A double means-end relationship for data

Bas van Gils
Strategy Alliance
Amersfoort, The Netherlands
bas.vangils@strategy-alliance.com

Hend*erik* A. Proper
TU Wien
Vienna, Austria
henderik.proper@tuwien.ac.at

***Abstract*—It appears that, in an increasingly digital world, *data* is becoming more and more important for organizations. In several real-world cases in the Netherlands, we have seen that organizations struggle with issues such as: *a*) What data do we have? *b*) Which data do we need for value creation? *c*) Which strategic choices do we make around using data? *d*) Which strategic choices do we make around managing data as an asset? And *e*) Which skills do our people need in light of the previous items?**

The *context* for these organizations is different, but the *challenges* are the same. Our position is that the use of *models* can help to get to grips with the complexity of the aforementioned challenges.

In this paper we explore *a*) the relationship between *data* and *models*, and *b*) a framework that connects data management as a means to care for data as an asset, and data as a means to achieve strategic objectives of the organization. In developing the latter, we follow a design science approach; we explore cases to understand challenges and requirements for such a framework as well as show how the framework helped to solve the challenges in these organizations.

***Index Terms*—data strategy, data management, semiotic triangle**

I. Introduction

It appears that data-related terms (data strategy, data science, artificial intelligence, becoming more data-driven, and managing data as an asset) are increasingly popular in business conversations and popular (management) literature. One stakeholder involved in a recent consultancy assignment[1] (end of 2023) lamented that "after years of riding the Lean and Six Sigma wave, we now seem to have discovered that we are about to drown in a lake of data if we're not very careful".

We agree that 'data' is important, yet also emphasize the need for a holistic overview [27] of (the essence of) the organization. For example, in the DEMO [7] method, the notion of communicative action is seen to be essential and equal attention must be paid to understanding the various facets of communicative action – which includes processes, data, and action rules [7]. This is also what the aforementioned sponsor alluded to: "We need to get grip on the organization *as a whole* and not just from a single perspective." Even more, it aligns with the findings in our recent book [30] where an integral view of organizations as enabler for digital transformation is presented. We expect that most stakeholders in organizations *know* this too, yet *struggle* to deal with the practical reality of having to deal with data-related challenges.

In [42], we have reported an earlier case in which we discussed the challenges of a pension fund provider in getting to grips with data as an asset in order to use data for value creation. Our hypothesis is that a better understanding of what data *is*, and an associated framework, will help organizations. This boils down to a belief that a better understanding of what data *is*, fundamentally an ontological question, complemented with a definition that positions the 'data buzzwords', and puts these in perspective, will help organizations to effectively leverage their data.

In this paper, we study three recent projects in light of the aforementioned hypothesis and as a basis for proposing a framework[2]: *a*) a (different from the one used in [42]) pension fund provider, *b*) a drinking water company, and *c*) a major harbor. We will first describe the three cases in Section II and explain what the challenges were that we encountered while working with these organizations. In Section III we will present an overview of the relevant literature in light of the objective of this paper. We develop the proposed framework in Section IV and reflect on this framework in light of the three cases in Section V. Conclusions and future research are presented in Section VI.

II. Three cases

In this section, we present three cases, which are all situated in the Netherlands, from the 2021-2023 period. The first author of this paper served as a consultant in each of these three cases. Names of people, products, etc. have been removed or altered for reasons of confidentiality. The details of these cases are reconstructed based on *a*) the agenda and meeting notes of sessions, *b*) the project assignment (statement of work), and *c*) the deliverables of the projects. We present the cases 'as they happened' based on this reconstruction. At the start of the next section, we will present preliminary observations/conclusion that will be the basis for the literature study.

A. Pension fund provider

We will use the abbreviation PFP as a place holder for the name of this pension fund provider. A pension fund provider typically has two lines of business: taking care of pensions of

[1] The assignment that is referred to was executed by the first author of this paper. The stakeholder and company wish to remain anonymous.

[2] The first author of this paper was one of the consultants that executed the projects.

participants and investment to increase the value of financial assets.

Initially, we were approached by (business) management with the question: please help us to become a *data-driven* pension fund provider. A few meetings later, we had discovered that PFP had no clear idea what they meant by the term *data-driven* yet there was a strong belief that it would be a good idea to use data to improve (efficiency of) processes, reduce cost by eliminating work, experiment with smart technologies (e.g. artificial intelligence), etc.

We started with an initial round of semi-structured interviews – which were intended to get a good understanding of the maturity of the organization in managing data as an asset, in using data for value creation, and for related foundational aspects and capabilities. In the semi-structured interviews, we used a fixed set of seed questions that were discussed in a time box. We used a time box to 'force' the stakeholders to present the most important insights first and to prevent them from sharing too many less important details. We guided the stakeholders along these questions and asked them to give an assessment of the current maturity level (low/medium/high) on various aspects of data management. This setup ensured that we could compare the insights from the various stakeholders. We did not need more detail and precision, so we kept it as simple as possible.

From these interviews, we learned two things. First, we found that stakeholders agreed that by and large the overall maturity was low[3]. Certainly it was higher in some areas due to local expertise of specific individuals, yet the average scores were low across the board. Second, we learned that stakeholders have a very poor understanding of data and its associated terminology. Basic terms such as 'data' and 'information' were used loosely and were misunderstood. As a result, there was confusion about the distinction between data management (loosely: the activities for managing data as an asset) and information management (which in the Netherlands is often positioned as the discipline that transforms the need for information systems to solutions, often taking classic business/IT alignment literature as a basis [19, 28]).

