Influence of brokering in Data Warehouse projects

Helena Vranesic

Goethe University Frankfurt, vranesic@wiwi.uni-frankfurt.de

Follow this and additional works at: http://aisel.aisnet.org/amcis2010

Recommended Citation

http://aisel.aisnet.org/amcis2010/178
Influence of brokering in Data Warehouse projects

Helena Vranesic
Goethe University Frankfurt
vranesic@wiwi.uni-frankfurt.de

ABSTRACT
Data Warehouse (DWH) projects bring together different communities of practice to create one body of knowledge and help increase companies’ competitive advantage. In this paper we argue that a “communities of practice” perspective can better envisage the problems often surfacing during the requirements elicitation in DWH development by drawing attention to the processes on the borders between involved communities. To investigate these phenomena more closely, exploratory, open interviews were conducted with experienced DWH-professionals developing DWHs for diverse business domains. Based on the results gathered, we argue that brokers have to be considered as relevant facilitators of the requirements elicitation process in any DWH project.

Keywords
Data Warehouse, Communities of Practice, Brokering, Boundary Objects

INTRODUCTION
In the past, researchers and practitioners have analyzed both social and technical factors affecting success of data warehouse (DWH) projects. In a recent study about the factors affecting DWH success, Hwang and Xu (2008) summarized past results (Chen et al. 2000; Shin 2003; Watson and Haley 1997; Wixom and Watson 2001) displaying a rare piece of evidence how operational factors and economic success factors have a positive effect on DWH quality and success. However, Rizzi et al. (2006) note the absence of effective techniques for collecting information requirements and quality-of-service requirements as well as for translating those requirements into conceptual models based on a common vocabulary between IT experts and decision-makers. Similarly, Jarke et al. (2009) argue that a deeper articulation of individual and group information requirements is missing in distributed projects. In response to the increased distribution of requirements analysis processes, Hansen and Lytyinen (2009) suggest a model to synthesize propagation between social and structural distributed cognition in contemporary requirements practice. However, their propositions only illustrate the current “state-of-affairs” and give no guidance for overcoming disagreements during the requirements elicitation. Due to the participation of different communities of practice (CoPs) (Wenger 1998) in the development process, these discontinuities are also revealed in the DWH domain. Nevertheless, despite a growing amount of research on DWH project management (Vassiliadis et al. 2001), success factors of DWH projects (Chen et al. 2000; Hwang and Xu 2008; Shin 2003; Watson and Haley 1997; Wixom and Watson 2001) and methodologies for DWH requirements analysis (Winter and Strauch 2003), we still lack a thorough understanding how interaction between these communities influences the development process per se.

We distinguish two major CoPs confronting each other in DWH projects:

1. operative system professionals (OSPs) with knowledge of legacy and source systems, and
2. business experts in decision-making fields (BEDFs) who define business requirements.

The connection between these two practices is made by introducing a third practice, the DWH-professionals responsible for the DWH development. As CoPs are characterized by shared histories of learning (Wenger 1998) and known to create discontinuities between those who participate in the community’s work and those who do not (Wenger 1998), DWH-professionals are confronted with communication gaps on both boundaries: towards BEDF and OSP communities. Therefore, we have concentrated our research on the DWH community as a main CoP in DWH projects, closely investigating the bordering conditions for DWH-OSP and DWH-BEDF borders. We also assume that the creation of a mutual understanding in the early phase of DWH projects, i.e., requirements analysis, plays an important role in its success. Consequently, we intend to answer the following research questions:

“What communication barriers emerge between communities involved in contemporary DWH projects?”

“What are necessary bordering conditions that help mitigate the communication barriers emerging between these communities?”
To investigate processes on the borders between communities, we conducted exploratory, open interviews with experienced DWH-professionals developing DWHs for diverse business domains. The interviewees’ responses provided information on multiple of their previous or current projects. For these projects, we observed phenomena that could help mitigate the communication barriers emerging between participating CoPs.

This paper represents a step toward the answer to the stated research questions by investigating the influence of brokering in contemporary DWH projects. Our contribution lies in the novel approach taken to identify communication barriers in DWH projects from a CoP perspective on both the DWH-OSP and the DWH-BEDF boundaries. The formulation of four specific hypotheses that are deduced plausibly from the exploratory study will serve as a basis for future research.

