructured mental models of systems into well-articulated models that act as base for conceptualization and theory building, this study explores open data capabilities and the relationships and the structure of the dependencies among these areas. Findings from this study reveal hitherto unknown knowledge regarding how the capability areas relate one another in these organizations. From the practical standpoint, the resulting architecture has the potential to transform capability management practices in open data organizations towards greater competitiveness through more flexibility and increased value generation. From the research point of you, this paper motivates theory development in this discipline.

1. Introduction

Open data is a vital organizational asset. While new start-ups have began to benefit from the vast potential of this resource [1], large numbers of these organizations fail to effectively use open data and fully leverage its potential [2][3]. Only a small percentage of organizations benefit from using open data and very few organizations can attribute a value [2]. Many do not yet know how to conduct data related activities more effectively [4]. One major reason for failure is the fact that these organizations do not clearly know

what data capabilities they require to effectively exploit open data for their business objectives [5]. As an example, an exploratory research of 33 Open Data Organizations (ODOs) in UK in 2016 [6] found that data capabilities remain unclear to many of these organizations. Lack of clear understanding of open data capabilities put open data utilizing organizations at huge risk [3]. Studies show that in order to compete and survive in the fast changing and competitive open data industry, ODOs are required to develop capabilities for generating value from open data, increasing agility and competitive advantage [7][3].

While few existing studies have investigated open data capability types and areas [8][9]–[11], to our knowledge, no previous scholarly work has attempted to comprehensively identify and investigate relationships and dependencies between different capability areas and, articulate a capability architecture for ODOs. Many studies [12][13][14] strongly suggest further research into open data capabilities and their dependencies. Therefore, in this study, we attempt to robustly address this research gap driven by two research questions (RQ): RQ1) what are the main elements of the open data capability architecture and RQ2) how do these elements relate to each other?.

To develop open data capability architecture, we adopt the ISM technique, which is a well-established methodology for developing relationships within a system of related elements [15]. We built upon existing studies on open data capabilities and included in this study expert experiences of Chief Executive Officers (CEOs) of 11 ODOs to explore relationships between different areas.

In this research, we define *open data organizations* as both non-profit and for-profit organizations that use, produce, or otherwise invest in open data as a key aspect of their operation for generating customer value and achieving organization's mission goals. In addition, adopting IEEE-1471 definition of architecture and the capability architecture definition given by [16], we propose that *open data capability architecture* identifies open data capabilities that are required to

support the organizational mission goals and provides a common language and framework to understand how to do things in ODOs to effectively harness the real value of data. The architecture also provides interdependencies or relationships between different capability areas.

2. Background

The background study of this research includes investigation of different types and areas of open data capabilities. Based on the well-known edicts of Capability-based-View and Dynamic-Capability Theory, we have identified three types of organizational capability: 1) *Value capabilities*, 2) *Dynamic capabilities*, and 3) *Competitive capabilities*. Below we present the three types and their associated areas.

Value Capabilities - The creation of customer 'value' is the key in every organization [17]. Value capabilities are source of value and can assist the organization to generate value proposition for the customers. While value capabilities are not the source of competitive advantage, they are necessary to produce customer value [18]. For value generation from open data, an analytical report by the European Data Portal [19] reports on four areas of capabilities including: 1) technical capabilities, 2) statistical capabilities, 3) analytical capabilities and personality, and 4) business insight and domain knowledge. Moreover, HM Government [13] highlights three overarching aspects to data capability which include 1) human capital, 2) tools and infrastructure, and 3) data re-use. In [8] and [20], authors developed an empirical investigation of ODOs to synthesize and introduce five value creating capabilities which are necessary for every ODOs. The five open data capability areas for generating value from open data include: 1) Open Data Individual Competences and Expertise, 2) Open Data Processes, 3) Open Data Organization, 4) Open Data IT and Technological Infrastructures, and 5) Management Capability and Data Governance.

Dynamic Capabilities - The majority of the studies on dynamic capability assert that dynamic capabilities are the ability of the organization to renew its capabilities to deal with rapidly changing environments (Helfat and Peteraf 2003). [21] defines dynamic capabilities as "a firm's capacity to deploy resources, usually in combination, using organizational processes, to effect a desired end". According to [22], in dynamic markets, it makes sense to use dynamic capabilities to build new resource configurations and move into new competitive positions. [22] further highlight that jettisoned resource combinations that no longer provide competitive advantage are also critical

capabilities as markets undergo change. According to [23] "dynamic capabilities are built rather than bought in the market", and they include organizational processes or 'routines' that are employed to reconfigure or to combine the organization's resources and paths which are the choices open to the organization today and likely to be in the future. [24] suggest that whilst market-based learning enables the organization to learn what the market needs, the organization must acquire knowledge from other sources to develop leading edge innovative products and services that will fulfill organization's needs. However, as highlighted by [25] and [26], the network relationship specifically with the suppliers plays a significant role in enhancing the supplier network, sensing and seizing opportunities, knowledge creation, resource configuration and integration and know-how exchange. Despite earlier studies on dynamic capabilities and how to achieve agility, very limited studies exist to investigate dynamic capabilities for ODOs. Empirical studies of open data dynamic capabilities in ODOs in [20] and [11] revealed five capability areas that facilitate and improve agility in ODOs. The areas include: 1) Open Data Process Innovation, 2) Open Data Knowledge Management and Organizational Learning, 3) Open Data Value Chain Performance, 4) Open Data Relationship Infrastructure, and 5) Management Functions of ODOs.