In the workshops for designing a strategic vision (in the form of a blueprint) and road-map for this organization, we used the DAMA DMBOK [20] as a basis in conjunction with the work on data strategy by DalleMule and Davenport in [5] as well as our own publication in this area [13]. During the workshops, we learned that the group struggled with the complicated nature of cause and effect in this context – which we translated to a double means-end relationship as illustrated in Fig. 1. The DM activities are a *means* towards the end of having a (healthy) Data Landscape, while the latter serves as an end towards higher order business goals such as Reduce cost, provide Better customer service and being More data-driven. This helped us to achieve the desired outcome for this project: once people understood this, it became easier to build the blueprint and road-map that will assist the organization in achieving its strategic objectives. The shared terminology and framework of means-end relationships provided the clarity that helped the group to create a shared blueprint/road-map.

In the years that followed, we continued to help the organization – mostly through training. In these training sessions, we again noticed the need to very clearly define key terms (i.e. *what is data?* and *why is that important?*): the baseline understanding of what we are dealing with in relation to the double means-end relationship was important and professionals in various roles (business managers, data owners, data stewards, analysts, etc.) all needed to hear that same story in a coherent and consistent manner.

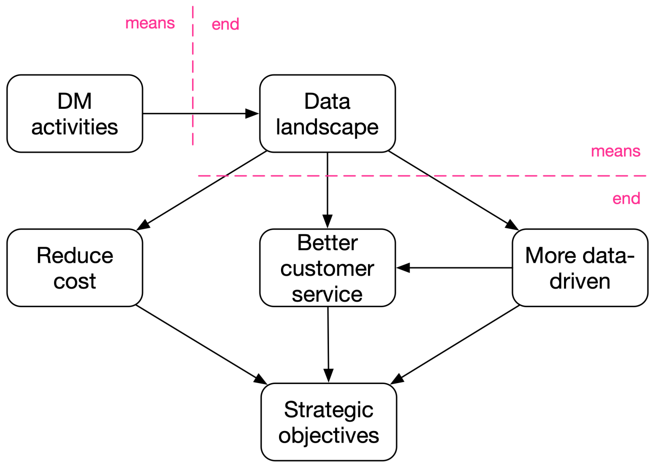

Figure 1. Double means-end

B. Drinking water company

The second case pertains to a drinking water company. We will use the name DWC to refer to this company. The mission of this company is to make water (drinking water as well as water for industrial applications) and distribute it to customers at an acceptable price. It is important to note that there are several autonomous drinking water companies in the Netherlands. They mostly face similar challenges (some variations on a theme, of course) and some are more innovation-minded than others.

Drinking water companies are good at engineering. Building a drinking water factory is highly complex. building a distribution network where quality and pressure of water is carefully arranged for is a daunting task. Even in a country with as much water as the Netherlands, managing the strategic reserve of (sweet) water makes matters even more complex.

In this organization there were 'mixed feelings' about (the use of) data. One the one hand it was said that "we make factories and bring drinking water to our customers, we're not an IT-company." On the other hand, it was also recognized that *a*) we have to report on the amount of data (and pressure) to the regulatory agency, *b*) we can use data to help predict when we need more/less capacity, and *c*) we can use data

[3]We did not compute average scores and standard deviations but used a qualitative assessment to arrive at this conclusion.

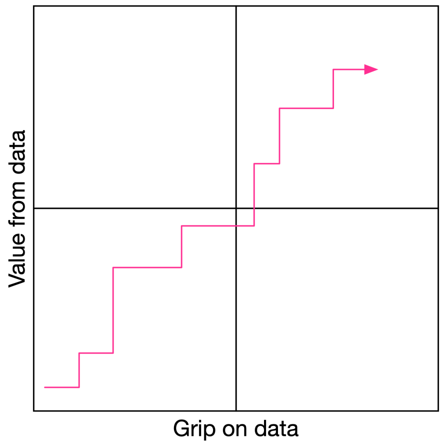

Figure 2. Balanced strategy

(particularly with data science applications) to detect not only that there may be a break down in the distribution system, but also where. Getting traction in such a setting – particularly in a setting where allocation of budgets is key to get things done – was far from easy to say the least.

The double means-end relation that was mentioned in the previous section was at play here too but in a different way: some stakeholders saw it whereas others did not. Meaning: some stakeholders recognized the need to manage data as an asset as an enabler for getting value from (the use of) data, but others did not. This also meant that only part of the group of stakeholders understood that using data *without* effective data management was a gamble at best (and pointless at worst). And at the same time, some stakeholders were willing to invest in basic capabilities for managing data as an asset in order to "get the data house in order" – which fits with an engineering mindset – whereas others were more interested in new ways of using data.

Here, too, as basic understanding of data was missing, particularly the fact that data can be used to give (partial) insight in what otherwise remains hidden (i.e. what happens in the distribution network underground at great distance). A related challenge was the fact that it is difficult to understand that data generated by operational technology (OT) and from administrative applications and 'regular' information technology (IT) can be combined depending on the purpose at hand. The third main challenge had to do with data *structures* and the capabilities of different types of systems. For example, administrative systems are good at keeping track of what happens but not so good at complicated historic data analyses. Classic data warehouse technology is perfectly suited for analyzing and reporting on historic data that is well curated, but perhaps not so much for machine learning and artificial intelligence applications. We again relied heavily on training sessions (both a core group that worked on the vision and road-map for this organization and for a selected group of business managers) to improve the knowledge position. With our core group, we developed a strategy that balanced between the two concerns: grip on data and value from data. This is shown in Fig. 2 which is inspired by [23]. The vision and road-map have been approved. Their realization is on-going.

C. Harbor

The third case pertains to a major harbor in the Netherlands. This case differs somewhat in content from the other two. We were approached to help in the following setting: *a*) the organization was working on a new long-term strategic plan, *b*) it was recognized that data and IT are key enablers for this plan, but at the same time *c*) there were some significant challenges with regard to IT. The request for help was threefold: help us to consider and make a plan for improving our IT processes as well as to help us create a vision a road-map for IT for the long term in light of current challenges and the future direction. Given the sustainable orientation of the strategy of this organization, we also recognize the need for a sustainable approach for the IT landscape (see e.g. [17] in which we published our thoughts on sustainable IS).

