Spring 6-10-2017

ARE YOU UP FOR THE CHALLENGE? TOWARDS THE DEVELOPMENT OF A BIG DATA CAPABILITY ASSESSMENT MODEL

Patrick Zschech
TU Dresden, patrick.zschech@tu-dresden.de

Kai Heinrich
TU-Dresden, kai.heinrich@tu-dresden.de

Marcus Pfitzner
TU Dresden, marcus.pfitzner@tu-dresden.de

Andreas Hilbert
Dresden University of Technology, andreas.hilbert@tu-dresden.de

ARE YOU UP FOR THE CHALLENGE? TOWARDS THE DEVELOPMENT OF A BIG DATA CAPABILITY ASSESSMENT MODEL

Research in progress

Zschech, Patrick, TU Dresden, Dresden, Germany, patrick.zschech@tu-dresden.de

Heinrich, Kai, TU Dresden, Dresden, Germany, kai.heinrich@tu-dresden.de

Pfitzner, Marcus, TU Dresden, Dresden, Germany, marcus.pfitzner@tu-dresden.de

Hilbert, Andreas, TU Dresden, Germany, andreas.hilbert@tu-dresden.de

Abstract

Utilizing Big Data scenarios that are generated from increasing digitization and data availability is a core topic in IS research. There are prospective advantages in generating business value from those scenarios through improved decision support and new business models. In order to harvest those potential advantages Big Data capabilities are required, including not only technological aspects of data management and analysis but also strategic and organisational aspects. To assess these capabilities, one can use capability assessment models. Employing a qualitative meta-analysis on existing capability assessment models, it can be revealed that the existing approaches greatly differ in their fundamental structure due to heterogeneous model elements. The heterogeneous elements are therefore synthesized and transformed into consistent assessment dimensions to fulfil the requirements of exhaustive and mutually exclusive aspects of a capability assessment model. As part of a broader research project to develop a consistent and harmonized Big Data Capability Assessment Model (BDCAM) a new design for a capability matrix is proposed including not only capability dimensions but also Big Data life cycle tasks in order to measure specific weaknesses along the process of data-driven value creation.

Keywords: Business IT-Alignment, Big Data Capabilities, Assessment Model, Capability Matrix

1 Introduction

The process of digitization leads to an increasing availability of data that is matched by the evolution of information technology (IT) to cope with this situation of large collections of high frequent data with multiple datatypes and structures. This phenomenon is commonly referred to as Big Data (Chen et al., 2012; Laney, 2001). Advantages of dedicated data utilization include improved decision support in terms of transparency, improved performance management based on data evidence and the development of new business models, products and services to cope with individual customer demands (Ketter et al., 2015; Wamba et al., 2015; Wang et al., 2015). To realise these potential improvements and enable business transformation one has to combine the technological possibilities with the contextual business principals of a company using Business-IT alignment strategies (Daniel and Wilson, 2003; Venkatraman, 1994). Context factors include business processes, IT infrastructure, and applications that need to be aligned within an adequate business model (Buhl et al., 2013). To ensure the combination of those factors and thus yield sustainable advantages from Big Data substantial guidance is necessary. Useful instruments that typically embody such guidance in a tool are capability models. They are established tools to evaluate organisational and IT capabilities and assess the current situation of organisations with regard to specific requirements (Becker et al., 2009) as they can arise with the evolutionary development of new situations such as the Big Data phenomenon. The necessity of viewing a company's value through its resources originated with the development of the resource-based view (Wernerfelt, 1984). The idea of some core competencies that distinguish a company from the competition was further developed into a capability based view, allowing for a dynamic view of strategy and market behaviour (Teece et al., 1997). In order to operationalize the measurement concepts based on this theoretic point of view, the concept of capability maturity models was introduced by the Information Systems (IS) research community (Chrissis et al., 2003). With the dynamic viewpoint of a company's capabilities as critical success factors and the necessity to measure the potential of a company to cope with this new Big Data situation, a conceptualized harmonic Big Data Capability Assessment Model (BDCAM) is necessary that aligns problem specific tasks with enterprise related capabilities. In a first step towards such a coherent assessment model as part of a broader research project, our research goal can therefore be stated as the design of a capability matrix as the foundation of a capability assessment model that reflects the dimensions "capabilities" and "tasks" as well as interdependencies among those dimensions.

2 Methodology

In order to derive a matrix as a proper foundation for our BDCAM, we employ a two-step approach:

(1) Assessment and conceptualization of existing Big Data capability models

(2) Design of a new capability matrix to overcome identified deficiencies

Screening the literature a wide range of Big Data capability models can be identified. The majority of them has been developed in industry by either consulting companies or technology vendors and thus results in a heterogeneous model diversity due to subjective biases. Nevertheless the models can be used to consolidate and harmonize the results and integrate them into a new coherent model. In regards to the methodological approach there exists some literature on how to design capability models and thus achieve step (1) and (2) in a systematic manner (Becker et al., 2009; de Bruin et al., 2005; Christiansen, 2009; Mettler, 2011; Pöppelbuß and Röglinger, 2011; Solli-Sæther and Gottschalk, 2010). These approaches delivered principles and guidelines for the design of capability assessment concepts. However, for the overall development process we adopted the approach by Becker et al. (2009) due to its systematic foundation and well-documented stepwise methodology. This methodology consists of eight phases of which the first four phases are considered: *p1) problem definition*, *p2) comparison of existing models*, *p3) development strategy*, and *p4) model development*. The remaining *phases p5-p8* deal with the implementation, application and evaluation of the constructed model and will be part of further research since we only address the design of the initial model structure and consider this work as research in progress. In the next section we define the core artefact of this research paper according to *p1* and thus introduce the concept of Big Data capabilities. Subsequently, in section 4 existing Big Data capability models are identified and assessed using a systematic review in combination with a qualitative meta-analysis according to *p2*. Section 5 is considered with *p3* and *p4* where a development strategy is determined and a new Big Data capability matrix is designed based on previously derived findings. The final section discusses limitations of the artefact and gives an outlook on the scope of the full research project.

3 Big Data Capabilities

The first step towards creating a capability assessment model starts with the initial problem definition (Becker et al., 2009). This will be done by introducing the concept of Big Data capabilities. The phenomenon of Big Data is a frequently discussed topic among scientists and practitioners within the IS domain. Basically, it describes a situation of an increasing availability of poly-structured data that is generated with increasing frequency (Laney, 2001). The level of data volume and complexity creates a necessity for specialized systems that are capable of storing, managing, analysing, and visualizing this data in a high performance manner (Chen et al., 2012). Reflecting various aspects behind the phenomenon of Big Data, de Mauro et al. (2015) conducted a comprehensive literature review and integrated the most important parts into a representative definition: "*Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical*

Methods for its transformation into Value.” (de Mauro et al., 2015, p. 13). Given this summary, it can be seen that the inherent attributes of Big Data are an integral part of its definition but furthermore it requires sophisticated resources to turn this data into value. These resources are not just limited to technical ones as it also asks for an alignment within the enterprise context including further aspects such as business models, organisational structures, processes, governance and more (Buhl et al., 2013; Gupta and George, 2016). According to the resource-based view, these additional technical and organisational resources can be expressed as dynamic capabilities of a company. Within the Big Data context the definition of Teece et al. (1997) is applicable who define dynamic capabilities as “*(…) the firm’s ability to integrate, build, and reconfigure internal and external competences to address rapidly changing environments.*” (Teece et al., 1997, p. 516). Given the Big Data environment, we define Big Data capabilities as the necessary resources to create business value out of Big Data situations. This involves capabilities to manage and analyse data (Agrawal et al., 2012; Gandomi and Haider, 2015), organisational aspects of a data driven culture, employee skills and technology awareness (Gupta and George, 2016). To utilize a company’s Big Data capabilities, it is inevitable that a company is able to compile and assess the resources mentioned above. Therefore, an assessment model enables an analysis of the state of capabilities and benchmarks them in the context of the company and the relevant market. The second important function of a capability assessment model is to envision possible development patterns in order to improve capabilities towards an anticipated state.

4 Meta-analysis of existing Big Data Capability Models

The second phase of the approach by Becker et al. (2009) involves a comparison of existing Big Data capability models. In order to collect and analyse existing models, we initially employed a systematic literature review (Kitchenham et al., 2009) by conducting a database search in the following databases: EBSCO Academic Search Complete, EBSCO Business Source Complete, Association for Computing Machinery (ACM) Digital Library, IEEE Xplore Digital Library, Emerald Insight, Science Direct, Springer Link, Web of Science Core Collection and DBLP Computer Science Bibliography. As a search string we used the term “big data” in combination with “capability model”, “capability maturity model” and “maturity model”. Without restriction the search yielded 234 papers (day of search: 12.07.2016). In the next step the results were filtered for formal criteria (research articles, free access and no duplicates) and only articles were kept that specifically address the development and/ or description of a concrete Big Data capability model. Based on this search strategy, solely one item remained (Sulaiman et al., 2015). Therefore, an extensive backward search was conducted, including not only scientific literature but also best practice models developed in industry. In this step only approaches were taken into consideration where the articles describing the models at least delivered some textual description with free access policy in English or German language. Furthermore, we explicitly excluded capability models from closely related (sub-) domains such as master data management, data governance or business intelligence (BI) from the review process in order to avoid conceptual diffusion at this stage of the development process. However, we acknowledge the existence of familiar and well established models, especially in the field of BI (e.g. Cates et al., 2005; Dinter, 2012; Lahrmann et al., 2011; Thamir and Poulis, 2015), and therefore utilize the capability concepts presented in those models as a comparative benchmark for our resulting capability matrix.

Eventually, the backward search returned another 13 models which reflect the overall relevance for Big Data assessment models. However, the majority of these models are based on industrial demands driven by inherent business models of either providing consulting or training services. Consequently the models lack of scientific rigor and objectivity with regard to the model development and application. Nevertheless, they pose an insight into relevant practice approaches on handling Big Data. Both structural elements and model contents specified by the different models can be regarded as critical success factors. Therefore, the models can be considered to fulfil industry-relevant requirements and thus should be used as inputs for the development of a consistent and harmonized BDCAM.

