2015 48th Hawaii International Conference on System Sciences

Guiding the Introduction of Big Data in Organizations: A Methodology with Business- and Data-Driven Ideation and Enterprise Architecture Management-Based Implementation

Martin Vanauer
University of Münster – ERCIS
vanauer@ercis.de

Carsten Böhle
University of Münster – ERCIS
boehle@ercis.de

Bernd Hellingrath
University of Münster – ERCIS
hellingrath@ercis.de

Abstract

Researchers and practitioners frequently assume that big data can be leveraged to create value for organizations implementing it. Decisions for big data idea generation and implementation need careful consideration of multiple factors. However, no scientifically grounded and unbiased method to structure such an assessment and to guide implementation exists yet. This paper describes a methodology based on IT value theory and workgroup ideation guiding big data idea generation, idea assessment and implementation management. Distinct business and data driven perspectives are distinguished to account for big data specifics. Enterprise Architecture Management and Business Model Generation techniques are used in individual steps for execution. A first prototypical application in the context of Supply Chain Management illustrates the applicability of the method.

1. Introduction

During the last three years, big data has advanced to one of the most debated trends in information systems. In a 2011 research report, McKinsey famously coined big data "[t]he next frontier for innovation, competition, and productivity" [1]. In fact, the topic's popularity – measured by internet search volume and publications – has increased exponentially [2]. In a nutshell, it is often assumed that big data can unleash new value and revolutionize industries [3]. New data sources can be leveraged to optimize current operations by allowing better analysis. Even more revolutionary, big data could alter existing and add new business models as firms could for instance sell data newly collected.

In practice however, organizations need to assess first whether new data is beneficial for their operations or enables the monetization of information. Second, IT infrastructure must be dimensioned accordingly to support the use of big data. As of 2014, technology firms and consultancies have published impact reports and case studies [4], [5] as well as generic methods for ideation, assessment and implementation of big data, e.g. [6]–[8]. These methods are a first stepping stone for cost-and-benefit as well as implementation roadmap decisions. Presented as one page checklists or figures, they however provide little guidance on how to approach raised questions (e.g. "Identify the use cases required to carry out your project."). Moreover, such generic methods are typically opaque when it comes to their development. Therefore, it remains unclear whether they are scientifically or empirically grounded, unbiased, and of general applicability.

In order to tackle this gap, this paper presents a method which is based on ideas from established scientific topics like IT value, workgroup ideation processes as well as Enterprise Architecture Management (EAM). The methodology describes a step-by-step approach developed for organizations to structure the introduction of big data, including steps to

(1) develop ideas for big data usage,
(2) assess these ideas with regard to their potential value as well as the required changes to the organizations architecture, and
(3) implement them coherently in the business.

For each step, methods are proposed which can support organizations in executing them – usually based on an enterprise architecture view on a firm.

To present the developed methodology, this paper is structured in seven sections. After the introduction, big data and the mentioned concepts are briefly introduced, providing a theoretical background to the development process of the methodology. The third section gives a general overview of all methodological steps, followed by sections four and five detailing the method's main blocks ideation and implementation. Section six exemplarily illustrates the method's applicability and feasibility. The paper concludes with a summary and an outlook on big data ideation, assessment and implementation research.

1530-1605/15 $31.00 © 2015 IEEE
DOI 10.1109/HICSS.2015.113

IEEE computer society

2. Theoretical foundations

2.1. Big data

Scoping the potential impact: As the name suggests, "big data" refers to data sets of significant size. The exact *volume* which makes a normal data set to big data, however, is subject to a moving definition. The core reason against a definite cutoff number – e.g. data larger than 500 petabytes – is the constant advancement of technology, which allows larger and larger data sets to be handled [9].

Beyond this most obvious dimension, the term big data encompasses two additional dimensions: variety and velocity [1], [10]. *Variety* primarily indicates that the number of data sources increases. Moreover, also different forms of data have appeared. These do not confine to structured data with well-defined data-definitions but also encompass semi-structured data as in geographical maps or Wikipedia entries, or unstructured data like social media content on Twitter or Facebook. *Velocity* refers to the (increasing) speed of data generation, as well as to its transfer and analysis. For many use cases, the possibility to analyze data in real time and integrate findings into ongoing processes or decision support systems is even more important than the vast data volume. Providing infrastructure which is able to handle, i.e. transport, store and compute (analyze) big data can become a challenge for organizations.

V*alue* and *veracity* are often cited as additional dimensions to characterize big data [9], [11]. Value addresses the usefulness of data to make decisions as the paradigm is not only about data itself but its use for analytical purposes. Veracity is connected to this and signifies the unpredictability and validity of particular kinds of data which are uncertain by their nature. Examples are data from public sources or even apparently reliable information: GPS sensor measurements may contain fuzziness, e.g. distortion by the reflection of high buildings nearby.

All five dimensions highlight that big data not only refers to data itself but also to technology enabling its use. This involves new analysis paradigms: e.g. instant data analysis without storage; or partitioning of analysis tasks with approaches like map-reduce [11].