We worked with a small group of stakeholders that, combined, represent the most important processes and departments of the organization. The CFO of the organization, who enlisted our help, was also part of our working team. In a series of workshops over a limited amount of time (approximately 8 weeks), we developed the requested deliverables.

In light of the topic of this paper, two things stand out. First, terminology was an issue. Basic terms such as 'business process', 'information system', 'data', 'functionality', etc. were very poorly understood. Particularly the discussion around 'data' versus 'information' turned out to be a tricky one: the latter was more or less equated to 'documents' by one part of our working group whereas the other part of the working group saw 'data' as a representation and 'information' as a mental construct (i.e., the information 'payload' for an actor is the knowledge increment of that actor by studying the document – a way of thinking that is close to that advocated in the FRISCO report [8], as well as elaborated in terms of infons [6] in [29]).

With this difference in understanding in our (small) working group, we felt it was impossible to clearly define the relevant IT-roles such as 'information management' (see again the remark in Section II-A) and document (life-cycle) management, let alone develop a blueprint and road-map for the future IT landscape. Even though we had a limited amount of time for the whole project, we spent a full session on synchronizing terminology.

Second, the means-end relationships were fairly well understood. Our group understood that people, processes, data, etc. are a means to an end. It took only very little discussion to see that people management (HRM), process management (BPM), data management (DM), etc. are means to take care of these 'ends'. This also very quickly lead to the realization that a discussion on the IT landscape does not make sense without considering the process context and vice versa.

This was also helpful in setting up the requested road-map. In light of the sustainability values in the organization *and* the realization that people, processes, data, etc. can be seen as assets with a clear life-cycle, we worked with a two-pronged approach in our road-map. On the one hand we worked with a 'repair list' where we improved existing assets in light of the new vision as expressed by the blueprint. On the other had, we worked with 'back casting' where we attempted to brainstorm gaps and improvement projects and clarify what would need to be done to go from *where we are* to *where we want to be*.

Our involvement with this organization is on-going and includes several aspects, notably: *a*) fleshing out the project charters and road-map in much more detail to help start up projects, *b*) support the project in which the big picture overview of processes and information systems is fleshed out in more detail, *c*) etc.

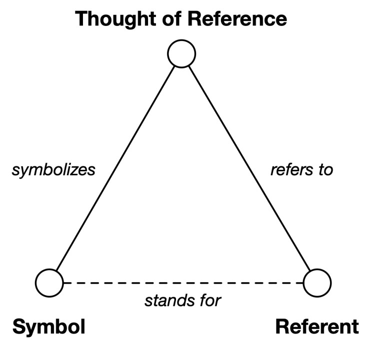

Figure 3. Semiotic triangle

III. LITERATURE

The take-away from the three cases is as follows. First, we found that (business) stakeholders have difficulty with 'basic terminology'. Key constructs such as the distinction between *data* and *information* are difficult. This is the source of further terminological confusion which seriously hampers the discussions/sessions that are intended to help the organization improve and achieve its goals. Given that our key concern is 'data', the first objective for the literature study is to define this elusive term (Section III-A). Second, we noticed that the *story* around the double means-end relationship was really helpful We will explore the literature to better understand this mechanism (Section III-B)

A. Defining data

We used *Web of Science* (WoS) to explore the literature around the definition of 'data' for the last 20 years. We filtered out articles that are available online (either in WoS or via Google Scholar) in English. The primary focus is academic work, yet we also considered leading authors from practice. We scanned abstracts (articles) and summaries (books) to identify the main schools of thought. Note that our objective here is to give an overview of these schools of thought rather than present a full critical literature survey (see e.g. [41]).

Before diving in, a small aside is required. We did find many sources that deal with (the definition of) some specific 'aspect' of data such as: *a*) What is data ethics? *b*) What are the five definitions of data integrity? *c*) What is the synthesis of definitions of big data characteristics? *d*) etc. While these are interesting, they do not directly answer our question: *What is data?* Through snowballing and our earlier work we came up with the overview presented in this section.

The prime author in the *philosophical information theory* school of thought is Floridi. He has published a series of three books on *The Philosophy of Information* (2011), *The Ethics of Information* (2013), and *The Logic of Information* (2019). As the titles suggest, the main topic of study is information. In the preface of the first book it is stated that:

> "*The essential message of the book is quite simple. Semantic information is well- formed, meaningful, and truthful data; knowledge is relevant semantic information properly accounted for; humans are the only known semantic engines and conscious inforgs (informational organisms) in the universe who can develop a growing knowledge of reality; and reality is the totality of information (notice the crucial absence of 'semantic').*," ([9, p. xiii])

Note that, according to this definition, semantic information *is* data (that has certain characteristics) but also that the universe is the totality of information. This begs the question: What do we mean with (non-semantic) information, and how can we define data? In chapter 8 of this work, Floridi acknowledges that a precise definition is difficult to say the least, yet sets out to provide one anyway, related to different underlying philosophies (including [37]). The augmented general definition of information (GDI$\star$) that he presents is as follows:

> "*σ is an instance of semantic information if and only if a) σ consists of n data (d) (for $n \geq 1$) b) that data are well-formed (wfd) c) the wfd are meaningful (mwfd $= \delta$) d) the δ are truthful.*," ([9, p. 104])

The 'normal' general definition of information (GDI) is the same but drops the last clause, suggesting that information need not be truthful. This leaves the definition of data. After a brief discussion which refers (loosely) to semiotics, the notion of 'data' is defined as:

> "*Datum $=_{def}$ x being distinct from y where x and y are two uninterpreted variables and the domain is left open to further interpretation.*," ([9, p. 85])

It is explained the *lack of uniformity* can be in the real world (i.e. the rings of a tree have different color which is a difference that has some meaning), between perception of symbols (i.e. a higher or lower charge in a battery), or between symbols (i.e. different letters in an alphabet). This definition is very general, and our interest is mainly in the latter type of difference: between symbols.