In order to assess the identified models and extract and conceptualize inherent expert knowledge and empirical results; we used a qualitative meta-analysis (Dixon-Woods et al., 2005; Greenhalgh, 1997). The basic scope of a capability assessment model comprises the fundamental model architecture with its dimensions and structural relationships between these dimensions reflected in a *capability matrix*. Another element, especially in the related context of maturity models, is the specification of maturity levels based on different stage characteristics and fulfilment criteria (Becker et al., 2009) reflected in a *measurement system*. These design elements can be used as a basis to analyse existing Big Data capability models. While the focus of this paper is the harmonization of different framework dimensions into a coherent capability assessment matrix, further research will consider necessary elements for a measurement system, following the ideas of maturity levels and fulfilment criteria (see discussion).

A closer look at the architecture of the 14 individual models reveals a highly heterogeneous model diversity due to their independent development and subjective biases. While the number of maturity levels solely varies from four to six stages, the assessment categories range from two to eight dimensions, addressing technological and business aspects to varying degrees (c.f. Appendix, Table 1). The aggregation of all individual dimensions yields a total number of 67 across all models. Due to varying terminology, structural familiarity and duplicates, this amount can be reduced to 34 different dimensions. Nevertheless, the remaining result set cannot be regarded as well-defined and distinct and therefore still allows overlapping and hierarchical subcategories (e.g. "*Operating Model*" (Knowledgent, 2014) vs. "*Organizational Capabilities and Resources*" (El-Darwiche et al., 2014) vs. "*Standards and Processes*" (SAP, 2013)). However, for the development of a generic and objective assessment model it is necessary to further consolidate this heterogeneous information base into consistent capability concepts. Therefore, we conduct an in depth content analysis of the heterogeneous dimensions and reduce them into ten Big Data related capability concepts:

- **Strategy**: Describes the strategic setting of Big Data initiatives due to measurable goals, visions and corporate culture as well as business model-related aspects such as investment and resource planning.
- **Performance Measurement**: Describes aspects of a transparent realization of business goals through the implementation of key performance indicators and measurement systems.
- **Skills**: Describes required expertise and experiences of human resources and therefore addresses related topics such as training, employee development and recruiting.
- **Organisation**: Describes organisational elements such as responsibilities, hierarchies, roles, leadership and team, department and management structures.
- **Processes**: Describes the operationalisation of strategic goals with defined standards, procedures, and workflows as well as established project management methods.
- **Technology**: Describes information and communication technology resources including applications, platforms, tools, infrastructures and architectures.
- **Data Management**: Describes tasks and processes involved in collecting, storing, processing and providing data including further aspects like data modelling, data quality and meta data management.
- **Data Analytics**: Describes tasks and processes for extracting knowledge using data analytics methods such as visualisations or descriptive, predictive and prescriptive analysis.
- **Information Management**: Describes the process of knowledge generation from data analysis results as a strategic asset towards the support of organisational decision making and action planning.
- **Governance**: Describes rules, guidelines and corporate policies to guarantee security, ethics, privacy or legal requirements.

In addition to those identified core concepts some aspects cannot be sorted into a category due to lacking support in the meta-analysis document base (e.g. "*Use Cases*" (SAP, 2013), "*Business Need*"

(Knowledgent, 2014), “*Business Value*” (Trost, 2015), “*Big Data Enablers*” (Sulaiman et al., 2015)). Those aspects aim at the assessment of company-specific conditions and promising benefits from Big Data initiatives that could not be aligned with a fundamental, theoretical base and therefore will not be considered at this stage of research. Nevertheless, we acknowledge that in accordance to Portela et al. (2016) such considerations could be used for a “pre-assessment” of a company’s individual context factors and business needs, evaluating to what extent Big Data initiatives are useful (see discussion).

Big Data Capability Models	Strategy	Performance Measurement	Skills	Organisation	Processes	Technology	Data Management	Data Analytics	Information Management	Governance
Dhanuka (2016)	x	x	x	x	x	x	x	x		x
El-Darwiche et al. (2014)	x		x	x		x	x	(x)		x
Halper and Krishnan (2013)	x		x	(x)	(x)	x	x	x		x
Infotech (2013)	x	(x)	x	(x)	x	x	x	(x)		x
Knowledgent (2014)	x		x	x	x	x	x	x		x
Meir-Huber and Köhler (2014)	x		x	x	(x)	x	x			x
Nott (2015)	x					x	(x)	x	x	x
Radcliffe (2014)	x	x	x	x	x	x	x	x		x
SAP (2013)	(x)		x	(x)	x	x	(x)	(x)	(x)	x
Sulaiman et al. (2015)	x	(x)	x	x	x	x	(x)	(x)	(x)	(x)
Thompson (2015)	(x)		x		x	x	(x)	(x)		
Trost (2015)	x	x	x	x		x	x	(x)		
van Veenstra (2013)	x					x	x	x		x
Vesset et al. (2013)	x	x	x	x	x	x	x	(x)	(x)	x
explicit count	**12**	**4**	**12**	**8**	**8**	**14**	**10**	**6**	**1**	**11**
total count	**14**	**6**	**12**	**11**	**10**	**14**	**14**	**13**	**4**	**12**

x ⇒ explicitly as dimension (x) ⇒ implicitly mentioned

Figure 1. Meta-analysis results of dimensions from observed Big Data capability models