2.2. IT value and IT strategy

Deducting starting points: The processes that capture, store and analyze big data are primarily performed by IT. With the advent of wide-scale IT deployment, researchers and practitioners alike were tasked with one question: if IT brings value to firms in the form of competitive advantage. Such theories explaining how IT (generically) brings value to organizations provide an overview of all potentially achievable benefits. Thereby, they help to understand what perspectives should be considered for evaluating specific IT topics like big data. Early research indicated that IT may help to lower costs by essentially making processes more efficient, so serving a typical business requirement. Beyond that, it was also recognized that IT can enhance differentiation and change the competitive scope, which effectively denotes the creation of new opportunities [12].

These two different perspectives are also echoed in IT strategy discussions where one view is based on the idea that IT supports the implementation of a business strategy. Business strategies for instance determine the business scope, competencies and – on an implementation layer – processes, skills and administrative infrastructure. All these aspects nowadays rely on IT for operations support. Vice versa, IT capabilities defined and developed by an IT strategy enable the execution of these business services, but also allow e.g. technology transformation of the business. For instance, business strategy can drive the business whereas IT has the sole role of supporting and implementing it. However, IT strategy can also drive the business with business management (as stakeholders associated with business strategy) solely serving as visionaries and potentially catalysts to realize the new ideas of IT [13], [14].

2.3. Workgroup ideation processes

Structuring idea development, assessment and implementation: Use cases where big data can be applied in organizations are – as most applications of "new" technology – usually not obvious instantly but are developed over time. In practice, the generation of ideas is a result of "innovators" working together. According to empirical observations, the effectiveness of such workgroups is an important factor for generating and implementing innovative solutions [15].

The innovation process can be distinguished into two distinct phases: an *ideation/creativity phase*, followed by the innovation *implementation phase*. While this linearity is helpful as a theoretical construct, loop backs, e.g. for idea refinement, may be found in practice. Both phases contain a transition and action (sub-) phase. Transition can be understood as a scoping and decision step, succeeded by action which denotes the application and realization of decisions [15], [16].

On a high level, the ideation transition phase focusses on the mission definition, involving the interpretation of issues and identification of opportunities. This transition is followed by the ideation action phase where creative solutions to the identified problem and opportunity are identified. The implementation transition phase focuses on the evaluation and selection of the most appropriate ideas whereas the subsequent action phase is about the realization of the ideas and solutions towards the identified problem or opportunity [15], [16].

In practice, the overall process might be instantiated multiple times. While some groups are ideating, others may already be implementing generated solutions and ideas.

2.4. Enterprise architecture management

Guiding the implementation of big data: Big data potentially changes all layers of an enterprise: new sensors or other hardware resources on an infrastructure level, software to store and analyze data on the application layer and processes on the business layer. As a result, the transition from an as-is state to a state which is enhanced with the application of big data (to-be state) needs to be guided carefully. Approaches like Business Process Reengineering focusing on just one aspect do not suffice.

In contrast, EAM is a methodology that specifically aims at the integrated modelling, analysis, and governance of business and IT [17], [18]. It is used for enterprise transformation which is "driven by experienced and/or anticipated value deficiencies that result in significantly redesigned and/or new work processes" [19]. EAM can be distinguished in enterprise architecture (EA) itself and its management.

EA encompasses information and technology services, processes, and infrastructure, expressed in terms of models. While elements considered to be part of an EA model differ between frameworks, a hierarchical layer structure is usually assumed [20], [21]. On top, a business layer represents the business architecture – i.e. strategy, governance, organization and business processes. The following technology layer encompasses e.g. applications. A bottom layer contains infrastructure, e.g. hardware and networks.

EAM uses these models to create a consistent view on impacted processes, applications, and technology. EAM frameworks provide comprehensive approaches for designing, planning, implementing, and governing such complex transformations affecting all layers of an enterprise [18], [22]. The frameworks offer guidance for organizational set-ups, e.g. for required roles and stakeholders, documents to create or the communication required in complex enterprise environments. Moreover, tools support the transformation by offering automatic support for qualitative and quantitative analysis, e.g. for gaps between an as-is and to-be state, or performance calculations.

3. Methodology overview

As with other new concepts, tools, or information brought into an organization, the introduction of big data resembles an innovative process as it introduces something new with an a priori not completely known benefit. Following a model of workgroup innovation explaining such processes (cf. section 2.3), the presented methodology (Figure 1) is structured in two phases: first an *ideation phase* with a subsequent *implementation phase*. Each phase is again structured into *transition* and *action (sub-) phases* which have different characteristic steps. Ideation is not a mere idea generation (creativity) but the generation of solutions by applying big data to new situations (innovation). Implementation encompasses an evaluation of developed solutions, which is not part of ideation, and a subsequent realization of use cases if a respective decision was made.

For the ideation phase specifically, two perspectives are defined. Insights gained from extensive research of IT value generation and realization (by means of IT strategy) theories can be transferred to the specific big data IT context. Main findings were that either business requirements can be better fulfilled by IT, or that IT opens up new business opportunities. Transferring the main ideas, an introduction process may a) either start with business requirements or b) with new aspects enabling opportunities which are for big data new data and analysis possibilities. This is reflected in two distinct approaches which are referred to as *Business First* (BF) or *Data First* (DF).