Competitive Capabilities – Competitive capabilities foster the organization's competitive advantage and allow organizations to stay competitive and outperform competitors [18] [27]. Although value and dynamic capabilities are important, we also need to recognize the role that the competitive external environment plays in the competitive advantage of ODOs [10]. Insights from a scenario, interviews, and a survey study [1] reveals three areas where the ODOs need to build capability to gain competitive advantage: 1) Information Technology, 2) Information and Data, and 3) Human. The Boston Consulting Group's study [28] also reveals six level open data capability areas in three top-level components: 1) Data Usage, 2) Data Engine, and 3) Data Ecosystem. While not many related works exist to study open data capabilities which are source of competitive advantage, [20] and [10] identified and described four dimensions of competitive capabilities required for ODOs. These are capability areas related to: 1) Open Data Enterprise Infrastructure, 2) Open Data Product and Service, 3) Open Data Business Development, and 4) Open Data Relational Rent. According to [10], open data competitive capability has strategic nature as the level of dynamics in the external environment increases it provides strategic alternatives that will differentiate the organization from



its competitors. Table 1, presents the open data capability areas identified from the literature.

Table 1. Open data capability areas from literature

Capability No.	Open Data Capability Types and Areas
Value Capabilities	
C1	Individual Data Competences and Expertise
C2	Open Data Processes
C3	Organization
C4	IT and Technological Infrastructure
C5	Management and Data Governance
Dynamic Capabilities	
C6	Data Process Innovation
C7	Knowledge Management and Organizational Learning
C8	Data Value Chain Performance
C9	Relationship Infrastructure
C10	Management Functions
Competitive Capabilities	
C11	Enterprise Infrastructure
C12	Data Products and Services
C13	Business Development
C14	Relational Rent

3. Methodology

3.1. Case selection

The selection criteria include ODOs that: 1) rely on open data as one of their main operational resources to achieve organization's mission goals, 2) the application of open data is primarily in developing new products and services, and 3) has long history of practice. Selected organizations must meet all the three criteria to be included in this study. Shortlisted organizations lead to diversity in our cases [29] in terms of the followings: geographical location, organizational size, sector, stream, data domain, data types, and data applications (see Table 2). Diverse cases bring unique perspective which enriches our understanding of open data capability architecture in which in our future research can lead to generalization [29].

Table 2. Organization's brief profile

Location	Size	Sector	Data Domain	Data Types	Data Applications
A Ireland	1-10	Profit Midstream	All (Client's Need)	Linked Open Data	New Product, Service, and Application
B USA	1-10	Profit Midstream	IoT, Consumer, Smart City	Open Data	New Product, Service, and Application
C UK	1-10	Profit Downstream	All (Client's Need)	Open Data	New Product, Service, and Application, Optimization, Advocacy, Strategic
D Belgium	1-10	Profit Midstream	Governance	Linked Open Data	Strategic
E UK	1-10	Profit Midstream	All (Client's Need)	Open Data	New Product, Service, and Application
F Belgium	1-10	Profit Midstream	Business, Research, and Consulting	Proprietary Data, Open Data	Organization Optimization
G Germany	1-10	Profit Downstream	Health	Open Data	New Product, Service, and Application
H USA	11-50	Non-Profit Upstream	All (Client's Need)	Open Data	New Product, Service, and Application, Optimization, Advocacy, Strategic
I UK	1-10	Non-Profit Midstream	Business, Research, and Consulting	Open Data	Advocacy
J Spain	51-200	Non-Profit Midstream	Data Research and Innovation	Open Data	New Product, Service, and Application
K Ireland	51-200	Non-Profit Upstream	Environment	Open Data, Linked Data	New Product, Service, and Application, Optimization, Advocacy, Strategic

3.2. Data collection and interviews

In the absence of adequate literature, the nature of the associations between the capability areas and the level of criticality in practice are not well understood. Therefore, to understand whether or not a relationship

exists between any two capability areas, we have included expert experiences of CEOs of ODOs.

To know and identify ODOs around the world, we request to access the third party's dataset of 685 ODOs. In 2016, this dataset was the largest and fast growing dataset of organizations using open data around the world. We shortlisted 43 organizations that meet all the criteria. Initial invitation email was sent to all 43 organizations informing them about the aim of the study and our purpose to conduct interviews. A follow up email was sent to all in two rounds, each within two weeks period. 11 organizations showed interest to participate in this study. We have emailed each participant, a document containing the cover letter, purpose of the study and interview, expected timing, interview structure, pattern of the interview questions, definition of concepts used in the interview, and trigger words and examples for each definition of capability area to ensure that all participants understand different concepts in a unified manner. Interested organizations were asked to provide their availability. Experiences were collected through one-to-one interview session with each expert [30]. The interview was designed to take no more 75 minutes and GoToMeeting application was used for all interviews.

The interview protocol lists the lines of inquiry that we wanted to explore in the course of an interview. The protocol served as a checklist during the interview to make sure that all relevant questions or issues are covered within the limited time of the interviewees as well as allowing individual perspectives to emerge [31]. The interview protocol developed series of questions categorized in three main parts: 1) Organizational background (questions such as type of the organization, sector of operation, business model and etc.), 2) Application of open data (questions around the use of open data in the organization such as the open data as the key resource, specific purpose to use open data, type of open data and etc.), and 3) Open data capabilities (this main part include questions to investigate adopted value, dynamic, and competitive data capabilities in the organization and possible relationships between them). Before the interview begins, permission was granted from interviewees and all interviews have been voice recorded. To prepare the interview protocol, we followed number of guides including: 1) Using scripts at the beginning and end of the interview, 2) Having open ended questions and free discussions, 3) Starting with basics and a more easy to answer questions, and 4) Using prompts for each question that can help us ensure that the used concepts and questions are understood by all the interviewees in a unified manner.

3.3. Coding and analysis

The lead researcher prepared single transcript document for each interview. Then, the two researchers involved in this study coded and analyzed all the interviews transcripts. For coding, NVivo was used which is a strong and comprehensive qualitative data analysis software platform which can be used to organize and analyze any types of qualitative data [32][33] and to “obtain rigor in dealing with such data” [33]. In this study, to capture all possible details, we analyzed data that shows that a relationship exist between any two capability areas. To code the possible relationships between the capability areas, we developed logic for coding. For example, we coded all the data that shows that a relationship exist between Capability A and Capability B under ‘A impacts B’ Relationship Node in NVivo. Through the coding process, we 1) select a particular expert experience and 2) assign this fragment to a specific relationship node. This allows us to investigate all the possible relationships between any two capability areas based on the experiences of experts. In addition, to increase rigor and prevent bias during coding, involved researchers independently conducted coding three times. We were not given consent to share the data for any purposes including analysis with researchers other than those involved in this study. To reduce data analysis bias, we have been open to all kinds of viewpoints that would ultimately help us to take better decisions and we have avoided analyses that only favor this study.

