Classification of Communication Defects in DWH Projects

Alexander Hoffmann  
*Johann Wolfgang Goethe Universitat Frankfurt am Main, alexhfm@wiwi.uni-frankfurt.de*

Wolf Behrmann  
*Johann Wolfgang Goethe Universitat Frankfurt am Main, wbehrmann@web.de*

Follow this and additional works at: [http://aisel.aisnet.org/amcis2008](http://aisel.aisnet.org/amcis2008)
ABSTRACT

Having completed the requirements engineering phase in a software engineering project, we usually refer to requirements as unambiguous, comprehensible and objective. In a software development case study conducted over a period of more than three years we observed several problems within and between the project teams resulting from different understandings of the software specification. Aim of the project was the implementation of a DWH that collects and evaluates significant data of an international banking institute in order to fulfill the requirements specified by Basel II. In order to classify and explain the observed communication defects that occurred during the implementation phase, we consider findings from communication theory and knowledge transfer (on the interaction setting). As a result, we propose a general schema for the classification of communication defects that may make a significant contribution to the handling of problems in communication issues.

INTRODUCTION

During the last twenty years many theses regarding success factors and management of data warehouse (DWH) projects were published. They mainly focus either on a technical or a business view of DWHs and not on the development process as a whole. While the technical papers deal with physical implementation, physical data models, schema design etc., the business oriented publications focus on a logical design and the specification of information requirements. In addition there are few authors which give a holistic view of the DWH development process. Although lots of modeling languages, process models and success factor analyses were published, in practice DWH projects remain critical and many well-known problems still occur.

In a case study we accompanied a DWH project of an international banking group over a period of more than three years. During this time we recognized several problems concerning communication between the participating parties and within teams. Since the observed problems lead to higher project costs and communication related issues are not being considered by DWH literature, we set up a research project in order to analyze communication problems within the project. Our aim was to structure and classify the observed communication defects and to find reasons for their occurrence.

As we will show in this paper, it is possible to combine rigor and relevance by using approved and broad accepted methods of case study research joined with well known theories from other domains like knowledge transfer and communication theory. By doing so it is possible to better understand our real life observations and to identify areas of action to increase performance in DWH projects and software engineering projects in general. Or as (Kilduff, 2006) puts it: „The route to good theory leads not through gaps in the literature but through an engagement with problems in the world that you find personally interesting.‟ By using this approach our conclusion base on a strong theoretical fundament.

This paper is structured as follows: The next section gives an overview of related work of the area of DWH projects. Afterwards, we present our case study and describe our research methodology. The fourth section gives a short introduction into communication theory and presents findings from the areas of knowledge transfer and language theory, which we use in the next section to analyze and to classify communication defects. Finally, we summarize our results and show the need for further research.
LITERATURE REVIEW

In 1968 the term “software engineering” initially appeared (Naur & Randell, 1968), but in spite of all subsequently invented methods to improve software engineering (Berry, 2004; Sommerville, 2001), software development still suffers from efficiency problems. 23% of software development projects are cancelled due to failure and 49% exceed project resources (Standish Group International, 2001), e.g. projects do not meet deadlines or simply get too expensive. Similarly, (Keil, 1995; Keil et al., 2000) state that between 30% and 40% of all software projects exhibit some degree of escalation.

It seems obvious, that a precise inquiry and specification of requirements is necessary for any software development project to be successful. Incomplete, wrong, or changing requirements lead to longer project runtime and higher costs. Thus, a lot of research has been done in the field of requirements engineering (RE). For example, (Kavakli & Loucopoulos, 2005) give a broad overview of current goal modeling approaches in RE. These approaches propose and offer procedures, methods and tools to cope with the process of RE. However, these approaches do not explicitly address the specific problems that occur after the specification is created. In contrast to existing approaches in RE, this work emphasizes a communication-based perspective that focuses on the time after RE, where the development team has to put the requirements into practice.

Since the 1970s a lot of publications have been made in the field of DWH. Most research addresses modeling languages and different process models to create the specification of a DWH project. Furthermore, these approaches refer to the organization of such projects. This implies the assumption that the major challenges of DWH projects lie in the elaboration and formal specification of the data. It is assumed that if a specification is correct in terms of the used formal specification language, the implementation will represent the desired result.

None of the aforementioned approaches explicitly address the personal interaction within project teams. Hitherto, only few publications focus on soft factors in DWH or on software development projects, e.g. in the field of knowledge management (He, 2004), learning processes (Pirinen & Pekkola, 2006) and communication (Gallivan & Keil, 2003). All mentioned research concerning DWH process models does not cope with the specific problems of interpersonal communication within DWH projects after RE phase is finished. This work gives a contribution to the aspect of communication within the field of DWH and software development.

CASE STUDY

Data Warehouse Development in Financial Industry

The following observations have been made during a Basel II project at an international banking group with a balance sheet total of more than 100 billion. The banking group consists of a head office (HO) and 15 major subsidiaries. Goal of the project is the fulfillment of requirements specified by Basel II, which is a list of recommendations on banking laws and regulations issued by the Basel Committee on Banking Supervision. The project started in 2001 and is planned to end in December 2007, so that the project has a running time of 6 years. Currently, the IS solution is being approved by the regular authority.