This brings us to the second school of thought from the field of *linguistics* in general and *semiotics* in particular. This view largely hinges on the notion of the 'semiotic triangle' as illustrated in Fig. 3. Simplifying, the main idea behind this triangle is that a symbol represents our understanding of some domain (referent) such that it can stand for that domain [24]. In theory, the notion of a symbol links nicely to the notion of *differences* in the theory of Floridi: a letter on a piece of paper is definitely a symbol and is definitely also a difference (between the white background and the black of the letter). This is often extended with the semiotic ladder [39] which connects the world of IT platforms (physical world, empirics, syntactics) with the world of human information functions (semantics, pragmatics, social world). Note that this way of thinking is more specific than the Floridi school-of-thought: it focuses very specifically on the IT-world which aligns with the focus of this paper.

The work on semiotics was a key component in the development of the FRISCO framework [8]. The *weltanschauung* behind this framework is based on the following assumptions (slightly paraphrased for brevity): *a*) The *world* exists, *b*) Humans are able to observe and perceive *parts* or *aspects* of the world, *c*) Humans are able to form conceptions in their mind. The collection of [...] conceptions in a persons mind is his or her knowledge, *d*) A perceived domain may be conceived as composed of identifiable components, *e*) Some things are conceived as having a static existence (states), while others are conceived as changes of some state (transitions). Hence, a perceived domain may be conceived as having an existence in a temporal context, *f*) Some transitions may be conceived as being performed or brought about by some active things, called actors. Such a transition, called an action, is performed by that actor on passive things, called actands, and *g*) Persons use representations to communicate their conceptions. These conceptions are represented in some language on some medium.

The interesting thing is that the FRISCO framework takes the semiotic triangle as a starting point and extends it. The actor (viewing the domain) is the starting point and views/observes the domain (perception). This leads to thoughts in the brain (conception) that may be represented using some language – which maps on the notion of a *sign* in the semiotic triangle.

Lately, the semiotic triangle has been extended further by various authors . For example, in [31], the notion of the semiotic triangle is extended to a *double semiotic triangle*. The key insight in this work is that one actor may *produce* a sign (that captures our understanding of a domain such that it can stand for that domain) but another actor may both observe the sign that we have produced *and* the original domain, and therefore has to *independently* verify whether the sign can indeed stand for that domain.

Based on this exploration, we conclude that in the context of organizations using data to create value, data: *a*) is a *representation*, *b*) represents our *understanding* of a domain (which is inherently subjective), *c*) such that it an *stand for* that domain, *d*) suggesting that it has both *structure* (syntax) and *meaning* (semantics) – otherwise the assessment that it can stand for a domain cannot be made. We will use this definition as the basis for the exploration of the double means-end relationship[4].

B. Double means-end

A first attempt to explore the literature on the double means-end relationship ('data management' is a means to achieve 'quality data' which is a means to achieve a business strategy) gave very little results. The query "double means-end relationship" lead to a 2003-paper in the area of psychology which is far from useful for our purposes. Reducing the query to "means-end relationship" yielded 123 results. After filtering for relevance, several useful articles remained, all pointing towards 'means-end chain theory'. We will give a brief overview of these theories and illustrate them in light of the objectives of this paper.

This remainder of this section is based on [1, 4, 21, 34, 40]. It should be noted that [1] presents a recent critical literature review of relevant theories.

The 'means-end chain theory' (MEC) stems from the field of customer behavior. It is intended as a theory that helps to understand consumer behavior by building/mapping a cognitive model of this behavior. It is based on a set of assumptions that differs slightly in wording between the various authors. The general idea is that

> "*MEC's main assumption is that people do not buy products for the products' sake, but for the benefit that their consumption can provide.*" ([4])

It connects consumer needs to the consumer perception of the product concept.

Consumers are presumed to be goal-oriented decision makers (This fits well with the purpose of this paper: in the cases that we described, decision makers have to investigate whether/how/how much/where to invest in data/data management.) Specific assumptions regarding consumers are: *a*) consumers buy and use products depending on their evaluation of self-relevant consequences of these behaviors and *b*) consumers are assumed to make voluntary and conscious choices between alternative objects. In light of the work of e.g. Simon [38], the results of MEC should be carefully interpreted: consumers (and decision makers, for that matter) may not be as rational as we think.

The typical approach for *investigating* behavior is the 'laddering technique' which is an interview technique. It is not particularly of interest for this discussion. The interested reader is referred to [21, Section 2.3] for an overview.

The outcome of the analysis is of more interest. Different structures have been described. For example, in [4] we see a mapping approach based on values, consequences, abstract attributes and concrete attributes. By contrast, in [40] we see personal values, consequences, and attributes and in [21] we see attributions, consequences, and values. A common principle that is used is that means-end relationships are mapped out in an hierarchical way as illustrated in Fig. 4.

[4]Note that this definition does not say anything about the shape or form of data. Data can come in the form of a tuple in a relational database, a key-value pair, an XML-tree, a word document, picture, video, etc.

This hierarchical assumption is challenged in [34] where it is concluded that a directed graph may be more appropriate.

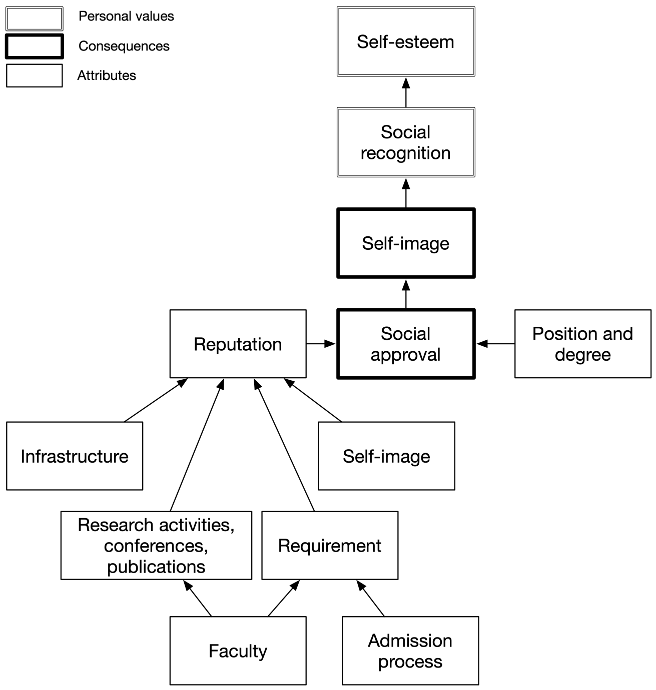

Figure 4. Hierarchical value map of educational services. Adapted (abbreviated) from [40]