3.4. Modeling

For modeling our dependency structure of different open data capability areas, we adopted ISM methodology. ISM is one of the most popular techniques for identifying and understanding mutual relationship among the elements or variables of a particular context [34] and is able to uncover hidden relationships between elements far more accurately than individual’s experiences. Because people are limited in their ability to address complex issues involving a significant number of elements at one time, the use of ISM can advance the collective understanding of such relationships [35]. It is used to establish dependencies among elements and to develop a dependency structure of a set of elements.

Followings are the steps for ISM methodology as discussed by Sushil [30][15], Jain and Raj [34] and Sindhwani and Malhotra [36]. Step 1: Identification of elements (open data capability areas) from the literature; Step 2: Identification of contextual relationship between the elements (driven by the RQs, “impact” is identified and later confirmed through the interviews); Step 3: Development of Structural Self-

Interaction Matrix (SSIM); Step 4: Development of Adjacency Matrix [35] or First Reachability Matrix (FRM) [34]; Step 5: Development of Reachability Matrix (RM) and transitivity check (an algorithm-based process developed and implemented by the researchers in MATLAB); Step 6: Level partitioning; Step 7: ISM Digraph is drawn; and Step 8: The final ISM model is drawn. Each of these steps is further illustrated in section 4.

4. Eliciting Relationships between Open Data Capability Areas

The various ISM steps, which lead to the development of the open data capability architecture, are illustrated below:

Step 1: Identifying Open Data Capability Areas

Open data capability areas have been identified through literature survey as shown in Table 1.

Step 2: Identifying Contextual Relationship and Interpretation

There is a need to find a type of contextual relationship between the capability areas [30]. The contextual relationship in this study is “impacts” which is identified from the literature review. The contextual relationship could be for example “capability area A impacts capability area B”. Contextual relationship was discussed and confirmed during the interviews session with the experts. Through ISM, the contextual relationship interprets the nature of the relationship and how that relationship really works [34].

Step 3: Structural Self-Interaction Matrix (SSIM) – Pairwise Comparison

Keeping in view the contextual relationship, the existence of a relation between any two capability areas (i and j) is checked from the coded data (experts’ experiences). Capability areas are aligned in row and column format where areas in row are denoted by “ i ” and those in column are denoted by “ j ”, the matrix hence obtained is the result of each pair of capability area being analyzed separately. SSIM is shown in Table 3. Four symbols are used to denote the nature of possible relationship between any two capability areas i and j (see Sushil [30][15] for the meaning of symbols).

Table 3. Structural Self-Interaction Matrix

SSIM		j elements													
		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
i elements	C1	0	0	0	V	0	0	X	0	0	0	0	0	0	0
	C2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	C3	0	0	0	0	0	0	A	0	0	0	0	0	0	0
	C4	0	0	0	0	X	0	X	0	0	0	V	0	0	0
	C5	0	0	0	X	0	0	0	0	0	0	0	0	0	0
	C6	0	V	0	0	0	0	A	X	0	0	0	X	X	X
	C7	X	0	V	X	0	V	0	0	V	0	0	0	0	A
	C8	0	0	0	0	0	X	0	0	0	0	0	0	0	0
	C9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	C10	0	0	0	0	0	0	0	0	0	0	V	0	V	0
	C11	0	0	0	0	0	0	0	0	0	A	0	0	A	0
	C12	0	0	0	0	0	X	0	0	0	0	0	0	A	0
	C13	0	0	0	0	0	X	0	0	0	A	V	V	0	0
	C14	0	0	0	0	0	X	V	V	0	0	0	0	0	0

Step 4: Adjacency Matrix or First Reachability Matrix (FRM)

Table 4 presents the FRM in which '0' indicates no relationship and '1' indicates relationship exists between *i* and *j* capability areas. The information in SSIM coded in V, A, X, O is transformed into 0s and 1s based on a set of rules (see Sushil [30][15] for the set of rules).

Table 4. First Reachability Matrix

FRM		elements													
		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
elements	C1	1	0	0	1	0	0	1	0	0	0	0	0	0	0
	C2	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	C3	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	C4	0	0	0	1	1	0	1	0	0	0	1	0	0	0
	C5	0	0	0	1	1	0	0	0	0	0	0	0	1	0
	C6	0	0	1	0	0	0	1	0	0	0	0	0	1	1
	C7	1	0	1	1	0	1	1	0	1	0	0	0	0	0
	C8	0	0	0	0	0	1	0	1	0	0	0	0	0	0
	C9	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	C10	0	0	0	0	0	0	0	0	0	1	1	0	1	0
	C11	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	C12	0	0	0	0	0	1	0	0	0	0	0	1	0	0
	C13	0	0	0	0	0	1	0	0	0	0	1	1	1	0
	C14	0	0	0	0	0	1	1	1	0	0	0	0	0	1

Note: For completeness, we define the (i, i) entry in the Adjacency Matrix as 1.

Step 5: Transitivity Check and Final Reachability Matrix (RM)

The transitivity of the relationship states that if element "A" is related to element "B," and "B" is related to element "C," then "A" is necessarily related to "C". The Final Reachability Matrix (RM) indicates whether a column variable can be "reached" from a row variable along a continuous, directed path [35]. RM is obtained from performing the transitivity check over the FRM [37], as shown in Table 5.