In a next step, it was examined which of the extracted core concepts are explicitly defined dimensions within the existing models and which of them are only implicitly addressed via indirect model descriptions (cf. Figure 1). Two results can be revealed: First, except the concepts of *Performance Measurement* and *Information Management* most of the observed models share a consensual view on the dimensions to be considered in a coherent BDCAM. Second, there is a high proportion of implicit aspects among most of the categories. Particularly those categories that address data-related tasks show interference with other dimensions. For example, in several models aspects of *Data Analysis* overlap with aspects of *Technology*, *Skills* and *Processes* (Infotech, 2013; SAP, 2013; Sulaiman et al., 2015; Thompson, 2015; Trost, 2015; Vesset et al., 2013). Hence, the dimensions suffer from a lack of discriminatory power and therefore assessment criteria cannot be assigned unambiguously. This can lead to misunderstandings during the assessment of capabilities which might lead to incorrect evaluation results (Pöppelbuß and Röglinger, 2011). De Bruin et al. (2005) summarize these requirements with principles of mutually exclusive and collectively exhaustive dimensions. Additionally, the observed models do not take dependencies into account nor do they introduce weights for the assessment dimensions. In contrast, the majority of the observed models either build parallel dimensions of strategic, technological and data-centric aspects or combine them into a non-distinguishable collection.

5 Design of a new Big Data Capability Matrix

Given the results of the model comparison, a new capability matrix is designed based on defined development strategy. According to Becker et al. (2009), one can decide whether to choose a development that results in a fundamentally new model or an enhancement of existing models by integrating given model elements. In order to use existing expert knowledge and empirical results with regards to critical Big Data success factors, we suggest to use the extracted core concepts from the given models and transform them into a new capability matrix that addresses the observed deficiencies.

In synthesis with scientific literature, a capability matrix to address the issues stated in the latter section should obey the principals of mutually exclusive and exhaustive dimensions. Therefore, we propose a *two space* capability matrix that, in contrast to the existing one-dimensional *dimension space* with conventional dimensions of a capability assessment model, also contains a *task space*. This second space contains all relevant tasks that reflect the Big Data life cycle from data collection and data storage to data analysis and knowledge generation (Chen and Zhang, 2014; Gandomi and Haider, 2015; Mountasser et al., 2015; Siddiqa et al., 2016) and therefore is divided into *Data Management*, *Data Analytics* and *Governance*. Those are the root hierarchy elements that can be subdivided into more relevant tasks depending on existing taxonomies (Siddiqa et al., 2016). With the separation of the two spaces, the routines in form of Big Data processes in the *task space* can be measured against the derived dimensions in the *dimension space*, yielding dynamic capability assessments (Constantiou and Kallinikos, 2015; Mountasser et al., 2015). Additionally, sequential dependencies between individual Big Data life cycle tasks (Mountasser et al., 2015) can be considered and assessed accordingly. For example, if weaknesses were found in the task of streaming data analysis, it can also be screened whether appropriate processes and systems for handling streaming data are established in earlier stages. Furthermore, the aspect of *Governance* is assorted a special role within the *task space*. In this way, it can be associated with single tasks like data security guidelines for a specific data management/ analysis task or with the general task of governance, allowing dependencies of different governance tiers within a company.

The *dimension space* on the other hand contains aspects of organisational relevance in order to assess the capabilities to fulfil Big Data tasks along the data life cycle while distinguishing between the different aspects extracted via our meta-analysis. According to Mettler (2011) and the commonly used CMMI approach (Kneuper, 2006), there are three basic dimensions used for capability assessment: *Processes*, *Objects* and *People*. Another model that also supports this concepts is the BIMM approach from Dinter (2012) as a representation of capability models within the BI context, involving the dimensions *Technology* and *Organisation* among other aspects. We synthesize those results with our extracted dimensions and transform the candidates *Technology*, *Processes* and *Skills* directly into final model dimensions within the dimension space. Additionally, we add the dimensions *Organisation* and *Strategy* to respect the necessity of roles and responsibilities as well as business goals (Chen and Zhang, 2014; Debortoli et al., 2014). Analogous to the task of *Governance* the dimension *Strategy* takes a superior function to enable dependencies between company vision and the other aspects (Journeault, 2016; Moore, 2001). The *Strategy* dimension explicitly contains all aspects of the previously extracted dimension *Performance Measurement* since it is an integral part of strategy to measure the magnitude of reaching strategic goals (Journeault, 2016). The candidate dimension *Information Management* extracted by the meta-analysis is not transformed into dimensional space since it focusses on the decision process that follows the data life cycle but barely is part of it. However, the structure of the capability matrix allows an ex post integration without loss of consistency if further model enhancements are required.