The remainder of the paper is structured as follows. The following section discusses challenges of a requirements phase reported for conventional information systems (IS) and projects these issues to the DWH domain. Section 3 presents selected methodology used throughout this research. Section 4 presents the results from our analysis. Consequently, we summarize our findings and limitations arguing that the existence of brokers accelerates the process of requirements elicitation. We conclude by giving an outlook on further research.

RELATED WORK

As early reported by Winter and Strauch (2003), information requirements analysis for DWH systems is generally comparable to the requirements and analysis phase for conventional IS, but it sometimes differs significantly. Certainly, though, there is still no consensus on what relevance and temporal priority should be assigned to the requirements analysis (Giorgini et al. 2008). A poor analysis can stem not only from an invalid design approach but also from the failure of interpreting and understanding the users’ terminology, concepts, viewpoints and goals (Nuseibeh and Easterbrook 2000). A recent field study conducted by Hansen and Lyytinen (2010) reports on the key challenges design professionals experience in the elicitation, specification, and management of IS requirements.

According to their study, the first group of challenges, the so-called limits of individual cognition, were presented as a key foundation for the socially-based challenges that emerge when multiple stakeholder groups interact in the design process. The challenges stem from inability of individual users or relevant stakeholders to articulate their needs concisely due to the differing perspectives between users and designers or the complexity of the system landscape. Another relevant limit of individual cognition is the difficulty of envisioning a future that differs substantially from their present environment. As DWH systems should support actual information needs by decision makers and knowledge workers as well as provide a means to meet future, presently unknown, information requirements, DWH development clearly contains all the above-mentioned challenges.

The second group contains challenges based on interpersonal processes through which requirements are identified, specified, and managed. Among these challenges, the problematic relationship between business and IT staff stands out: “It has repeatedly been observed that business and IT professionals ‘speak different languages’ and apply different yardsticks for desired outcomes” (Hansen and Lyytinen 2010). Therefore, limits of individual cognition (e.g. articulation) most certainly influence the business-IT relationship and the knowledge transfer between the two sides. The part of knowledge that can be articulated (explicit knowledge) can be captured, stored or sent, whereas the remaining tacit knowledge is less quantifiable (Hildreth et al. 1999). Seeing explicit and tacit knowledge as a duality, rather than opposing sides of knowledge, can be a fertile ground for explaining the failure of requirements elicitation within the analysis phase of the DWH project if taken as a simple capturing of articulated user’s knowledge. By pure reification of the captured requirements, the explicit aspect of knowledge gets codified and stored, neglecting the tacit aspect (Hildreth and Kimble 2002). The two constituent processes of participation and reification have to be in balance (Wenger 1998). While in social communities participation shapes our experience, through the process of reification aspects of human experience and practice are congealed into fixed forms and given the status of objects. If knowledge is predominantly only tacit, the participation proportion of the duality will be higher. Conversely, as the explicit part of the knowledge prevails, a larger proportion of reification is possible. Although the process of reification produces objects that can move across boundaries (Wenger 1998), they can be understood only by participants of a community sharing the same tacit knowledge. Likewise, as objects can cross the boundaries of communities, members of one CoP can also participate in multiple CoPs at once and help to introduce those reified objects from one CoP to another whenever required. There are, thus, two forms of boundary connections: boundary objects and brokering (Brown and Duguid 2001; Wenger 1998).

Star and Griesemer (1989) define boundary objects as “both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites”. In the context of DWH projects, the artifacts (e.g. shared documents) exchanged between CoPs (Brown and Duguid 2001) can potentially become boundary objects if they belong to (at least) two different CoPs. Brokering provides missing coordinating connection between
CoPs and includes activities by individuals that involve facilitating transactions and the flow of knowledge between CoPs (Pawlowski et al. 2000). In DWH projects, DWH-professionals have brokering roles. They translate, coordinate and align different perspectives on both boundaries to OSPs and BEDFs and prevent communication gaps. To fill in gaps, DWH-professionals actively facilitate a negotiation-of-meaning process between all involved parties with help of boundary objects. We assume that boundary objects supporting this process could foster quicker discovery of misunderstandings between the participants and thereby accelerate the process of creating a joint language community (Kamlah and Lorenzen 1984) enabling the consequent transfer of knowledge between CoPs. We also assume that the brokering role is not delegated only to DWH-professionals, but could also be performed by other participants of DWH project's communities.