Table 5. Reachability Matrix

RM		elements														Driver
		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	
elements	C1	1	0	1*	1*	1*	1*	1*	0	0	1*	1*	1*	1*	1*	11
	C2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
	C3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
	C4	1*	0	1*	1	1*	1*	1*	0	0	1*	1*	1*	1*	1*	11
	C5	1*	0	1*	1	1*	1*	1*	0	0	1*	1*	1*	1*	1*	11
	C6	1*	1	1*	1*	1*	1*	1*	0	0	1*	1*	1	1	1	12
	C7	1	0	1	1*	1	1*	1	1*	1	0	1*	1*	1*	1*	12
	C8	1*	0	1*	1*	1*	1*	1*	1	0	1*	1*	1*	1*	1*	11
	C9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
	C10	1*	0	1*	1*	1*	1*	1*	0	1	1	1*	1	1*	1*	12
	C11	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
	C12	1*	0	1*	1*	1*	1*	1*	0	0	1*	1	1*	1*	1*	11
	C13	1*	0	1*	1*	1*	1*	1*	0	0	1	1	1	1*	1*	11
	C14	1*	0	1*	1*	1*	1*	1	1	1	0	1*	1*	1*	1	11
Dependent		10	2	11	10	10	10	10	2	1	11	10	10	10	117	

Note: * denotes the values which are changed from "0" to "1" during transitivity check in MATLAB

Step 6: Level Partitions on the Reachability Matrix

Level Partitioning is mandatory to identify the levels of each element (resulting the hierarchical model) affecting any system or problem. ISM-based level partitioning is obtained from the RM [36]. The RM must be processed to form different levels. From the RM, we determine the reachability set (RS), antecedent set (AS), and Interaction Set (IS) $RS \cap AS$ for all the elements [15][37].

The element for which the RS and IS are the same is the top-level elements in the ISM Digraph [35]. For example, the first row of the RM shows that C1 occurs for all variables except C2, C9, and C10, indicating that C1 can reach or impact all capability areas except

C2, C9, and C10. Therefore, in this case, the $RS(C1) = \{1,3,4,5,6,7,8,11,12,13,14\}$. Similarly, in the first column, C1 occurs for all variables except C2, C3, C9, and C11, indicating that C1 can only be reached or impacted by itself, C4, C5, C6, C7, C8, C10, C12, C13, and C14. Thus, the $AS(C1) = \{1,4,5,6,7,8,10,12,13,14\}$. The IS of the RS, and the AS (the common elements in both sets), results in $RS \cap AS = \{1,4,5,6,7,8,12,13,14\}$.

Once the top-level element/s is/are determined, it is separated out from the other elements. The same process is then repeated to constitute the next level [35]. This iteration continues until the level of last remaining element is derived from the process [37]. The elements in the top level of the hierarchy will not reach any elements above their own level [15].

This analysis was completed in three iterations (as shown in Table 6) which means that analysis revealed three levels of the ISM digraph. First iteration suggests Level 1 = {C2, C3, C9, C11}, second iteration suggests Level 2 = {C1, C4, C5, C6, C7, C8, C12, C13, C14}, and the third and the last iteration suggests Level 3 = {C10}.

Table 6. Level Partitioning

Iteration	Element No.	Reachability Set (RS)	Antecedent Set (AS)	Intersection (IS) IS = RS ∩ AS	Set (L _i)	Level (L _i)			
Iteration 1	C1	1,3,4,5,6,7,8,11,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14	L1={C2, C3, C9, C11}	1			
	C2	2	2	2					
	C3	3	1,3,4,5,6,7,8,10,12,13,14	3					
	C4	1,3,4,5,6,7,8,11,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C5	1,3,4,5,6,7,8,11,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C6	1,3,4,5,6,7,8,11,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C7	1,3,4,5,6,7,8,9,11,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C8	1,3,4,5,6,7,8,11,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C9	9	9	9					
	C10	1,3,4,5,6,7,8,10,11,12,13,14	10	10					
	C11	11	1,4,5,6,7,8,10,11,12,13,14	11					
	C12	1,3,4,5,6,7,8,11,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C13	1,3,4,5,6,7,8,11,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C14	1,3,4,5,6,7,8,11,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
Iteration 2	Element No.	Reachability Set (RS)	Antecedent Set (AS)	Intersection (IS) IS = RS ∩ AS	Set (L ₂)	Level2 (L ₂)			
	C1	1,4,5,6,7,8,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14	L2={C1, C4, C5, C6, C7, C8, C12, C13, C14}	2			
	C4	1,4,5,6,7,8,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C5	1,4,5,6,7,8,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C6	1,4,5,6,7,8,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C7	1,4,5,6,7,8,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C8	1,4,5,6,7,8,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C10	1,4,5,6,7,8,10,12,13,14	10	10					
	C12	1,4,5,6,7,8,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C13	1,4,5,6,7,8,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	C14	1,4,5,6,7,8,12,13,14	1,4,5,6,7,8,10,12,13,14	1,4,5,6,7,8,12,13,14					
	Iteration 3	Element No.	Reachability Set (RS)	Antecedent Set (AS)			Intersection (IS) IS = RS ∩ AS	Set (L ₃)	Level3 (L ₃)
		C10	1,4,5,6,7,8,12,13,14	1,4,5,6,7,8,10,12,13,14			1,4,5,6,7,8,12,13,14	L3={C10}	3

Step 7: Digraph for ISM

The ISM model or Digraph (Directed Graph) is useful to interpret the dependencies between all capability areas and hierarchies pictorially [35]. The Digraph is generated from the level partitions and RM [37]. First, the elements are arranged graphically in levels based on the analysis in step 6. Second, the directed relationships are drawn as per the relationships shown in the RM (*i-j* link with "1" entry). In case of ISM, all the transitive links (*i-j* link with "1*" entry) are dropped in the digraph [30]. The model is generated using nodes or vertices and lines connecting the nodes, as shown in Figure 1. The lines depicts relationship or dependencies [37]. Once the transitive relations are

removed from the digraph the resultant Digraph is the initial ISM Digraph.