A third conceivable approach, Technology First, is practically not feasible because of two reasons: First, technology should not be introduced as an end in itself but in such a way that it provides value to the business [23], [24]. The question should therefore not be where new analysis (technology-based) could be applied but which data contains value if analyzed. The latter is already reflected in DF. Second, big data is not a new technology trend per se but different trends coming together. Thus, it would be difficult to clearly distinguish which technology artifacts are starting points: software like Apache Hadoop, or specific algorithms.

All in all, the ideation phase either resembles the identification of business needs in the BF approach; or the development of new business models based on an identification of available data in DF. For all approaches, an EA model of the enterprise serves as an input to many steps. If such a model does not yet exist, it should therefore be created in a preliminary phase.

In BF, the ideation transition phase is mainly about defining which enterprise goals need to be fulfilled and which current challenges hinder the company to do so. Modelling and Requirements Engineering (RE) is used to support this step [25]. The action phase then starts the development of one or more business use cases which resolve identified challenges or incrementally improve business with regard to the goals.

For DF, the ideation transition phase does not start with business requirements but with the identification of key resources – essentially specific types of big data like sources and the data captured by them, or the resulting analysis and insights gained thereby. Afterwards, the action phase requires the derivation of a value proposition. DF is primarily supported by business model design concepts, like the Business Model Canvas [26] applied in our case.

The two approaches are also relevant for transition in the implementation phase, where – in a first step – the financial and organizational feasibility is checked. Therefore, either a cost/benefit analysis (in BF) or a value proposition fit assessment (in DF) is performed. Afterwards, both approaches merge for a second step in implementation transition where the focus is shifted to technical feasibility. The order is intentional as both cost/benefit and value proposition fit assessment are qualitative or significantly rely on assumptions. For implementation however, a detailed assessment of actual required transformation effort is possible.

This second transition step also marks the beginning of the extensive use of EAM methods [17], [18], [21] for impact assessment and rollout. First, a model of the as-is and targeted to-be state of the organization's architecture is used to identify changes and perform a technical feasibility assessment for the implementation of big data. The subsequent implementation action phase is also unified and first starts with deducting an implementation roadmap and, finally, the enterprise transformation.

The following short descriptions of steps focus on the most important aspects to consider. Additionally, structured tables are provided to compare the different approaches and steps in detail. Table 1 compares the BF and DF ideation phases. The implementation phases' steps are compared in Table 2 (BF/DF) and 3 (unified).

4. Ideation

4.1. Business First

Transition – objectives decision and challenges identification: The BF approach starts with business requirements and is based on the assumption that those existing requirements can be fulfilled by big data. The main premise is the improvement of current operations by collecting and analyzing the new data available. In a first step, business requirements are identified in the form of organizational goals and existing and known operational challenges.

Enterprise goals, typically expressed in a strategy, are primarily identified to scope the search for use cases. For instance, innovators in an organization with a focus on costs may favor ideas to improve efficiency, e.g. improving forecast accuracy with better data. Measurability of goals is required for evaluation purposes. The identification of challenges is another scoping tool. Challenges are known obstacles in operations which could not be "solved" by current technology, processes etc. Big data may be an approach to resolve the challenges.

Models support both tasks. Strategy understanding is facilitated by modelling the associated goals by means of EAM [25]. Having such a model, it becomes possible to analyze to what extent the enterprise meets the goals. One example is the detection of conflicting interests and solutions, where e.g. one goal is to assign personal assistants to customers whereas another requires personnel reduction. Those issues are challenges which can be identified by goal analysis techniques of RE [27].

The BF transition step is completed when the subsequent use case development is clearly scoped.

Action – business use case development: In the action step, use cases – ideas for the application of big data in the organization – are developed based on the scope and context provided by the transition step.

Guiding the development, two general concepts exist, into which the developed ideas should fit. First,