Last but not least, it should be noted that in [40], Veludo-de-Oliveira, Ikeda, and Campomar observe that some people experience difficulties when dealing with different abstraction levels during laddering interviews. We hypothesize that this is also true for interpreting the outcome in the form of the aforementioned hierarchical maps. This is particularly relevant because we often see a similar effect when discussing data management aspects as illustrated by the cases in Section II: the difference in abstraction level between processes and tools for collecting specific meta-data, related to the processes where this data is used, the decisions that are based on them, and the (strategic) outcomes that result from these decisions is often hard to understand for professionals.

IV. Framework

From the review of the literature, we have seen that, in the given context, it makes sense to use a definition of 'data' based on semiotics: data is a representation of our understanding of some domain, such that it can stand for that domain. From the review of the literature, we have also seen that means-end thinking is – at least in the field of consumer preferences – common. A useful insight is that the hierarchical nature of such means-end relationships is challenged and that a networked-approach may make more sense.

This means that we have two ingredients for the required framework in place: *a*) a definition of data and *b*) a way to connect data management to data to business goals using the means-end relationships. In light of the three cases, we believe that two things are missing still. First, we need a way to help (business) stakeholders to better understand data. Our experience is that stakeholders find data *a*) abstract, *b*) complex, and often *c*) overwhelming in the amount of data as well as the amount of types of data. This can be overcome by creating useful abstractions that are visualized and explained in a meaningful way. We believe that models (more specifically: conceptual domain models) can help in doing so. In earlier work [31] we presented the definition of a 'domain model' that is also based on the semiotic triangle. Key elements from this definition are: *a*) a model is a *representation* that is externalized, *b*) it is a model because stakeholders/modellers have assessed the fact that the model can stand for (their understanding of) the domain under consideration, and *c*) models are created with different modeling goals in mind, allowing us to differentiate between e.g. models for understanding, models for decision making, etc.

With this in mind, we can essentially[5] define a conceptual data model as "a representation of (our understanding of) a domain, created for the purpose of understanding/conceptualizing that domain, such that it can stand for that domain." If we are careful to take the *understandability* of the model into account, then such conceptual data models should help us to achieve our goal of appointing a tool to help stakeholders understand data. A cursory review of the literature suggests that, when creating the representation/model, we have to take into account such factors as *a*) the language that is used, *b*) consistency in verbalization and abstraction levels, *c*) the size (number of concepts (elements and relationships) in the model, *d*) modularity and consistent abstraction mechanisms, *e*) etc. [2, 36, 43]. In our framework, we do not advocate for a specific meta-model, framework, or even modeling language (e.g. ORM2 [18] versus ER [3] versus UML[25]). Our sole claim is that models can help overcome the hurdle of understanding data.

Second, we need to take into account that we should also cater for non-hierarchical means-end relationships. Given our experience in using the hierarchical model in the three cases, we are inclined to leave this hierarchical nature intact at least partially. The solution lies in extending the way of thinking, and using a stratified approach, as follows:

- In the top layer of the model, we map out the mission, vision, strategy of the organization. This represents the ultimate 'end' that we wish to achieve.
- In the second layer of our model, we map out the 'means' that we can use to achieve that end: people, processes, money, reputation, and of course data. These are only useful if *a*) they are of sufficient quality and *b*) when they are used in harmony, with a clear underlying principle/architecture.
- The former point suggests that taking care of these 'means' is important. This means that underlying disci-

[5]A more nuanced definition is provided in [31].

plines such as HRM, process management, finance, and data management are means to achieve that end.

Taking this into account, we have developed the framework as shown in Fig. 5.