Step 8: ISM Final Model – Open Data Capability Architecture

The open data capability architecture which is a structural model is derived from the connective information contained in the digraph. The details of capability areas are indicated in the respective boxes with indicated relations as worked out in the digraph. The architecture depicts the open data capability areas and their reachability to the higher level capability areas and, provides a clear picture with an understanding of the dependencies among the capability areas. The open data capability architecture is shown in Figure 2.

5. Findings and Interpretation

The main finding of this research is the open data capability architecture shown in Figure 2. The architecture emerged from ISM analysis. We summarize and organize the findings based on the two research questions.

RQ1. What are the elements of the Open Data Capability Architecture?

Answer to this question identifies and improves our understanding about open data capability areas for generating value from open data, improving agility and competitive advantage in ODOs. The resulting open data capability architecture shows 14 elements. Capabilities required to generate value from open data include: Individual Competences and Expertise; Open

Data Processes; Organization; IT and Technological Infrastructure; and Management and Data Governance. Capabilities required to enable and improve agility in the organization include: Process Innovation; Knowledge Management and Organizational Learning; Value Chain Performance; Relationship Infrastructure; and Management Functions. Capabilities required that are necessary for competitive advantage include: Enterprise Infrastructure; Product and Service; Business Development; and Relational Rent. Two research findings stand out in this part which include: 1) while literature indicates the importance of ‘Management Functions’ capability area on the agility of the organization only [25][24], findings suggest that, this area is the most influential capability area in ODOs, but are relatively underdeveloped in these organization and 2) looking at the resulting open data capability architecture shows that many open data capability areas - from all the three capability types - are placed at the middle layer of the architecture which indicates that, these capability areas are factors of instability, since any action towards them has consequences not only on them but also on other capability areas.

RQ2. What is the relationship between the elements of the Open Data Capability Architecture?

Answer to this question identifies relationships between open data capability areas and improves our understanding on how these areas influence one another and, what are the dependencies between them. The resulting open data capability architecture

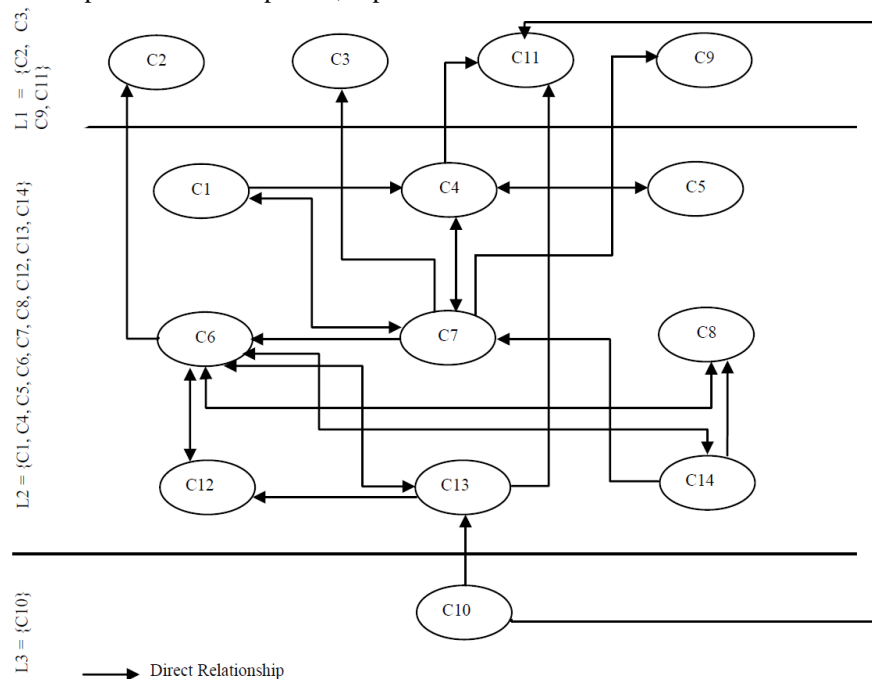


Figure 1. ISM digraph

identifies 26 relationships in which each indicates that a direct relationship exists between two particular open data capability areas. One major finding stand out in this part which includes: 1) open data capability areas are dependent on one another both within one capability type and across types. This means that, the three types are related to one another and they can influence development of one another. As a result of this dependency, our finding shows that, clear structure in the relationships exists. In ODOs, capabilities can be developed as the result of these dependent relationships and the two-ways communication among them. This communication between capability areas enables capability implementation loop where capabilities are assessed and innovated to improve agility in the ODOs. In addition, implementation loop generates knowledge and improve learning within the organization which can also contribute to the agility of the organization. As demonstrated in our architecture, Knowledge Management and Organizational Learning capability is at the very centre and has many relations to other capability areas. This can also generate and improve feedback loop where knowledge is actively generated and used to improve open data capabilities toward achieving organization's mission goals.

Open data capability architecture shows that, 'Management Functions' as one influential dynamic capability plays a critical role in enabling and improving agility in ODOs. We argue that, traditional management approach where management power is centralized cannot adequately tackle the open data market agility. Now more than ever management teams in ODOs are required to develop innovative approaches and encourage inter- and intra organizational collaborative environment where knowledge, creative ideas and power can be freely shared and used within the organization. We further argue that, management agility as a capability has been of concern within the context of open data. We believe that management capability in the area of agility is poorly developed. As many new start-ups are emerging to tap into the vast potentials of open data, data collected from the interviews show that these relatively young managers may not have the adequate levels of knowledge and understanding about management capabilities to deal with the fast changing demands in the open data market in particular. Consequently, when the management teams in the ODOs fail to address the required changes effectively, the changes may turn out to be crises, which could lead to organizational failure. However, when decisions regarding the needed changes are made by a competent management, they may conveniently be used to improve capabilities in