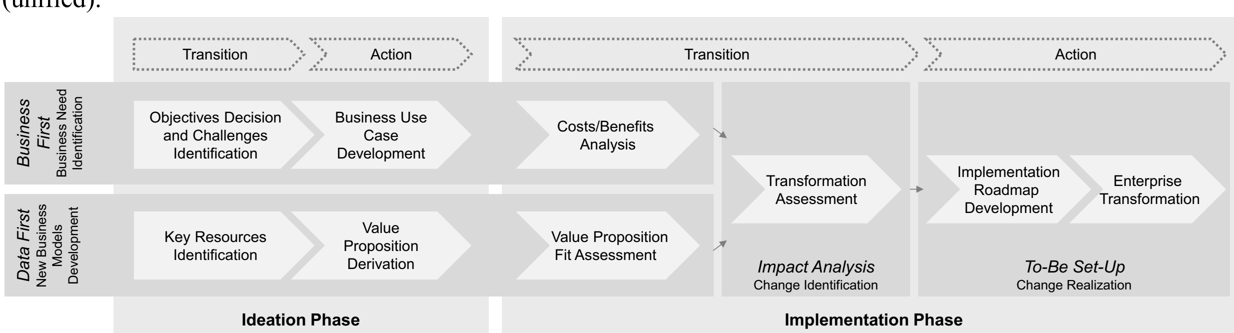

Figure 1. Big data ideation, assessment and implementation methodology

Table 1. Ideation phase: Comparison of Business First and Data First

Approach	Business First	Data First
Phase goal	Improving current operations by use of big data, expressed as use case ideas	Creating sellable services using big data, expressed as coherent business models
Sub-phase objectives	*Transition*: Identification of enterprise objectives and challenges for improvements *Implementation*: Development of use cases	*Transition*: Identification of key resources *Implementation*: Derivation of value propositions based on key resources
Primary innovators	*Trans.* (objectives): Senior mgmt., strategists *Trans*. (challenges): First-/mid-level mgmt. *Implementation*: First- and mid-level mgmt., operations personnel, data scientists	*Transition*: First- and mid-level management, enterprise architects (if EA model exists) *Implementation*: Senior management, strategists, sales, data scientists/analysts
Main inputs	Strategic goals of organization, current operational challenges (i.e. issues known to affect performance)	Knowledge on organization and potentially available data (i.e. may not yet be captured), market expertise regarding sales opportunities
Method support	*Transition*: EA models of strategy and challenges, goal analysis techniques *Implementation*: EA model, creativity tech.	*Transition*: BMC, EA model with modeled data objects *Implementation*: BMC, creativity techniques
Phase result	Use case measurably improving identified objectives by e.g. resolving challenges	Potentially sellable value proposition (service) with customer segments, etc. (cf. BMC)

use cases may partially or completely resolve challenges identified in the transition. Second, a use case may also contribute to a part of an organization where no explicit need for improvement (i.e. no challenge) was recognized beforehand. The improvement shall be associated to one of the a priori defined measurable goals.

The use case development primarily requires domain knowledge of operational processes, which is the reason why first- and mid-level management and personnel are primary innovators. These are brought together with scientists and analysts to foster discussions on operations as well as data (analysis) potentials.

Use case development processes may abstractly be supported by creativity techniques. From an EA perspective, views which are partial models tailored to the needs of specific stakeholders are promising tools. They can provide additional context to innovators by creating transparency on e.g. the as-is situation as well as available data, interfaces etc. The action step is completed when innovators have identified a number of big data use cases or have concluded that none exist.

4.2. Data First

Transition – key resources identification: In contrast to the requirements driven BF approach, Data First starts with the newly available data (and associated analysis possibilities). It aims at creating services which can be sold to other entities. While the focus on operational improvement (BF) could be seen as evolutionary, DF is per se revolutionary as it revolves around designing a new business model (BM).

The term BM subsumes the idea of a service, as well as the presumed customers, the resources needed to realize the service etc. [28]. One of the most prominent BM meta models is the Business Model Canvas (BMC, Figure 2) which contains 9 building blocks [26]. The key building block is the central value proposition which is marketed and sold to specific customer segments via channels. Internally, the value proposition is realized by key resources and activities using those resources, possibly relying on key partners.

Data (and its use) are localized in the key resources (key activities) blocks. In the transition step, innovators identify data which is captured and therefore can be a key resource for potential value propositions. Such identification may be supported by EA models which typically contain explicitly modeled data objects.

However, the identification process may not necessarily focus on data already captured in IT only but also needs to include what data could be available. For instance, a retailer may currently not store all point of sales data but could connect all check registers to a collecting server to use it for analysis. Consequently, the list of data to be prepared is not required to be very detailed (e.g. speed, position, etc. of trucks) but can be on a higher level of granularity (e.g. fleet movement).

The DF ideation transition step is completed when a high level overview of the company's potentially capturable data is prepared.

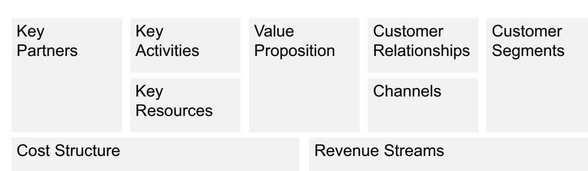

Figure 2. Business Model Canvas

Table 2. Implementation transition phase: Comparison of Business First and Data First

Approach	Business First	Data First
Phase goal	Selecting financially beneficial big data use cases for transformation assessment and implementation	Selecting financially beneficial business model which value proposition fits to the overall organizational goals (i.e. not negatively affecting it)
Sub-phase objectives	Assessment of financial benefit of use cases (business case)	Assessment of value proposition with regards to strategic match, available expertise (to realize BM) and financial viability (business case)
Primary innovators	First- and mid-level management (supported by e.g. strategists for business case calculation)	Senior management (supported by e.g. strategists for business model creation)
Main inputs	Use cases, models including KPIs of current situation, accounting numbers (e.g. salaries)	Business models, organizational goals
Method	Controlling	EA models of strategy, goal analysis techniques
Phase result	Prioritized use cases for transformation assessment (incl. business cases for use cases)	Prioritized business models for transformation assessment (incl. business cases)

Action – value proposition derivation and business model development: Key resources allow an enterprise to create and offer value to customers in the form of a value proposition. In the action phase of the DF ideation approach, such a value proposition needs to be derived from the theoretically available data. Creating a value proposition requires an ability "to imagine what does not exist". The required creativity can however be guided by techniques, e.g. visual thinking, prototyping, storytelling or scenarios [26].