Creating an orthogonal capability matrix, we place the *dimension space* as columns whereas the *task space* forms the rows of the matrix (cf. Figure 2). Furthermore, we suggest a hierarchical structure for both task and dimension elements in order to allow the subdivision into more specific aspects; e.g. the subdivision of *Data Analytics* tasks into descriptive, predictive and prescriptive analytics with individual subtasks (Delen and Demirkan, 2013; Evans and Lindner, 2012) or the subdivision of the *Organisation* dimension into more detailed elements such as roles, responsibilities and entity structures. The cells of the matrix can then be seen as granular aspects for the assessment of capabilities along different tasks of the Big Data life cycle considered from different organisational and technological viewpoints. Hence, measuring dynamic capabilities alongside routines (Constantiou and Kallinikos, 2015) builds the core concept of our BDCAM approach.

The assessment itself can be realised with a suitable measurement system which as indicated before will be part of further research. Nevertheless, we already propose an initial draft for a possible direction in order to convey the basic idea towards the application of our BDCAM approach. For each cell within the capability matrix, specific assessment criteria can be set up based on a catalogue of assessment questions, fulfilment indicators and a scoring system (Yucalar and Erdogan, 2009). The measurement system then implements a general weighted function ***f(t,d)*** of an element ***t*** of the *task space* and an element ***d*** of the *dimension space*. For example, taking the intersection between the *Data Analytics* task "predictive analytics" with its subtasks and the *Organisation* dimension, it could be assessed whether or not dedicated predictive analytics teams are established, how many analysts are involved in creating and evaluating predictive models and whether or not clear responsibilities are assigned. Further examples of such assessment questions are shown in Figure 2. At this point, the assessment is not limited to only fine-granular cells, but also assessment criteria for inter-cellular and superordinate blocks are possible (e.g. general aspects of *Data Management* such as the availability of standard workflows for the management of metadata). Based on the fulfilment of each question, standardized scoring points can be derived assessing the capability level of a dimension with regard to a certain task. The assessment scores for superior hierarchy levels can then be calculated by an aggregation mechanism as a function of the subordinate levels. Thus, observed high level capability weaknesses can be broken down into granular details in order to derive practical guidelines to resolve the issues and identify possible development paths. To visualise the degrees of fulfilment and make weaknesses transparent to model users, it is helpful to apply diagrams such as bar charts as illustrated in Figure 2 for demonstration purposes.