6. Discussion

6.1. Implications of findings

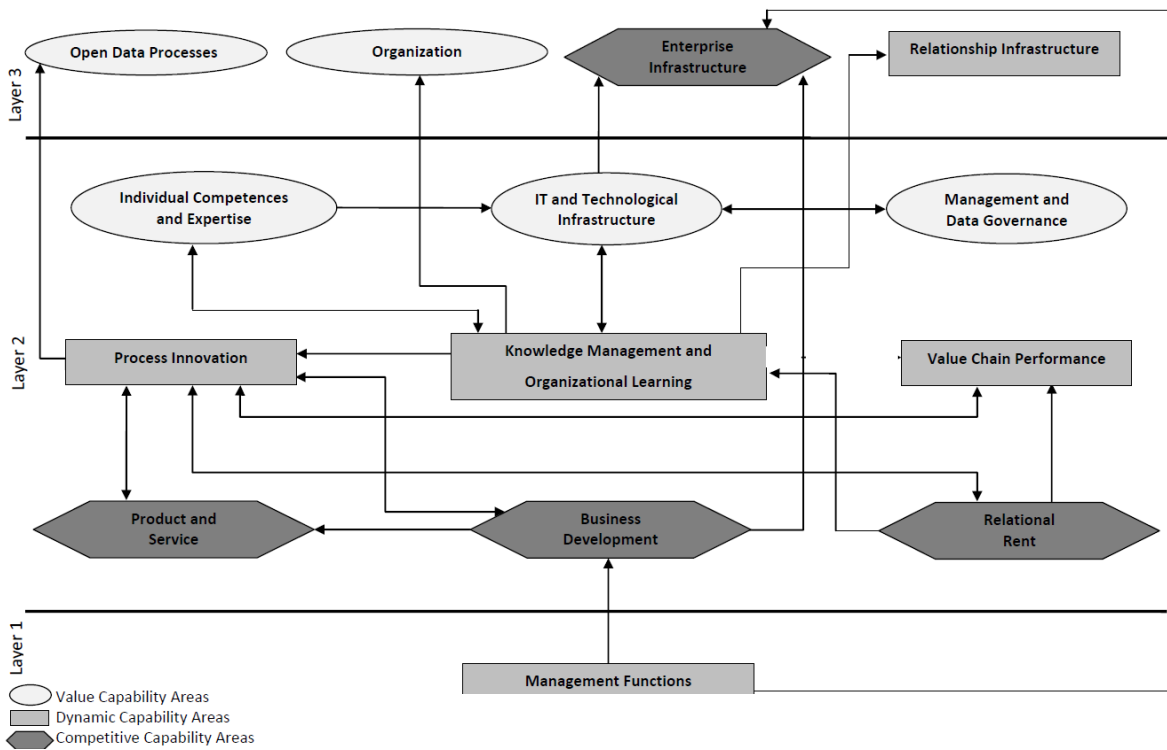


Figure 2. The Open Data Capability Architecture

the organization. We state that, this could be one key reason why many of these start-ups fail when they are still at their early stage. *As a result and supported by the developed architecture in this study, we argue that, Management Functions capabilities serve as 'input' to the development of open data capabilities in the ODOs.*

As highlighted in our findings section, nine open data capability areas located in the middle layer of the architecture are factors of instability in the ODOs. This means that, any change towards them has consequences beyond them to other capability areas and also on the produced open data products and services. *Hence, based on our findings, we claim that, organizational agility is established and improved in this layer where, different capability areas are impacting each other to enable development of valuable open data products and services. Therefore, we have labelled this layer as the 'interaction' layer of the open data capability architecture.* In addition, we argue that, considering the importance of this layer, it is very vital that management team is capable of providing sufficient inputs to enable agility by adopting and developing appropriate innovative capabilities.

Top layer of the architecture is where the final transformation happens. Therefore, capability areas located in this layer are the 'outcomes' of this effort. Taking into an account their autonomous nature, they still get influenced by the interactions and the feedback loop that are established in the middle layer. In Figure 3, we illustrate the interaction between the three layers of the open data capability architecture.

6.2. Relating findings to some existing works

Insights from a scenario, interviews, and a survey study [1] reveals three areas where ODOs need to build capability: 1) IT: having access to the Internet,

adopting cloud computing to run software and applications without having to own, manage, and operate the internal cloud resources and capabilities, and adopting tools for data processing, linking, cleaning, and other tools; 2) Information and Data: having database with open dataset for commercial open data use, developing company database to obtain commercial benefits, and using data to develop new company products and services; 3) Human: computer skills, finding and accessing open data, tool selection and use, data and result interpretation, and stakeholder network management. The resulting open data capability architecture and findings from our study confirm the importance of the three capability areas and adds to it by outlining other open data capability areas that are equally important to the competitiveness of ODOs. In addition, the relationships identified in our study show that there is a relationship between the three capability areas presented in study by Zuiderwijk et al. [1].

Agarwal et al. (2014) [28] claim that by developing three core capability components organizations can put in place a framework for enabling and succeeding with data and big data. The three components are: 1) Data Usage (Identifying Opportunities and Building Trust), 2) Data Engine (Laying the Technical Foundation and Shaping the Organization) and 3) Data Ecosystem (Participating in a Big-Data Ecosystem and Making Relationships work). We confirm that the reported capabilities in Agarwal et al. [28] can contribute to realizing the potential of open and big data. In this regard, Identifying Opportunities and Building Trust can both relate to Business Development Strategic capabilities; laying the Technical Foundation and Shaping the Organization can each relate to Technological Infrastructure Capabilities and Organization capabilities;