In that step, innovators need to understand that data cannot be seen isolated but that a value proposition may imply a combination of multiple data. Moreover, potentially only portions of needed data are available whereas others cannot be captured in-house but could be purchased. In that case, key partners would help to realize a value proposition. A complete development of a BMC should be pursued as it is helpful for further assessment. Following the distinction that the ideation phase is about idea generation only, all potential value propositions should be developed, disregarding whether they are feasible or for instance match the company's culture of disclosing data etc. Such a check is performed in the implementation phase.

The DF ideation action step is completed when innovators have developed sketches of complete BMC which contain data potentially able to be captured as a key resource. Alternatively, it is also completed if an informed decision was made that no value proposition could be derived from possible key resources.

5. Implementation

5.1. Business First

Transition – costs/benefits analysis: The transition sub-phase of implementation is mainly concerned with the evaluation and selection of the best use cases developed in the preceding ideation phase. The list of business use cases resulting from the BF ideation phase needs to be assessed with regards to financial validity for long-term use.

Overall, the use cases need to enhance the fulfillment of enterprise objectives as measured by the associated metrics. As a result, business cases are set-up and calculated, involving aspects like potential saved costs or enhanced revenue, but also running costs for the use of the data, e.g. new salaries of data scientists or maintenance of sensor networks. Costs for implementation are not considered as they are part of the second transition step. In addition, non-monetary aspects like higher customer satisfaction (as a potential business objective) have to be considered. Financial and non-financial dimensions are then aggregated to decide on a priority list of use cases to realize.

The BF implementation transition step is completed when the use cases are prioritized according to their impact on the objectives of the organization.

5.2 Data First

Transition – value proposition fit assessment: In contrast to the simple costs/benefits analysis of the BF approach, DF requires a more extensive check of the derived ideas. In addition to a financial viability assessment, also a fit assessment of the new BM must be performed with regard to the established BM.

The new value proposition must be included into that model and realized by existing organizational capabilities. For instance, even if all key resources and activities are available within the firm, but no sales channel exists yet to the targeted user group, it must be created. Although companies theoretically could build-up these resources, it may not be justified with regard to potential benefits compared to the overhead created. This argument mirrors insights gained from traditional strategy literature which states in general that concentration on one strategy or core ability is beneficial [29], [30]. Similarly, companies may have a value proposition to sell certain data, but cannot do so as having data exclusively is either a reason for having competitive advantage or may even be required legally.

The DF implementation transition step is completed when the business models are prioritized, including a category of "must not be realized".

5.3 Impact analysis

Transition – transformation assessment: The second transition step of the implementation phase marks the shift from an either data or business driven perspective to a unified one. The unified perspective is assumed because both to-be implemented solutions have to be integrated into an existing EA which they are validated against. Primarily, two different aspects need to be determined: the required kind and number of resources to realize the big data use case and the complexity of transitioning to a to-be state. For this purpose, the interrelation of processes, data, and technology is investigated.

Foremost, an existing IT infrastructure may need to be changed to account for the new requirements. Large data volumes have to be transferred from the point of creation to a point of (temporary) storage and analysis. Such analyses require considerable computing capacities. In a nutshell, storage space, network capacity and computing power must be dimensioned and distributed correctly to avoid creating bottlenecks.

For this reason, first, the amount of data created should be estimated, e.g. by approximating the number of sources, and which data size is created per time. Architecturally, it then is necessary to determine where the data is processed as this determines the storage space for very large data. As argued by [9], stored data (e.g. collected in a year) may become so large that it cannot be transferred to a separate point of analysis in reasonable time. However, also continuously created data needs to be transferred to a point of storage/ analysis, e.g. from sensors to a server analyzing it instantly. While networks within a facility may be capable to handle additional data load, an internet connection may become a non-identified bottleneck.

Potential options include the extension of in-house resources or – if assessed as reasonable for big data – the use of external infrastructure as e.g. in cloud computing. For both, the calculated input is relevant – either to invest in the servers (in-house) or to select the most suitable cloud computing provider. Quantitative analysis of EA models has been proposed as a solution to determine the amount of data in the organization, and whether the infrastructure (i.e. number of servers realizing storage space and computing power) and all networks are sized correctly for the new data load [31].

The complexity of the change is expressed as the differences between an as-is state without and a to-be state with big data. The current architecture modeled in an EA modelling language as well as use cases and business models are inputs for modelling the full to-be situation, including the determined size of resources, their ownership and location. By modelling this state, the impact of the intended change on the organization is defined. Narrating the architecture work, EAM frameworks ensure that all relevant outputs and deliverables are developed (e.g. TOGAF ADM B-D).

A gap analysis, automatically generated by EAM tools, identifies unchanged, eliminated, and new elements. Upon completion, a list of the gaps, i.e. the elements that must be added for the to-be architecture, is available. Using the identified gaps, the resulting change can be quantified. In addition to the number of elements to change itself, it is also possible to infer how many connections are affected. With increasing complexity measured by connections, difficulty increases too, reducing feasibility.