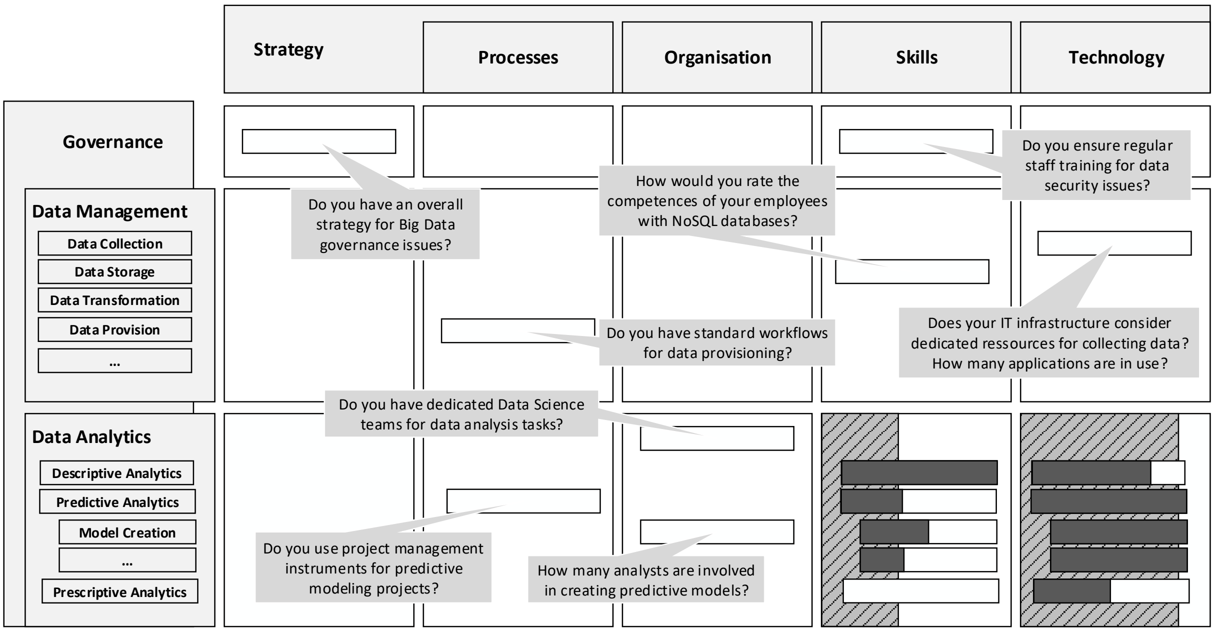

Figure 2. Design of a new capability matrix for a coherent BDCAM

6 Discussion and outlook

In this paper, a capability matrix as the foundation for a BDCAM was synthesized from the results of a qualitative meta-analysis based on heterogeneous, non-exhaustive and non-exclusive dimensions. The results could be separated into a *task space* and a *dimension space*. While the dimension space reflects general capability dimensions derived from established capability models, the task space reflects Big Data tasks that are derived from a Big Data life cycle. Therefore, our BDCAM synthesizes the aspects of general capability models with the specific needs of Big Data environments. The synthesis and integration into the proposed *two space* structure results in distinguishable elements in both the task and dimension space with higher discriminatory power and the possibility to consider interdependencies in the form of a hierarchical structures in both spaces. The process of designing the model follows the phases suggested by Becker et al. (2009) while we focused on the first four steps towards the development of a comprehensive capability assessment model. The remaining steps of the methodological approach include the development of further model elements and the implementation, application and evaluation of the overall model which will be part of further research. The design of the capability matrix is the result of a first iteration. In addition, we proposed an initial draft for a measurement system. Thus, the next steps that will be conducted within the full research process involve (i) confirming and enhancing the conceptual and structural elements of both spaces with empirical studies and expert knowledge and (ii) the fundamental development of a suitable measurement system for the capability assessment, containing a catalogue of standardized assessment questions, a systematic scoring system, aggregation functions and a weighting mechanism. Furthermore, it is planned to develop a pre-assessment step as an integral part of our BDCAM, evaluating pre-conditional context factors and business needs to estimate to what extent Big Data initiatives are useful (Portela et al., 2016). Depending on the company's individual situation it is not always promising to establish static encompassing Big Data capabilities if the specific environment does not ask for it. For example, if a company produces large amounts of poly-structured data which, however, is not generated in high frequencies, as supposed to be for instance with streaming data, the company will not need particular capabilities for it. Thus, such circumstances should be considered within the overall assessment in order to avoid biased evaluation results that lead to misleading conclusions about a firm's capability of handling their specific Big Data situation. As a consequence, the idea is to integrate a context-sensitive pre-assessment system which will lead to a situational customization of the cellular blocks within the previously designed capability matrix by adjusting underlying assessment and fulfilment criteria accordingly.

References

Agrawal, D., P. Bernstein, E. Bertino, S. Davidson, U. Dayal, M. Franklin, J. Gehrke, L. Haas, A. Halevy, J. Han, H. V. Jagadish, A. Labrinidis, S. Madden, Y. Papakonstantinou, J. M. Patel, R. Ramakrishnan, K. Ross, C. Shahabi, D. Suciu, S. Vaithyanathan and J. Widom (2012). *Challenges and Opportunities with Big Data - A Community White Paper Developed by Leading Researchers Across the United States.* URL: http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf (visited on 11/01/2016).

Becker, J., R. Knackstedt and J. Pöppelbuß (2009). "Developing Maturity Models for IT Management: A Procedure Model and its Application." *Wirtschaftsinformatik* 51 (3), 249–260.

de Bruin, T., M. Rosemann, R. Freeze and U. Kulkarni (2005). "Understanding the Main Phases of Developing a Maturity Assessment Model." In: *Proceedings of the 16th Australasian Conference on Information Systems*, Sydney: Australia, 8–19.

Buhl, H. U., M. Röglinger, F. Moser and J. Heidemann (2013). "Big Data: A Fashionable Topic with(out) Sustainable Relevance for Research and Practice?" *Business & Information Systems Engineering* 5 (2), 65–69.

Cates, J. E., S. S. Gill and N. Zeituny (2005). "The Ladder of Business Intelligence (LOBI): A Framework for Enterprise IT Planning and Architecture" *International Journal of Business Information Systems* 1 (1/2), 220–238.

Chen, H., R. H. L. Chiang and V. C. Storey (2012). "Business Intelligence and Analytics: From Big Data to Big Impact." *MIS quarterly* 36 (4), 1165–1188.

Chen, P. C. L. and C.-Y. Zhang (2014). "Data-intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data." *Information Sciences* 275, 314–347.

Chrissis, M. B., M. Konrad and S. Shrum (2003). *CMMI Guidlines for Process Integration and Product Improvement*. URL: http://dl.acm.org/citation.cfm?id=773274/ (visited on 11/01/2016).

Christiansen, S.-K. (2009). *Methode zur Klassifikation und Entwicklung reifegradbasierter Leistungsbewertungs- und Leistungssteigerungsmodelle*. Paderborn: Heinz Nixdorf Institut.

Constantiou, I. D. and J. Kallinikos (2015). "New Games, New Rules: Big Data and the Changing Context of Strategy." *Journal of Information Technology* 30 (1), 44–57.

Daniel, E. M. and H. N. Wilson (2003). "The Role of Dynamic Capabilities in E-Business Transformation." *European Journal of Information Systems* 12 (4), 282–296.

Debortoli, S., O. Müller and J. Vom Brocke (2014). "Comparing Business Intelligence and Big Data Skills: A Text Mining Study Using Job Advertisements." *Business & Information Systems Engineering* 6 (5), 289–300.

Delen, D. and H. Demirkan (2013). "Data, Information and Analytics as Services." *Decision Support Systems* 55 (1), 359–363.

Dhanuka, V. (2016). *Hortonworks Big Data Maturity Model*. URL: http://hortonworks.com/wp-content/uploads/2016/04/Hortonworks-Big-Data-Maturity-Assessment.pdf (visited on 01/11/2016).