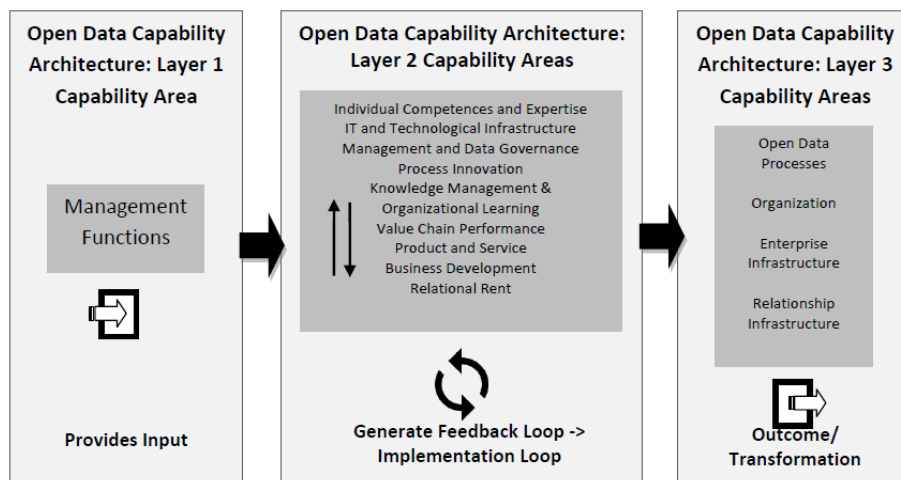


Figure 3. Communication between the three layers of the architecture

Participating in a Big-Data Ecosystem and Making Relationships Work can each relate to Knowledge Management and Organizational capabilities and Relational Rent Strategic capabilities in our work.

In the empirical study completed by Dremel et al. (2017) [38], a capability model was developed for big data analytics to address all relevant facets of a company that performs big data analytics to deliver new products and services or to improve existing ones. The model includes eight capability areas or what the authors called as 'competence fields'. The eight areas are: 1) Customer Relationship Management, 2) Partner Life Cycle Management, 3) Product/Service Life Cycle Management, 4) Enterprise Risk Management, 5) Strategy Development, 6) Transformation Competence, 7) Enterprise Architecture and 8) Process Management, and Information Management. Authors highlighted that the capability model provides a generic library of capabilities that can be used to assess a company's ability to successfully perform big data analytics. However, the model comprises no relationship or link between the capabilities. Despite the differences that exist in open data and big data domains, this study [38] and our thesis develop some similarities and overlapping elements or capabilities. Customer Relationship Management and Partner Life Cycle Management can be interpreted as Relationship Infrastructure Capabilities; Product/Service Life Cycle Management can be interpreted as Open Data Value Chain Performance; Both Enterprise Risk Management and Strategy Development can fit into Open Data Business Development Strategies and Other Strategic Capability Areas; Transformation Competence could be similar to Knowledge Management and Organizational Learning capabilities and other capabilities associated with Dynamic Capabilities; Enterprise Architecture can be categorized as Organizational Capabilities; and Process Management and Information Management can fit into Open Data Process Capabilities. The authors can use the relationships identified in this study to refine their model.

Nevertheless, we understand that no existing study developed open data capability architecture with clear link and dependencies between the areas.

7. Conclusion

Although open data belongs to one of the most intensively discussed topics today, few research efforts have investigated the capabilities required for generating value from open data, improving agility and competitive advantage in organizations utilizing this resource to meet their mission goals and, no

study has attempted to articulate open data capability architecture.

Building upon the identified capability areas from the related literature and expert experiences collected from the CEOs of 11 ODOs, we have developed open data capability architecture which gives an initial, yet unique and empirically grounded view of the capabilities that ODOs require to generate value from open data, improve agility and obtain competitive advantage. The architecture includes main open data capability areas, relationships between the capability areas and, clear dependency structure between the areas.

The developed architecture will help ODOs and organizations or start-ups whose aim is to use open data to meet their business objectives to understand open data capabilities, how different capability areas are related to one another and their dependencies. By utilizing and implementing the open data capability architecture, ODOs can create a solid foundation for effectively harnessing open data.

This study does not claim to be completely exhaustive. The findings are largely based on the case analysis of the interviewed ODOs. To empirically generalize conclusions, the research of other organizations in open data industry is needed. Therefore we are unable to claim that factors beyond the scope of this study will not have an influence on capability development and competitive advantage of ODOs. In addition, through multiple data coding iterations, we avoid any data coding bias in this research and, to reduce data analysis bias, we have been open to all kinds of viewpoints that would ultimately help us to take better decisions. However, we recognize that possible bias may exist for for-profit ODOs due to the higher number of organizations participating in this study. The findings of this study were reviewed by peers to provide confirmation that these research conclusions are sound and reasonable given the data.

We anticipate that future research will aim to 1) take into consideration the transitive relationships and develop interpretation of each transitive relationship, 2) apply MICMAC technique (applied widely with ISM) to classify capability areas as driver, linkage, dependent and autonomous areas [30], 3) categorize capability areas based on the different application of open data (organizations using open data, producing open data, and investing in open data) and 4) to develop a better understanding of the nature of organizations who had conflicting experiences.