RESEARCH METHODOLOGY

We collected data through open interviews conducted with nine DWH developers having on average eight years of DWH development experience. All participants have a degree in computer science/informatics. The interviewees were encouraged to express their thoughts on any topics they felt were relevant to the requirements analysis processes. To gain insights based on data analysis, we focused on similarities and differences in fourteen DWH projects of various sizes, ranging approximately from four to thirty participants. Each project the interviewees worked on represented the unit of data analysis. The interviewees reported on their experiences with current or recent projects in the following economy branches: energetics (3), telecommunications (3), public health insurance (1), higher education (2), trading (1), banking (3), and insurance (1).

The aim of the data analysis was to identify and understand what role diverse objects and brokering processes could have in communication between bordering communities participating in requirements analysis in different DWH projects. Since this was an exploratory study, we did not approach the data with a preconceived theoretical perspective. Rather, we drew upon our findings from interviews.

The open interviews, which lasted 30-90 minutes (60 minutes on average), were transcribed to support formal analysis of the data. Interview transcripts were coded using MAXqda (MAXQDA 1995). The interview protocol served as the preliminary coding structure for the data. However, as specific themes began to surface in the coding process, additional interviews were conducted via chat and e-mail. The code structure was iteratively revised until the researchers determined that all relevant themes or issues were reflected (Eisenhardt 1989). After each round of coding, we conducted separate (two-person) coding using the same set of codes on the transcripts. Differences were resolved by detailed discussions.

PRELIMINARY FINDINGS

As we analyzed interviews from the CoP perspective, we have investigated both the DWH-OSP and the DWH-BEDF boundaries.

Boundary between DWH-professionals and OSP community

In several analyzed projects, members of OSP and DWH communities were employees of the same company. In order to establish if this had any effect on bordering processes we have categorized DWH development projects into two groups: in-house projects carried out by the DWH department within the company and outsourced projects carried out by an external consulting company. Nine out of fourteen projects were identified as outsourced projects (Table 1).

Even though outsourced DWH development projects involved DWH consultants external to the company’s community and misunderstandings or misinterpretations of transactional system data were thus expected to occur in a greater number than with in-house projects, results of our analysis revealed that interviewees (members of DWH community) commonly reported only minor communication problems in both cases. If misinterpretation of exchanged objects had occurred, DWH-professionals usually obtained help by directly contacting the responsible people. We assume that the reason for this lies in the fact that both groups share a common computer science background. One interviewee even added: “We understand each other very well with the IT department, because we speak ‘the same language’”. With reference to Hansen and Lyytinen (2010) it is, therefore, plausible to assume the communication border between OSP and DWH community to be positioned on the syntactical level, where, according to Carlile (2004), a common lexicon suffices to specify the differences and dependencies of consequences at the boundary. On this level, if communities seek better mutual understanding, they only need to communicate more (Carlile 2002). Nevertheless, relevant technical, political or organizational bordering issues were reported in almost all interviews, e.g. problems with data delivery from source systems, data quality, non-cooperating external transactional database contractors etc., as frequently mentioned in literature (Shin 2003; Vassiliadis et al. 2001). Table 1 lists the causes for projects exceeding the planned deadlines.

Based on the previous discussion, we summarize:
(H1) The border between OSP and DWH participants is syntactical and participants share a same common lexicon.

(H2) Problems of technical, political or organizational nature are present on the border between OSP and DWH participants and can influence project success.