Dixon-Woods, M., S. Agarwal, D. Jones, B. Young and A. Sutton (2005). "Synthesising Qualitative and Quantitative Evidence: A Review of Possible Methods." *Journal of Health Services Research & Policy* 10 (1), 45–53.

El-Darwiche, B., V. Koch, D. Meer, R. T. Shehadi and W. Tohme (2014). *Big Data Maturity*. URL: http://www.strategyand.pwc.com/reports/big-data-maturity/ (visited on 11/01/2016).

Evans, J. R. and C. H. Lindner (2012). "Business Analytics: The Next Frontier for Decision Sciences." *Decision Line* 43 (2), 4–6.

Gandomi, A. and M. Haider (2015). "Beyond the Hype: Big Data Concepts, Methods, and Analytics." *International Journal of Information Management* 35 (2), 137–144.

Greenhalgh, T. (1997). "Papers That Summarise Other Papers (Systematic Reviews and Meta-Analyses)." *BMJ: British Medical Journal* 315 (7109), 672–675.

Gupta, M. and J. F. George (2016). "Toward the Development of a Big Data Analytics Capability." *Information & Management* 53 (8), 1049–1064.

Halper, F. and K. Krishnan (2013). *TDWI Big Data Maturity Model Guide*. URL: https://tdwi.org/white-papers/2013/10/tdwi-big-data-maturity-model-guide/asset.aspx (visited on 11/01/2016).

Infotech (Ed.) (2013). *Big Data Maturity Assessment Tool*. URL: https://www.infotech.com/research/it-big-data-maturity-assessment-tool (visited on 11/01/2016).

Journeault, M. (2016). "The Integrated Scorecard in Support of Corporate Sustainability Strategies." *Journal of Environmental Management* 182, 214–229.

Ketter, W., M. Peters, J. Collins and A. Gupta (2015). "Competitive Benchmarking: An IS Research Approach to Address Wicked Problems with Big Data and Analytics." *ERIM Report Series Research in Management Erasmus Research Institute of Management*. URL: http://repub.eur.nl/pub/79158/ (visited on 11/09/2016).

Kitchenham, B., O. P. Brereton, D. Budgen, M. Turner, J. Bailey and S. Linkman (2009). "Systematic Literature Reviews in Software Engineering – A Systematic Literature Review." *Information and Software Technology* 51 (1), 7–15.

Kneuper, R. (2006). *CMMI: Verbesserung von Softwareprozessen Mit Capability Maturity Model Integration*. 2nd Edition. Heidelberg: Dpunkt.

Knowledgent (Ed.) (2014). *Big Data Maturity Assessment*. URL: https://bigdatamaturity.knowledgent.com/ (visited on 11/01/2016).

Lahrmann, G., F. Marx, R. Winter and F. Wortmann (2011). "Business Intelligence Maturity: Development and Evaluation of a Theoretical Model". In: *Proceedings of the 44th Hawaii International Conference on System Sciences.* Kauai: USA, 1–10.

Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity and Variety*. URL: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (visited on 11/01/2016).

de Mauro, A., M. Greco and M. Grimaldi (2015). "What is Big Data? A Consensual Definition and a Review of Key Research Topics." In: *Proceedings of the 4th International Conference on Integrated Information*. Ed. by G. Giannakopoulos, D. P. Sakas and D. Kyriaki-Manessi. Madrid: Spain, 97–104.

Meir-Huber, M. and M. Köhler (2014). *Best Practice für Big Data Projekte*. URL: http://www.bmvit.gv.at/service/publikationen/innovation/downloads/big_data_leitfaden.pdf (visited on 11/01/2016).

Mettler, T. (2011). "Maturity Assessment Models: A Design Science Research Approach." *International Journal of Society Systems Science* 3 (1/2), 81–98.

Moore, J. I. (2001). *Writers on Strategy and Strategic Management: Theory and Practice at Enterprise, Corporate, Business and Functional Levels* Penguin UK. URL: https://books.google.de/books?hl=de&lr=&id=D3kgzSFm8bMC&oi=fnd&pg=PT5&ots=1y02KL9hj-&sig=lfDWuq5o_49EogVsICzj_8-u0DE (visited on 08/10/2016).

Mountasser, I., B. Ouhbi and B. Frikh (2015). "From Data to Wisdom: A New Multi-Layer Prototype for Big Data Management Process." In: *Proceedings of the 15th International Conference on Intelligent Systems Design and Applications*. Marrakech: Morocco, 104–109.

Nott, C. (2015). *A Maturity Model for Big Data and Analytics*. URL: http://www.ibmbigdatahub.com/blog/maturity-model-big-data-and-analytics (visited on 11/01/2016).

Pöppelbuß, J. and M. Röglinger (2011). "What Makes a Useful Maturity Model? A Framework of General Design Principles for Maturity Models and its Demonstration in Business Process Management." In: *Proceedings of the 19th European Conference on Information Systems.* Helsinki: Finland, p. 28.

Portela, F., L. Lima and M. F. Santos (2016). "Why Big Data? Towards a Project Assessment Framework." *Procedia Computer Science* 98, 604–609.