8. References



- capabilities: An exploration of how firms renew their resource base," *Br. J. Manag.*, vol. 20, no. SUPP. 1, pp. 1–41, 2009.
- [1] A. Zuiderwijk, M. Janssen, K. Poulis, and G. van de Kaa, "Open Data for Competitive Advantage: Insights from Open Data Use by Companies," in *Proceedings of the 16th Annual International Conference on Digital Government Research (dg.o 2015)*, 2015, pp. 79–88.
 - [2] C. Reid, R. Petley, J. McClean, K. Jones, and P. Ruck, "Seizing the Information Advantage: How organizations can unlock value and insight from the information they hold," 2015.
 - [3] HEDIIP, "Data Capability: A call to action," UK, 2016.
 - [4] M. Janssen and a. Zuiderwijk, "Infomediary Business Models for Connecting Open Data Providers and Users," *Soc. Sci. Comput. Rev.*, Mar. 2014.
 - [5] J. M. Gathungu and J. K. Mwangi, "Dynamic capabilities, talent development and firm performance," *DBA Africa Manag. Rev.*, vol. 2, no. 3, pp. 83–100, 2012.
 - [6] J. Fawcett, "Examining open data as a source of competitive advantage for big businesses," in *Open Data Research Symposium*, 2016.
 - [7] Booz Allen Hamilton, "Tips for Building a Data Science Capability," Hamilton, 2015.
 - [8] F. Ahmadi Zeleti and A. Ojo, "Open data value capability architecture," *Inf. Syst. Front.*, 2016.
 - [9] F. Ahmadi Zeleti and A. Ojo, "Capability Development in Open Data-Driven Organizations," in *Government 3.0 – Next Generation Government Technology Infrastructure and Services*, 1st ed., A. Ojo and J. Millard, Eds. Springer, Cham, 2017.
 - [10] F. Ahmadi Zeleti and A. Ojo, "Competitive Capability Framework for Open Government Data Organizations," in *Proceedings of the 18th Annual International Conference on Digital Government Research*, 2017, pp. 250–259.
 - [11] F. Ahmadi Zeleti and A. Ojo, "Critical Factors for Dynamic Capabilities in Open Government Data Enabled Organizations," in *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*, 2016, pp. 86–96.
 - [12] The World Bank, "Open Data for Business (OD4B) Tool," Washington DC, 2015.
 - [13] HM Government, "Seizing the data opportunity: A strategy for UK data capability," UK, 2013.
 - [14] B. Klievink and M. Janssen, "Realizing joined-up government - Dynamic capabilities and stage models for transformation," *Gov. Inf. Q.*, vol. 26, no. 2, pp. 275–284, 2009.
 - [15] Sushil, "Interpreting the interpretive structural model," *Glob. J. Flex. Syst. Manag.*, vol. 13, no. 2, pp. 87–106, 2012.
 - [16] Reply, "Capability Architecture: Mind the Gap!," 2012.
 - [17] A. Lindgreen and F. Wynstra, "Value in business markets: What do we know? Where are we going?," *Ind. Mark. Manag.*, vol. 34, no. 7 SPEC. ISS., pp. 732–748, 2005.
 - [18] 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.
 - [19] E. van S. Wendy Carrara, Sander Fischer, "Analytical report: E-skills and Open Data," Brussels, 2015.
 - [20] F. Ahmadi Zeleti and A. Ojo, "Qualitative Structural Model for Capabilities in Open Data Organizations," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019, pp. 2902–2911.
 - [21] S. Fainshmidt, A. Pezeshkan, M. L. Frazier, A. Nair, and E. Markowski, "Dynamic Capabilities and Organizational Performance: A Meta-Analytic Evaluation and Extension," *J. Manag. Stud.*, vol. 53, no. 8, 2016.
 - [22] K. M. Eisenhardt and J. a. Martin, "Dynamic capabilities: what are they?," *Strateg. Manag. J.*, vol. 21, no. 10–11, pp. 1105–1121, Oct. 2000.
 - [23] V. Ambrosini, C. Bowman, and N. Collier, "Dynamic capabilities: An exploration of how firms renew their resource base," *Br. J. Manag.*, vol. 20, no. SUPP. 1, pp. 1–41, 2009.
 - [24] J. Weerawardena, G. S. Mort, P. W. Liesch, and G. Knight, "Conceptualizing accelerated internationalization in the born global firm: A dynamic capabilities perspective," *J. World Bus.*, vol. 42, no. 3, pp. 294–306, 2007.
 - [25] M. Augier and D. J. Teece, "Dynamic Capabilities and the Role of Managers in Business Strategy and Economic Performance," *Organ. Sci.*, vol. 20, no. 2, pp. 410–421, 2009.
 - [26] D. J. Teece, "Explicating dynamic capabilities: The nature and microfoundation of (sustainable) enterprise performance," *Strateg. Manag. J.*, vol. 28, no. August, pp. 1319–1350, 2007.
 - [27] P. Oliveira, A. V Roth, and W. Gilland, "Achieving competitive capabilities in e-services," *Technol. Forecast. Soc. Chang.*, vol. 69, pp. 721–739, 2002.
 - [28] R. Agarwal, E. Baltassis, J. Brock, and J. Platt, "Enabling Big Data: Building Capabilities that Really Matter," Boston, 2014.
 - [29] K. M. Eisenhardt, "Building Theories from Case Study Research," *Acad. Manag. Rev.*, vol. 14, no. 4, p. 532, 1989.
 - [30] Sushil, "Modified ISM/TISM Process with Simultaneous Transitivity Checks for Reducing Direct Pair Comparisons," *Glob. J. Flex. Syst. Manag.*, vol. 18, no. 4, pp. 331–351, 2017.
 - [31] C. Boyce and P. Neale, "Conducting in-depth interviews: A Guide for designing and conducting in-depth interviews," 2006.
 - [32] Social Science Data and Software, "Using NVivo for Qualitative Data Analysis," vol. 1. Stanford University, Stanford, pp. 7–7, 2011.
 - [33] A. H. Hilal and S. S. Alabri, "Using NVivo for Data Analysis in Qualitative," *Int. Interdiscip. J. Educ.*, vol. 2, no. 2, pp. 181–186, 2013.
 - [34] V. Jain and T. Raj, "Modeling and analysis of FMS flexibility factors by TISM and fuzzy MICMAC," *Int. J. Syst. Assur. Eng. Manag.*, vol. 6, no. 3, pp. 350–371, 2015.
 - [35] Z. Guo, K. J. Stevens, and Y. Li, "Analyzing Students' Technology Use Motivations: An Interpretive Structural Modeling Approach," *Commun. Assoc. Inf. Syst. Vol.*, vol. 30, no. April 2012, pp. 199–224, 2012.
 - [36] R. Sindhwani and V. Malhotra, "A framework to enhance agile manufacturing system - A total interpretive structural modelling (TISM) approach," *Benchmarking An Int. J.*, vol. 24, no. 2, pp. 467–487, 2017.
 - [37] D. Jain and D. K. Banwet, "Modelling critical elements of selection for strategic alliance partner for network managed services using interpretive structural modelling (ISM)," *J. Model. Manag.*, vol. 8, no. 3, pp. 41–49, 2013.
 - [38] C. Dremel, S. Overhage, S. Schlauderer, and J. Wulf, "Towards a Capability Model for Big Data Analytics," *Proc. der 13. Int. Tagung Wirtschaftsinformatik*, pp. 1141–1155, 2017.