<table>
<thead>
<tr>
<th>Project &amp; Subprojects</th>
<th>DWH brokers</th>
<th>BEDF brokers</th>
<th>In-house (I)/ Outsourced (U)</th>
<th>Process finished:</th>
</tr>
</thead>
<tbody>
<tr>
<td>TELCOM01_subproject1</td>
<td>Yes</td>
<td>I</td>
<td></td>
<td>O:Change in requirements definitions</td>
</tr>
<tr>
<td>TELCOM01_subproject2</td>
<td>Yes</td>
<td>Yes</td>
<td>I</td>
<td>IT</td>
</tr>
<tr>
<td>PUBSER04</td>
<td>Yes</td>
<td>O</td>
<td></td>
<td>O:Problem with delivery of source system data; Dirty data</td>
</tr>
<tr>
<td>NUTCOM01</td>
<td>Yes</td>
<td>O</td>
<td>IT</td>
<td></td>
</tr>
<tr>
<td>MONINST02</td>
<td>Yes</td>
<td>O</td>
<td>IT</td>
<td></td>
</tr>
<tr>
<td>INSCOM01</td>
<td>Yes</td>
<td>O</td>
<td>O:Problem with delivery of source system data; Dirty data</td>
<td></td>
</tr>
<tr>
<td>TELCOM02</td>
<td>Yes</td>
<td>O</td>
<td>IT</td>
<td></td>
</tr>
<tr>
<td>MONINST03</td>
<td>Yes</td>
<td>O</td>
<td>O:New transactional system was introduced; Dirty data</td>
<td></td>
</tr>
<tr>
<td>PUBSER03</td>
<td>Yes</td>
<td>Yes</td>
<td>I</td>
<td>IT</td>
</tr>
<tr>
<td>PUBSER02</td>
<td>Yes</td>
<td>Yes</td>
<td>I</td>
<td>IT</td>
</tr>
<tr>
<td>TELCOM01_subproject3</td>
<td>Yes</td>
<td>I</td>
<td>O:Unclear source data definition</td>
<td></td>
</tr>
<tr>
<td>SOFTCOM01</td>
<td>Yes</td>
<td>I</td>
<td>IT</td>
<td></td>
</tr>
<tr>
<td>ENERCOM01_subproject1</td>
<td>Yes</td>
<td>O</td>
<td>IT</td>
<td></td>
</tr>
<tr>
<td>ENERCOM01_subproject2</td>
<td>Yes</td>
<td>O</td>
<td>IT</td>
<td></td>
</tr>
<tr>
<td>ENERCOM01_subproject3</td>
<td>Yes</td>
<td>O</td>
<td>IT</td>
<td></td>
</tr>
<tr>
<td>MONINST01</td>
<td>Yes</td>
<td>O</td>
<td>IT</td>
<td></td>
</tr>
<tr>
<td>ENECOM02</td>
<td>Yes</td>
<td>O</td>
<td>O:Change in reporting tool</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Listing of relevant findings from the analyzed projects

**Boundary between DWH-professionals and BEDF community**

Since DWH-professionals usually cannot receive completely determinable sets of requirements specifications (Winter and Strauch 2003), requirements elicitation is a process depending upon the actual or perceived need for information to reduce uncertainty, therefore involving the integration of the DWH-professionals’ own knowledge with the knowledge of the BEDF community. Hence, requirements elicitation is not merely a technical task, but rather a task that requires a great deal of work in creating and maintaining social relationships and often intrinsically involves a process of brokering. In order to pursue further analysis of the collected data, we took into account Wenger’s definition of brokering as: “use of multimembership to transfer some element of one practice into another” (Wenger 1998) and investigated the DWH-BEDF boundary in more detail.

Participation in requirements elicitation necessitates “a clear definition of business needs” (Hwang and Xu 2008) exposing DWH-professionals to problems that could be outside the realm of their competence in the customer’s business domain. However, several of our interviewees working either on in-house or outsourced DWH development projects reveal that the DWH community responded to this challenge by bringing in individual members of the community who were most familiar with the customer’s business. Those individual members are brokers according to the above definition – we refer to them...
more specifically as **DWH brokers** – as they assumed the boundary role according to Aldrich and Herker (Aldrich and Herker 1977) in requirements elicitation, acting as both filter and facilitators:

“If you have experience in the customer’s business branch, then you partially know what customers want from you. Although they have problems expressing themselves, you can adjust their statements and produce more fitting reports.” (Filtering aspect)

“We knew roughly how the DWH model should look like so we asked the ‘right’ questions.” (Facilitating aspect)

We assume that in the initial phase of requirements elicitation as well as in phases when the customer’s business becomes more complex and/or diverges from previous knowledge possessed by the DWH community, having the right DWH broker could moderate the negative effects for the project from exceptions or changes in the customer’s requirement definitions. In consequence, requirements elicitation should generally become less difficult and complex, leading to fewer iteration cycles of DWH design validation.