Radcliffe, J. (2014). *Leverage a Big Data Maturity Model to Build Your Big Data Roadmap*. URL: http://www.radcliffeadvisory.com/research/download.php?file=RAS_BD_MatMod.pdf (visited on 11/01/2016).

SAP (Ed.) (2013). *Getting Big Value from Big Data*. URL: https://www.sapvirtualagency.com/FileExplorer/Partners/HANA/Big%20Data%20Opportunity/Guide%20-%20Getting%20Big%20Value%20from%20Big%20Data.pdf (visited on 11/01/2016).

Siddiqa, A., I. A. T. Hashem, I. Yaqoob, M. Marjani, S. Shamshirband, A. Gani and F. Nasaruddin (2016). "A Survey of Big Data Management." *Journal of Network and Computer Applications* 71, 151–166.

Solli-Sæther, H. and P. Gottschalk (2010). "The Modeling Process for Stage Models." *Journal of Organizational Computing and Electronic Commerce* 20 (3), 279–293.

Sulaiman, H., Z. C. Cob and N. Ali (2015). "Big Data Maturity Model for Malaysian Zakat Institutions to Embark on Big Data Initiatives." In: *Proceedings in the 4th International Conference on Software Engineering and Computer Systems*. Kuantan: Malaysia, 61–66.

Teece, D. J., G. Pisano and A. Shuen (1997). "Dynamic Capabilities and Strategic Management." *Strategic Management Journal*, 509–533.

Thamir, A. and E. Poulis (2015). "Business Intelligence Capabilities and Implementation Strategies." *International Journal of Global Business* 8 (1), 34–45.

Thompson, S. (2015). *Big Data Maturity Model*. URL: http://www.cbigconsulting.com/wp-content/uploads/2015/08/CBIG-Consulting-Big-Data-Maturity-Model.pdf (visited on 11/01/2016).

Trost, U. (2015). *Big Data Future - Chancen Und Herausforderungen für Die Deutsche Industrie*. URL: http://www.mhp.com/fileadmin/mhp.de/assets/studien/MHPStudie_BIG-DATA.pdf (visited on 11/01/2016).

van Veenstra, A. F. (2013). *Big Data in Small Steps*. URL: http://www.idnext.eu/files/TNO-whitepaper–Big-data-in-small-steps.pdf (visited on 11/01/2016).

Venkatraman, N. (1994). "IT-Enabled Business Transformation: From Automation to Business Scope Redefinition." *Sloan Management Review* 35 (2), 73–88.

Vesset, D., G. Girard, J. Feblowitz, M. Versace, C. Burghard, A. O'Brien, C. W. Olofson, D. Schubmehl, B. McDonough, A. Woodward and S. Bond (2013). *Big Data Maturity Model*. URL: https://www.idc.com/getdoc.jsp?containerId=255138 (visited on 11/01/2016).

Wamba, S. F., S. Akter, A. Edwards, G. Chopin and D. Gnanzou (2015). "How 'Big Data' Can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study." *International Journal of Production Economics* 165, 234–246.

Wang, X., White. Leroy and X. Chen (2015). "Big Data Research for the Knowledge Economy: Past, Present, and Future." *Industrial Management & Data Systems* 115 (9), 1566–1576.

Wernerfelt, B. (1984). "A Resource-Based View of the Firm." *Strategic Management Journal* 5 (2), 171–180.

Yucalar, F. and S. Z. Erdogan (2009). "A Questionnaire Based Method for CMMI Level 2 Maturity Assessment." *Journal of Aeronautics and Space Technologies* 4 (2), 39–46.

Appendix

Big Data Capability Models	Dimensions	Big Data Capability Models	Dimensions
Dhanuka (2016	- Sponsorship - Data and Analytics Practices - Technology and Infrastructure - Organization and Skills - Process Management	Radcliffe (2014)	- Vision - Strategy - Value & Metrics - Governance, Trust & Privacy - People & Organization - Data Sources - Data Management - Analytics & Visualization
El-Darwiche et al. (2014)	- Technical capabilities/infrastructure - Organizational capabilities and resources - Data availability and governance - Sponsorship - Data driven descision making culture	SAP (2013)	- Use Cases - Information and Application Architecture - Standards and Processes - People and Skills - Governance
Halper and Krishnan (2013)	- Organization - Infrastructure - Data Management - Analytics - Governance	Sulaiman et al. (2015)	- Business Goals - Big Data Enablers - People - Processes - Technology
Infotech (2013)	- People - Process - Technology - Data	Thompson (2015)	- People - Process - Technology
Knowledgent (2014)	- Business Need - Technology Platform - Operating Model - Analytics - Information Management	Trost (2015)	- Business value - Organisation - Technology - Data
Meir-Huber and Köhler (2014)	- Kompetenzentwicklung - Infrastruktur - Daten - Prozessumsetzung - Potenziale	van Veenstra (2013)	- Organizational Capablities - Technical Capabilities
Nott (2015)	- Business strategy - Information - Analytics - Culture and Execution - Architecture - Governance	Vesset et al. (2013)	- Intent - Data - Technology - People - Processes

Table 1. Overview of existing Big Data capability models and considered dimensions