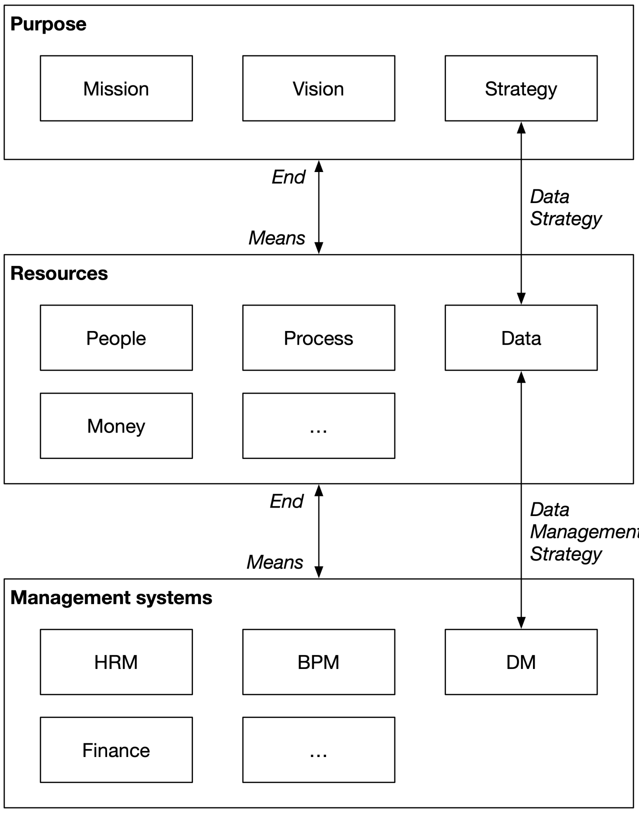

Figure 5. Framework

Note the following key points. First, we have built up our model in three layers: *purpose* of the organization, *resources* that are needed to achieve this purpose, and *management systems* to take care of these resources. This layering visualizes the double means-end relationship. Second, note that in each layer we have listed key elements that we believe to make up this layer. For example, in the resources layer we have listed people, process, data, money etc. In line with the literature on means-end-chains, we take into account that we need to go beyond mere 'vertical' means-end relationships: alignment within each of the layers is also required. We have omitted the visualization of this 'horizontal' alignment for reasons of readability of the diagram. We do wish to point out that links on the resources layer (i.e. data being related to processes, because processes require and produce data) should also be reflected in the layer of the management systems (i.e. BPM and DM should be aligned). We hypothesize that this is where disciplines such as enterprise architecture/business design come in [22, 26, 32, 35].

The framework allowed us to also more clearly define often-used terms: 'data strategy' and 'data management strategy'. The key publication – also used in our own work [13, 14] – in this area is [5]. In this work, data strategy is positioned to have an offense-aspect (value creation with data) as well as a defense-aspect (taking care of data). We argue that this should be split: the data strategy is defined as "the set of strategic choices that are made about using data as an asset to realize a business strategy" whereas a data management strategy is defined as "the set of strategic choices on how to manage data as an asset."

The framework is *abstract* and *high-level*. In the three cases, we learned that it is *sufficiently concrete* to open the discussion about data, and data management, as a means to an end and to help stakeholders understand the double means-end relationship as well as related terminology. The framework was never intended as a guideline for implementation yet it did help in making design decisions about implementations.

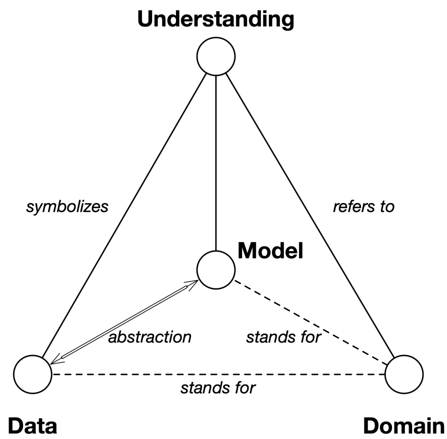

Figure 6. Double semiotic triangle

This leaves the notion of using *(conceptual) models* as a means to understand data, such that it can be used for value creation. Popular data management frameworks such as the DAMA DMBOK (e.g. [20]) have a functional area for data modeling and design. This functional area deals with the techniques to model data (at rest, in motion) in various contexts (e.g. conceptual models for understanding a domain, relational models for the design of transaction systems, star schemas for the design of systems in analytical contexts). We have defined a conceptual data model to be "a representation of understanding of a domain, created for the purpose of understanding/conceptualizing that domain, such that it can stand for that domain." This definition is very similar to our definition of data: "a representation of understanding of a domain such that it can stand for that domain." Both are representations, yet the abstraction level is different. This is shown in Fig. 6. Positioning this double semiotic triangle in our framework (Fig. 5), we observe that the discipline of (conceptual) data modeling, as part of data management, is a means to achieve the end of understanding data – one of the key resources to achieve a business strategy.

V. Reflection

This leaves us with a reflection of the used framework in light of the three cases that were presented in Section II. We will follow the same structure as before.

A. Pension fund provider

One of the key insights in this project was the double means-end relationship. If we look more closely at Fig. 1 then we can already see hints of the fact that a simple hierarchical double means-end (as was traditional in means-end chains) is too simplistic. The diagram shows that data management activities are a means to achieve the end of an effective data landscape which, in turn, is the enabler to attain certain goals. Moreover, these goals are also interrelated: "becoming more data driven" is related to "better customer service." The emphasis of the engagement was on data. Other aspects (i.e. processes, systems) were placed out of scope – deliberately. However, they did show up during the engagement more than once: it seems impossible to talk about (managing and using) data without considering the context in which people use/produce data (processes, systems). In the follow-up engagements we mainly focused on training. During these training sessions, particularly with the data owners of the organization, we discussed this 'broader' perspective more than once.

The framework, particularly with the double means-end relationship, as proposed in this paper would certainly have helped to position and align certain discussions. Whether the explanation about (conceptual) data models would have helped remains to be seen. Our assessment is that there was such a big aversion against modeling that the topic was more or less 'off limits' for a reason.

B. Drinking water company

The case of the drinking water company was very different, if only because the culture of such organizations is different: we typically deal with engineers. The whole topic of data/data management was a difficult one yet – opposite to what happened at the pension fund provider – the use of (conceptual) models was better understood (and in some cases: accepted).

One of the key discussions that we had pertained to the balanced strategy between grip on data (data management strategy) and value creation with data (data strategy). We did not have the diagram of Fig. 5 available at the time, but we essentially told this story. One of the 'horizontal linkages' that was important in this project was the link between IT and OT as explained previously. With simple models (both data models and architecture models), we were able to show the linkages between these two worlds as well as the need for an integral approach to managing data as an asset.

Truth be told: the expectation was that we would create more (conceptual) data models as part of this engagement. Different priorities and difficulties in getting the topic of *data* on the management agenda prevented us from doing so. Whether more elaborate models would have helped (or potentially: hindered) our cause remains to be seen.

C. Harbor

In the harbor case, we *had* to deal with the definition-discussion, getting a shared understanding of data, information, information system, the information-handling part of business processes, etc. We did rely on a (previous incarnation) of our double semiotic triangle as shown in Fig. 6 as the foundation for this discussion. Our diverse stakeholder group (including a project manager, analyst, security specialist and other roles) quickly saw that a clear definition of terms would help in speed and quality of the discussions. This was mentioned explicitly in the after-action review.