Recordings also revealed that BEDF members who have already conducted data analysis and were familiar with meaning of the data required for further DWH development – we refer to them as **BEDF brokers** – were able to articulate future system requirements better and thereby help DWH-professionals elicit requirements:

“Perhaps they have worked with similar [reporting] systems before or they have managed to get the same functionality they later received with the new system by using, e.g., Excel. The goal is the understanding of data, and you generally have to do that on your own. Sometimes such people unexpectedly help you.”

The above discussion can be summed up by the following two hypotheses:

(H3) Requirements elicitation is conducted with fewer iterations and more accurately when DWH-professionals are also part of the business community, i.e., when there are DWH brokers involved.

(H4) Requirements elicitation is conducted with fewer iterations and more accurately when BEDF members are familiar with data coming from transactional source systems, i.e., when there are BEDF brokers involved.

In the next section we introduce taxonomy of brokering in DWH projects and analyze the role of DWH and BEDF brokers in reported DWH Projects.

**Brokering in DWH projects**

The results from our analysis revealed two groups of brokers: (1) DWH brokers – DWH-professionals with an understanding of the customer’s business domain and (2) BEDF brokers – members of the BEDF community with an understanding of the data delivered by existing IS needed for further DWH development (Table 1).

However, based on the collected data, we were not able to discern how well DWH brokers understand the customer’s business, nor how well BEDF brokers understand data delivered by existing IS needed for further DWH development. Nevertheless, we found evidence that some degree of brokering does exist. In case of in-house project PUBSER03 (Table 1), the DWH team contained employees from the contracting company, implying therefore a good understanding of the customer’s business domain within the team:

“The DWH development team also included the employees of the company, for which we were developing the DWH in the first place.”

On the contrary, in most cases we found evidence of DWH-professionals who thought to be familiar with the customer’s business, but turned out to be wrong:

“We had a rough idea how the DWH should be built even before we met with the clients. After the meeting, we had to adjust it to their specifics”

We had an opportunity to analyze the role of the DWH brokers in the reported projects from the perspective of DWH-professionals because all interviewees were DWH-professionals. As we intended to find a link between brokering and the project’s course, we focused our analysis on the relation between:

1. the DWH-professionals’ perceived knowledge of the customer’s business process (X-axis of Figure 1.a) and
2. the DWH-professionals’ estimation of the customer’s requirements (Y-axis of Figure 1.a).
We were not able to analyze the role of BEDF brokers in DWH projects, due to the fact that no interviews with BEDFs were conducted. However, influence of BEDF brokers cannot be excluded. Therefore, if BEDF brokers were mentioned in interviews, we have included them in our analysis.

**Dynamics of DWH brokering - DWH-professionals’ perspective**

The X-axis from Figure 1 represents the degree of the DWH-professionals’ familiarity with the users’ business, while the Y-axis represents the degree to which DWH-professionals estimated the customers’ requirements.

The DWH projects start out in quadrant I or IV in case DWH-professionals perceive business processes as familiar, thus indicating the presence of DWH brokers within the DWH community. Concurrently, DWH-professionals may also perceive that some members of the BEDF community show an understanding of the data required for further DWH development (BEDF brokers). However, both perceived knowledge about the business process as well as identification of BEDF brokers do not assure the presence of a shared common lexicon between DWH and BEDF communities which is needed for a good estimation of the customer’s requirements (e.g., according to our interviews, INSCOM01 from Table 1. starts in quadrant IV and then moves with the time to quadrant I).

Due to the complex environment of DWH projects, the problem of mutual understanding is not just a question of syntax: to assume that all the stakeholders have an aligned understanding of the used terms from the start is misleading. Briefly stated, a semantic heterogeneity can exist when data is defined differently by different users (March and Hevner 2007), which is often the case in DWH projects. Basically, the problem is not syntactical, but one of *pragmatically creating* a mutual understanding and of *semantically* describing the data. In this context, “mismatches” occur due to the misalignment between the type of boundary faced and the capacity of the knowledge process occurring on the boundary (Carlile 2004). DWH-professionals and BEDFs will try to use the old syntax to transfer knowledge, but they actually need to transform and align the meanings of the terms they use first. Therefore, they initially attempt to agree upon common meaning of the terms (semantical border). If a resulting terminology does not prevent poor estimation of customers’ requirements definition, communities confront on a pragmatic boundary, and proceed with negotiation and transformation of both common knowledge and domain-specific knowledge used in the past (Carlile 2004). To facilitate these processes, DWHs and BEDFs use boundary objects. Depending on the type of boundary faced, though, boundary objects with different capacities are required (Carlile 2004). Therefore it is not surprising that our analysis revealed that DWH projects with DWH brokers might as well start with poor requirements estimations then proceed through several prototyping circles (boundary objects on pragmatic border (Carlile 2004), before they finally get estimated as acceptable:

---

**Figure 1. Dynamics of DWH brokering - DWH-professionals’ perspective**
“After 7 years working for the DWH team in this company, I have still made a mistake in every single report’s definition [in this project]. The problem was the definitions sent by a user supposing to help us ‘translate’ business requests coming from his department. He didn’t add any additional value, but only transferred original requests…Our specification of the requirements was quite bad, which caused a lot of change requests.

Q: Did you go through many iterations?

For sure, one iteration would be great.” (TELCOM01)

However, our analysis also suggests that with DWH and BEDF brokers present, the number of cycles should presumably be lower than without them. In case of NUTCOM01, the interview reported very good estimation of customer’s requirements:

“Meeting reports were sent to all participants, for the confirmation of their content. We wanted to check, if we’ve all well understood their requirements. Few had comments, but the majority didn’t.

Q: Did they ignore you or didn’t they have nothing to add?

I think we’ve defined everything very well. They had nothing more to add.”

If the alignment finally proceeds successfully, DWH-professionals will be able to establish a common lexicon with BEDFs, estimate requirements well and move from quadrant IV to I (Figure 1.b).

In DWH projects starting out in the quadrant III (e.g., ENECOM02), DWH-professionals perceive the business process as unfamiliar and, as in the aforementioned case, they may identify BEDF brokers. Through the communication with BEDFs, DWH-professionals could gradually acquire knowledge about the business and move to quadrants I or IV (Figure 1.b). Although results from our analysis lack transition details, it seems plausible that projects can move from quadrant III to quadrant I either via quadrant IV, if developers gain knowledge of customer’s business but poorly estimate requirements, or transition directly when both processes are in balance.

Even though the case when DWH-professionals perceive the business process as unfamiliar and eventually succeed to estimate the requirements well (quadrant II) appears improbable, our analysis illustrates an example where this was exactly the case. In the second subproject of ENERCOM01, DWH-professionals reported the presence of BEDF brokers, who not only were familiar with data coming from transactional sources, but had apparently participated in their development. The “power BEDF brokers” were able to define requirements by directly referring to the source data, as well as to help define the extraction, transition and loading (ETL) process, hence exhibiting membership to both BEDF and OSP communities:

“We’ve worked with quite an advanced user that had a degree in math, worked at first in the production department, afterwards in the IT department and de facto developed their information system, and now works in the department of strategic planning. He understands both IT and the company’s business process, being able to sketch examples of reports he expected. We pretty much understood it all.”

CONCLUSION

In this paper we observed bordering conditions that help mitigate the communication barriers emerging between communities involved in contemporary DWH projects. Firstly, we argue that the border between OSP and DWH participants is syntactical. However, having a common lexicon does not rule out problems of technical, political or organizational nature occurring and influencing project success. Secondly, we observed brokering on the border between BEDF and DWH participants. We argue that, if DWH and BEDF brokers participate in the initial phase of requirements elicitation, they act as facilitators and should, according to our analysis, help accelerate the process of elicitation. The results were formulated into four concise hypotheses.

Apparently, this study has some limitations with regard to the base data. We conducted only nine interviews so that generalizing the analysis results is difficult. However, as this was an exploratory study, fourteen reported DWH projects of different sizes and types have succeeded to cover seven different branches. This diversity is expected to have a positive effect on the ability to generalize the study. Moreover, the study itself, although illustrative, does not in any way test the taxonomy being developed. In order to alleviate these issues, we plan to conduct additional rounds of interviews to offer a reliable setting from which the research questions could be rigorously pursued and the formulated hypotheses may be verified.

REFERENCES