Before implementation, the envisioned to-be architecture needs to be validated, e.g. by consulting experts. In order to only address them with aspects regarding their sphere of expertise, EAM enables "filtering" by using so called viewpoints. For example, an application usage viewpoint contains only selected elements from the business and application layers, but none from the infrastructure layer. Anyone interested in the usage of applications within business processes could therefore use this viewpoint to extract the necessary information while hiding information that is irrelevant in this context. Therefore, the correctness of a certain change can be validated with experts in specific fields – while still maintaining a consistent picture in the overall architecture.

The transition step concludes when, based on an understanding of necessary resources, the to-be scenario is modeled, and a model-analysis based decision is made to realize or not realize it.

5.4 To-be set-up

Action – implementation roadmap development: Given that a decision to implement the big data to-be state was made, realization needs to be planned. While small scale changes may be implemented easily, larger changes require a detailed roadmap which for instance could include multiple transition architectures. Guided by EAM frameworks like the well-established TOGAF [21], realizing a target architecture involves the identification and prioritization of work packages and mentioned transition architectures.

Big data is associated with specific challenges which need special *consideration* when developing the

Table 3. Implementation phase: Unified transition and action steps

Approach	Impact analysis (transition)	To-be set-up (action)
Phase goal	Determining if big data use case/business model should be realized based on required extent of necessary changes to organizations architecture	Realizing the planned transformation from an as-is state to the envisioned big data to-be state
Sub-phase objectives	Determination of required infrastructure resources (storage, computing, network) and development of a to-be architecture incl. big data use to determine extent of changes (gaps)	*Roadmap development*: Development and set-up of implementation and migration plan *Enterprise transformation*: Governing, managing and realizing impl. and migr. plan
Innovators	Enterprise architects, relevant innovators (e.g. from changing business units management)	*Both*: Enterprise architects, relevant innovators (e.g. from affected business units)
Main inputs	EA model of relevant parts of enterprise, use case/business model to realize	*Roadmap dev.*: architectural inputs *Enterprise transf.*: impl. and migration plan
Method support	TOGAF ADM phases B-D, EA views, architect-ture analysis (e.g. gap, quantitative EA analysis)	*Roadmap dev.*: TOGAF ADM phase E-F *Enterprise transf.*:TOGAF ADM phase G-H
Phase result	To-be EA model including assessment of neces-sary changes, decision to realize to-be scenario	*Roadmap dev.*: Impl. and migration plan *Enterprise transf.*:Realized transformation

implementation roadmap. The most apparent challenge concerns technical and data security governance. Technically, big data ecosystems can become complex. Hadoop for instance is comprised of 14 packages. Here, well-defined transition architectures can support understanding the whole eco system step by step instead of trying to realize it in one-step only. Moreover, big data inherently refers to integrating and correlating as much data as possible. However, from a legal perspective, certain data may not be allowed to be stored or used for certain kinds of analysis. The planning of implementation and associated governance has to ensure that the protection of data is guaranteed without exception. This also applies to decisions related to hosting as data may not be allowed to stretch over national borders. Similarly, with increasing size of a data repository, it is more likely to be target of theft as it becomes more valuable. Such security issues must not be neglected when deciding on governance mechanisms in the roadmap.

The first implementation action step is completed when an implementation roadmap is developed and matched to the identified capabilities of the company.

Action – enterprise transformation: In the last step of the methodology, enterprise transformation is realized. Governed by EAM methods, organizations follow the implementation roadmap to achieve the target state. Finally, after realizing architectures, performance is measured and compared to the pursued objectives as identified in ideation.

6. Application example

While the preceding sections focused on the abstract description of the methodology, the following paragraphs will outline – based on the presumed case of a mid-sized manufacturing company – how it can be applied in practice. This manufacturing company uses complex machines with high operating costs. As-is models of the organization's high level processes and infrastructure, modeled in the ArchiMate language [32], are available to innovators. Due to the type of the scenario, a description of the realization (enterprise transformation) of the developed solution is omitted.

Ideation – transition: The firm only sells physical goods and has no intentions to offer complementing services. Therefore, no data-based business models can be identified, but a constant need for improved operations exists, leading to a BF approach. In the first BF step, by using a model of the enterprise's goals, it is recognized that unplanned failures or maintenance shut-downs of machines are a core challenge. They negatively impact two major goals. First, outages significantly affect customer service because of production on-demand so that no inventory is available to offset delays in production. Second, low reliability, which is measured by hours of unplanned downtime per week, is associated with high repair costs.

Ideation – action: In the use case development step, the use of technology to predict potential failures for preemptively planned maintenance is envisioned. Intelligent maintenance systems (IMS) use data from machines' sensors to discover patterns indicating a forthcoming breakdown (Figure 3).