The topic for this engagement was broad. Initially we had a focus on the IT landscape and associated processes. Here, we explicitly took the main business processes/value streams into account in order to have sufficient *context* to understand IT and data. Using a(n previous) incarnation of the thinking behind the double means-end framework did help in coming up with a balanced solution to the case. The ability to connect the dots between processes, data, and systems on the one hand, and supporting functions such as process management, data management, and IT management on the other helped in the quality of the final result.

VI. Conclusion

The motivation for writing this paper lies in the fact that we see many organizations struggle to 'get value from their data'. The line of thinking that we have developed in practice is: in order to help organizations to tackle their (data) challenges, we need to get across that (making strategic decisions on) taking care of data is key for (making strategic decisions on) getting value from data. In our practical work, we had good experience with using the line of thinking that we dubbed *the double means-end relationship*. To help stakeholders see the point, we have relied heavily on models, mostly of the boxes-and-arrows sort, to convey key points. The objective of this paper is to develop a framework that unifies these ideas such that it becomes a basis to guide the thinking of stakeholders when dealing with (data) challenges.

From the literature review, we learned that *a*) defining key terms, even as basic as 'data' and 'information' is far from easy, *b*) it makes sense to see 'data' as a representation of our understanding of a domain such that it can stand for that domain, *c*) means-end thinking is common and useful, yet we should combine *vertical* and *horizontal* linkages to get a full understanding. The last point (needing also *horizontal* linkages) is a considerable extension to our thinking. We are very much aware of the *architecture perspective* (paraphrased: the way of thinking that focuses on the essentials, the big pictures, from various perspectives) as can be seen from our earlier publications (e.g. [12, 15, 16, 30, 33]). Yet, we had not yet connected that sufficiently to the proposed framework.

Reflecting on the three cases, we conclude that the framework is useful when organizations embark on their data/data management journey.

To extend the framework, we propose to at least stick to the three layers that we presently have: this seems sufficiently rich to describe most aspects of the enterprise (Pareto rule) while simple enough for business stakeholders to understand. The main challenge is to work in the horizontal linkages. We intend to do this by the architecture discipline in the framework to ensure that both horizontal and vertical linkages are considered when addressing (data) challenges of the enterprise.

Further validation of the framework is a two-sided challenge. On the one hand, we want an 'expert perspective' on the framework to assess completeness, correctness, and expected usefulness of the framework. The group of experts should straddle academia and practice to get a sufficiently rich perspective. On the other hand, we also need to take into account that 'the proof of the pudding is in the eating.' Meaning: we need to road-test our framework by applying it in the field and assess its usefulness as objectively as possible. This requires a multiple case study and is part of future research.

References

[1] E. Borgardt. "Means-End Chain Theory: a Critical Review of Literature." In: *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 64.3 (2020), pp. 141–160.

[2] Hans F. M. Burg and R. P. van de Riet. "The impact of linguistics on conceptual models: consistency and understandability." In: *Data & Knowledge Engineering* 21.2 (1997), pp. 131–146. ISSN: 0169-023X.

[3] Peter Pin-Shan Chen. "The Entity-Relationship Model: Toward a Unified View of Data." In: *ACM Transactions on Database Systems* 1.1 (Mar. 1976), pp. 9–36.

[4] A. I. A. Costa, M. Dekker, and W. M. F. Jongen. "An Overview of Means-End Theory: Potential Application in Consumer-Oriented Food Product Design." In: *Trends in Food Science & Technology* 15 (2004), pp. 403–415.

[5] L. DalleMule and T. H. Davenport. "What's your Data Strategy? The key is to Balance Offense and Defense." In: *Harvard Business Review* (May 2017), pp. 112–121.

[6] Keith Devlin. *Logic and Information*. Cambridge Tracts in Theoretical Computer Science. Cambridge, UK: Cambridge University Press, 1991. ISBN: 0-521-49971-2.

[7] Jan L. G. Dietz and Hans B. F. Mulder. *Enterprise Ontology: A Human-Centric Approach to Understanding the Essence of Organisation*. The Enterprise Engineering Series. Berlin, Germany: Springer, 2020. ISBN: 978-3-030-38854-6. URL: https://doi.org/10.1007/978-3-030-38854-6.

[8] E. D. Falkenberg et al. *A Framework of Information Systems Concepts*. IFIP WG 8.1 Task Group FRISCO, IFIP, 1998.

[9] L. Floridi. *The Philosophy of Information*. Oxford University Press, 2011.

[10] L. Floridi. *The Ethics of Information*. Oxford University Press, 2013.

[11] L. Floridi. *The Logic of Information*. Oxford University Press, 2019.

[12] B. van Gils. "Strategy and Architecture - Reconciling Worldviews." In: *Advances in Enterprise Engineering II - First NAF Academy Working Conference on Practice-Driven Research on Enterprise Transformation (PRET 2009), held at CAiSE 2009*. Ed. by H. A. Proper, F. Harmsen, and J. L. G. Dietz. Vol. 28. LNBP. Springer, 2009.

[13] B. van Gils. *Data Management, a Gentle Introduction. Balancing Theory and Practice*. 1st ed. Van Haren Publishing, 2020.

[14] B. van Gils. *Data in context. Models as Enablers for Managing and Using Data*. The Enterprise Engineering Series. Springer, 2023.

[15] B. van Gils and S. van Dijk. *The Practice of Enterprise Architecture - Experiences, techniques, and best practices*. BiZZdesign, 2014.

[16] B. van Gils and S. van Dijk. *ArchiMate : from theory to practice*. BiZZdesign, 2015.

[17] B. van Gils and H. Weigand. "Towards sustainable digital transformation." In: *2020 IEEE 22nd Conference on Business Informatics*. Vol. 1. IEEE. 2020, pp. 104–113.

[18] Terry A. Halpin and Tony Morgan. *Information Modeling and Relational Databases*. Morgan Kaufmann, 2010.