Implementation – transition: Recognizing that algorithms become more accurate when more data

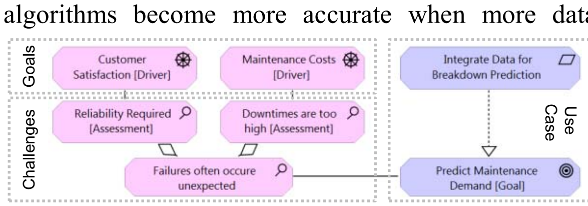

Figure 3. Business First ideation: Objective and deducted use case

throughout the machine's lifecycle is available, the firm plans to use an advanced IMS, integrating sensor data but also semi-structured information as working principles [33] which is data with distinct big data characteristics (Table 4). This decision is confirmed in the costs/benefits assessment. First, scheduling maintenance activities are faster and less expensive than unplanned repairs. Moreover, the improved reliability is also calculated to improve customer satisfaction – offsetting the IMS costs.

In the unified transformation assessment, a full model of the enterprise's to-be setup is developed – including the internal storage system and networks from the different machines to the server environment. Using quantitative analysis on the model, after adding the new machines, an increase of current network infrastructure (Gigabit LAN) usage from 10% to 200% is detected. More detailed assessment identifies that all machines would transmit about 200 MB/second over the network. As a result, the to-be situation is adapted: production receives additional network capacity to transmit data to an in-place analysis server which is dimensioned accordingly and where it is analyzed instantly. Persistent storage of all data is not pursued. However, a selection of the data is stored on a cloud IMS storage and analysis environment where a knowledge base of machine health is build up for learning opportunities [33]. In this cloud, a data technology stack is required, e.g. a Hadoop deployment consisting of different components such as a HDFS database, Flume for collecting stream data, or Hive for managing and querying the dataset.

Internally, processes aligned with an operations reference model, are adapted accordingly. The process "Schedule Asset Management Activities" makes use of a service provided by an Enterprise Asset Management System (EAMS). While the EAMS formerly just returned vendor's recommended intervals, it is now connected to the cloud as well as to the analysis server. Additionally, processes can now be triggered by the EAMS if the analysis server identified an upcoming breakdown. An excerpt of the to-be model is shown for illustration in Figure 4. Extended graphical models are available from the authors on request.

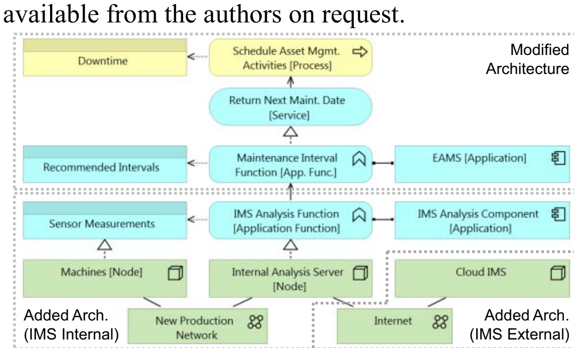

Figure 4. High level IMS to-be architecture

Table 4. Smart maintenance data characteristics

Dim.	Characteristics
Volume	*~6.3 PB/year* (200 machines; 71 sensors/ machine; each sensor ~54 MB/hour [34])
Velocity	*~200MB/second* data generation by sensors, addl. data irregularly
Variety	Integration of structured sensor data and semi-structured data from other systems (system configuration, physical knowledge, working principles)
Veracity	Risk of misinterpreting abnormal sensor values (e.g. due to defect sensors)
Value	Unplanned disruptions have high impact on business, reduction potential by IMS

7. Conclusion and outlook

The implementation of big data will continue to shape organizations by changing processes, applications and infrastructure. Due to the complexity associated to such changes, the technology should not be introduced instantly but assessed first with regards to advantageousness and implementation complexity.

In this paper, a methodology for big data ideation, assessment and implementation was presented. In contrast to other methodologies, its structure is based on scientific theories: a) workgroup ideation for the overall approach to create, evaluate and realize solutions and b) IT value for distinguishing different perspectives on how big data can bring value to organizations. All steps of the method moreover rely on established bodies of knowledge to guide users in application. Implementation and the ideation phase of BF are supported by means of EAM. The data driven business model development (DF) is supported by the BMC which is one of the most prominent BM ideation techniques. Relying on such well-researched and practically adapted knowledge ensures that researchers and practitioners alike can easily apply the method without the need to familiarize with new special interest topics.

While the praxis example highlighted the applicability of the method in one specific context, further application is beneficial for validation and refinement. A large business process outsourcing provider (BPOP) has agreed to participate in a case study, which is currently being conducted and planned to be completed until April 2015. In the ongoing case study, we study how the methodology can be applied to a range of application areas and industries as multiple divisions of the BPOP create and handle large data for different customers. First results point to more research into the question of integration where multiple internal and external data sources have to be managed and governed accordingly. An additional test is discussed with a logistics service company. We expect this test to be specifically interesting due to the distributed nature of data creation: e.g. by a large fleet of vehicles equipped with sensors.

In line with this research plan, other studies may also investigate if and how different steps are adapted in practice. More application would provide more validation of the proposed innovator groups' set-up and an estimation of required time for ideation and implementation. Detected best practices could be transferred back to methodologies used: namely EAM and BMC development. Overall, the importance of assessing big data first, before implementing it, must remain a core notion to firms.

8. References

[1] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation competition and productivity," 2011.