[19] C. Henderson and N. Venkatraman. "Strategic Alignment: Leveraging Information Technology for Transforming Organizations." In: *IBM Systems Journal* 32.1 (1993), pp. 472–484.

[20] D. Henderson, ed. *DAMA DMBOK - Data Management Body of Knowledge*. Technics Publications, 2017.

[21] Y. Jung and H. Kang. "User Goals in Social Virtual Worlds: A Means-End Chain Approach." In: *Computers in Human Behavior* 26 (2010), pp. 218–255.

[22] J. Lapalme. "Three schools of thought on enterprise architecture." In: *IT Professional* 14.6 (2011), pp. 37–43.

[23] N. van der Meulen, P. Weill, and S. Woerner. "Managing Organizational Explosions During Digital Business Transformations." In: *transformation* 1 (2020), p. 2.

[24] C. K. Ogden and I. A. Richards. *The Meaning of Meaning – A Study of the Influence of Language upon Thought and of the Science of Symbolism*. Oxford, United Kingdom: Magdalene College, University of Cambridge, 1923.

[25] OMG. *UML 2.0 Superstructure Specification – Final Adopted Specification*. Tech. rep. ptc/03-08-02. Needham, Massachusetts, USA: Object Management Group, Aug. 2003.

[26] Martin Op 't Land et al. *Enterprise Architecture: Creating Value by Informed Governance*. The Enterprise Engineering Series. Berlin, Germany: Springer, 2009. ISBN: 978-3-540-85232-2. URL: https://doi.org/10.1007/978-3-540-85232-2.

[27] Martin Op 't Land et al. "Positioning Enterprise Architecture." In: *Enterprise Architecture: Creating Value by Informed Governance*. The Enterprise Engineering Series. Berlin, Germany: Springer, 2009. Chap. 3. ISBN:

978-3-540-85232-2. URL: https://doi.org/10.1007/978-3-540-85232-2_3.

[28] M. M. Parker and R. J. Benson. "Enterprisewide Information Management: state-of-the-Art Strategic Planning." In: *Journal of Information Systems Management* 6.3 (1989), pp. 14–23.

[29] Henderik A. Proper and Peter D. Bruza. "What Is Information Discovery About?" In: *Journal of the American Society for Information Science* 50.9 (July 1999), pp. 737–750. URL: https://doi.org/10.1002/(SICI)1097-4571(1999)50:9%3C737::AID-ASI2%3E3.0.CO;2-C.

[30] Henderik A. Proper, Bas van Gils, and M. Kazem Haki, eds. *Digital Enterprises – Service-Focused, Digitally-Powered, Data-Fueled*. The Enterprise Engineering Series. Berlin, Germany: Springer, 2024. ISBN: 978-3-031-30214-5. URL: https://doi.org/10.1007/978-3-031-30214-5.

[31] Henderik A. Proper and Giancarlo Guizzardi. "On Domain Conceptualization." In: *Advances in Enterprise Engineering XIV – 10th Enterprise Engineering Working Conference, EEWC 2020, Bozen-Bolzano, Italy, September 28, October 19, and November 9-10, 2020, Revised Selected Papers*. Ed. by David Aveiro et al. Vol. 411. Lecture Notes in Business Information Processing. Berlin, Germany: Springer, 2021, pp. 49–69. ISBN: 978-3-030-74196-9. URL: https://doi.org/10.1007/978-3-030-74196-9_4.

[32] Henderik A. Proper, Roel Wagter, and Joost Bekel. "On Enterprise Coherence Governance with GEA: A 15-Year Co-Evolution of Practice and Theory." In: *Software and Systems Modeling* (2022). ISSN: 1619-1374. URL: https://doi.org/10.1007/s10270-022-01059-0.

[33] Henderik A. Proper et al., eds. *Architectural Coordination of Enterprise Transformation*. The Enterprise Engineering Series. Berlin, Germany: Springer, 2018. ISBN: 978-3-319-69584-6. URL: https://doi.org/10.1007/978-3-319-69584-6.

[34] J. van Rekom and B. Wierenga. "On the Hierarchical Nature of Means-End Relationships in Laddering Data." In: *Journal of Business Research* 60 (2007), pp. 401–410.

[35] J. Ross, C. M. Beath, and M. Mocker. *Designed for digital: How to architect your business for sustained success*. Management on the Cutting Edge. MIT Press, 2019.

[36] M. Serrano et al. "Metrics for data warehouse conceptual models understandability." In: *Information and Software Technology* 49.8 (2007), pp. 851–870.

[37] C. E. Shannon. "A Mathematical Theory of Communication." In: *Bell system technical journal* 27.3 (1948), pp. 379–423.

[38] H. Simon. "A behavioral model of rational choice." In: *Models of man, social and rational: mathematical essays on rational human behavior in a social setting*. Wiley, 1957.

[39] R. K. Stamper. "The semiotic framework for information systems research." In: *Information systems research: Contemporary approaches and emergent traditions* (1991), pp. 515–528.

[40] T. M. Veludo-de-Oliveira, A. A. Ikeda, and M. C. Campomar. "Discussing Laddering Application by the Means-End Chain Theory." In: *The Qualitative Report* 11.4 (Dec. 2006), pp. 626–642.

[41] J. Webster and R. T. Watson. "Analyzing the Past to Prepare for the Future: Writing a Literature Review." In: *MIS Quarterly* 26.2 (June 2002), pp. xiii–xxiii.

[42] M. van der Winden, B. van Gils, and H. Weigand. "Improving the Data Management Capability at a Dutch Pension Fund Service Provider." In: *2021 IEEE 23rd Conference on Business Informatics (CBI)*. Vol. 1. IEEE. 2021, pp. 181–190.

[43] S. Zugal et al. "Assessing the impact of hierarchy on model understandability–a cognitive perspective." In: *Models in Software Engineering: Workshops and Symposia at MODELS 2011, Wellington, New Zealand, October 16-21, 2011, Reports and Revised Selected Papers 14*. Springer. 2012, pp. 123–133.