[2] Google, "Search term comparison: Cloud Computing, Big Data," 2014. [Online]. Available: http://www.google.com/tremds/explore#q=bigdata%2Ccloudcomputing. [Accessed: 11-Jun-2014].

[3] T. Davenport, P. Barth, and R. Bean, "How 'big data'is different," MIT Sloan Manag. Rev., vol. 54, no. 1, 2012.

[4] IBM, "IBM Big Data: A collection of Big Data client success stories," 2012.

[5] T. H. Davenport and J. Dyché, "Big Data in Big Companies," 2013.

[6] C. Hagen, K. Khan, M. Ciobo, J. Miller, D. Wall, H. Evans, and A. Yadav, "Big Data and the Creative Destruction of Today's Business Models," 2013.

[7] Intel, "Getting Started with Big Data: Planning Guide," 2013.

[8] T Systems, "Big Data Readiness Assessm.," 2013.

[9] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big Data: Issues and Challenges Moving Forward," in 46th Hawaii International Conference on System Sciences, 2013, pp. 995–1004.

[10] Meta Group, "3D Data Management: Controlling Data Volume, Velocity, and Variety," 2001.

[11] A. E. Gattiker, F. H. Gebara, A. Gheith, H. P. Hofstee, D. A. Jamsek, J. Li, E. Speight, J. W. Shi, G. C. Chen, and P. W. Wong, "Understanding System and Architecture for Big Data," 2012.

[12] M. E. Porter and V. E. Millar, "How information gives you competitive advantage," Harv. Bus. Rev., no. July-August, pp. 149–160, 1985.

[13] J. Henderson and N. Venkatraman, "Strategic alignment: Leveraging information technology for transforming organizations," IBM Syst. J., vol. 32, no. 1, pp. 472–484, 1993.

[14] M. Mocker and A. Teubner, "Information Strategy - towards a comprehensive model of information strategy," in ECIS, 2005.

[15] J. Farr, H. Sin, and P. Tesluk, "Knowledge Management Processes and Work Group Innovation.," in The International Handbook on Innovation, L. V. Shavinina, Ed. 2003, pp. 574–586.

[16] M. M. Hammond, N. L. Neff, J. L. Farr, A. R. Schwall, and X. Zhao, "Predictors of individual-level innovation at work: A meta-analysis.," Psychol. Aesthetics, Creat. Arts, vol. 5, no. 1, pp. 90–105, 2011.

[17] M. M. Lankhorst, Enterprise Architecture at Work: Modelling, Communication and Analysis, 2nd ed. Springer, 2013.

[18] M. Op't Land, E. Proper, M. Waage, J. Cloo, and C. Steghuis, Enterprise Architecture: Creating Value by Informed Governance. Springer, 2009.

[19] W. Rouse and M. Baba, "Enterprise transformation," Commun. ACM, vol. 49, no. 7, pp. 67–72, 2006.

[20] R. Winter and R. Fischer, "Essential layers, artifacts, and dependencies of enterprise architecture," in 10th IEEE International Enterprise Distributed Object Computing Conference Workshops, 2006.

[21] The Open Group, TOGAF Version 9. 2009.

[22] K. Winter, S. Buckl, F. Matthes, and C. M. Schweda, "Investigating the State-of-the-Art in Enterprise Architecture Management Methods in literature and Practice.," in MCIS, 2010.

[23] N. Carr, "IT doesn't matter," Harv. Bus. Rev., 2003.

[24] D. Kiron and R. Bean, "Organizational Alignment Is Key to Big Data Success," MIT Sloan Manag. Rev., 2013.

[25] D. Quartel, W. Engelsman, H. Jonkers, and M. Van Sinderen, "A goal-oriented requirements modelling language," in 13th IEEE International Enterprise Distributed Object Computing Conference, 2009.

[26] A. Osterwalder and Y. Pigneur, Business Model Generation. 2008.

[27] A. van Lamsweerde, "Requirements engineering in the year 00: a research perspective," in 2000 International Conference on Software Engineering ICSE 2000.

[28] S. M. Shafer, H. J. Smith, and J. C. Linder, "The power of business models," Bus. Horiz., vol. 48, no. 3, pp. 199–207, 2005.

[29] J. Quinn, T. Doorley, and P. Paquette, "Technology in services: rethinking strategic focus," Sloan Manage. Rev., 1990.

[30] M. E. Porter, Competitive Strategy: Techniques for Analyzing Industries and Competitors. 1980.

[31] M. Vanauer, C. Böhle, and B. Hellingrath, "The Impact of Big Data on Organizations: Using Enterprise Architecture Management and Quantitative Analysis for Implementation Planning," Working Paper (submitted for review), 2014.

[32] The Open Group, ArchiMate® 2.1 Specification. 2013.

[33] J. Lee, E. Lapira, B. Bagheri, and H. Kao, "Recent advances and trends in predictive manufacturing systems in big data environment," Manuf. Lett., vol. 1, no. 1, pp. 38–41, Oct. 2013.

[34] A. Vons, "Big Data from Sensors: Engine for Competitive Differentiation," SAP, 2014. [Online]. Available: http://scn.sap.com/community/high-tech/blog/2014/08/19/big-data-from-sensors-engine-for-competitive-differentiation.