To fulfill the requirements of Basel II and to enable a group-wide calculation of the required values one main task of the project was to implement a central DWH. It is designed to serve as the data source for the Basel II calculation engine and has to store all relevant raw data required for the calculation of so called Risk Weighted Assets of the banking group. Primarily, data of transactions, collaterals, customers and rating information of both HO and subsidiaries have to be delivered into the central DWH. To achieve this, each subsidiary had to develop extraction jobs for their local databases which create specification conform flat files that are then sent to HO and imported into the DWH.

The specification of the flat files was developed by a core team in HO consisting of five persons, including business and DWH experts to ensure a broad support of the local units. The team also supported and controlled the implementation process of the database extraction jobs. The specification was created using XML Schema Definition (XSD) in order to define the required data in a formal and unambiguous way. At the beginning of the project, an initial training workshop was conducted where the XSD format was explained to representatives of the subsidiaries. Since the specification was assumed to be unambiguous due to the use of XSD, it was sent to each subsidiary via e-mail without further descriptions. Furthermore, HO assumed that the persons in the subsidiaries share the same knowledge about the banking domain, that each of the used terms and descriptions would be understood without any problems.

During the implementation phase, a contact person in HO was nominated for each subsidiary to clarify misunderstandings and technical problems. Support through HO was realized by using several communication channels including e-mail and telephone. Since after the distribution of the specification all support requests of the subsidiaries could be satisfied, HO team assumed that each subsidiary had understood the specification in the intended way. But completeness checks of the flat files
that were delivered by the subsidiaries showed several problems resulting from a different understanding of the specification. These misunderstandings led to delays in the project timetable and unforeseen, additional costs, since on-site-visits had to be arranged about once per month in order to discuss revealed problems, the actual implementation status, and acute problems in face-to-face situations. The different understandings could neither be predicted nor managed via telephone and e-mail. It seemed as if HO and subsidiaries each spoke different languages, since the interpretations of the published specification varied to a considerable degree, although the required data and data structures were formalized unambiguously due to the use of XSD and additionally extensive textual descriptions were used.

Aim of our research was to structure the observed problems and to find reasons for their occurrence. During our research we did not use findings from the area of requirements engineering, since a specification was already at hand and problems appeared not until the publication of the specification. Instead, we classified and explained the observed misunderstandings using findings from communication theory and language theory, since the observed problems occurred due to misunderstandings of the already finalized specification.

Research Methodology

Our research is performed according to the case study method described by (Yin, 2003). This methodology was used because we have examined complex organizational phenomena in their natural setting. The goal of our examination is to understand communication related problems between different project members in a large DWH project. Therefore we have to deal with how and why questions about a contemporary set of events over which we have only little control (Yin, 2003 p 20). By analyzing real life phenomena it is ensured that we deal with relevant questions and by using the broad accepted method of Yin the rigor of our results is guaranteed as well. These criteria led us to choose case study as our preferred research method.

One of the authors is still involved as a consultant in the project described in chapter 3.1. The observations concerning communication defects were documented by him and reconciled with other involved employees of the banking groups. The observations contain descriptions of the occurred problems, the identified causes for the problems (normally communication defects) and details about how and when the problems were discovered as well as involved persons or units. The facts were collected on-site and through interrogations of subsidiary contact persons and HO team members by one of the authors. Furthermore the project documentation (meeting protocols, presentations of lessons learnt sessions, data models, calendars, etc.) was used to confirm the documentations. Additionally unstructured information and non-formal documentation (e.g. e-mail correspondence) was evaluated. The observed facts were then discussed and analyzed by both authors. To ensure an adequate distance and to increase the objectivity of the case, all findings were also discussed in our research team with non-involved people.
THE “COMMON GROUND” IN COMMUNICATION

In our case study we observed communication between HO team and the subsidiaries necessary to fulfill the specified requirements. During this communication, knowledge was transferred from HO to the subsidiaries. Most problems we observed in our case study are based on misunderstandings concerning the concepts that are represented by data structures in the specification. That is, defects in knowledge transfer from HO to subsidiaries occurred. In order to explain the observed communication defects and to be able to put them into a structure that allows us to find strategies for more successful communication in software engineering projects, we will explain the idea of common ground, which is a prerequisite for successful knowledge transfer, and give a short insight into the basics of human communication insomuch as they are necessary to understand the evolution of common ground.

(Clark, 1996) emphasizes that communication acts can be seen as joint activities with two actors, a sender and a receiver, being involved. For successful communication they need to coordinate their activities. Basis for this coordination is shared knowledge between the two actors: They must share a common ground. Any utterances will be formed in relation to the assumptions made by the sender about the common ground. By each of his or her utterances the sender wants to extend the common ground step by step. This extension of the common ground is called grounding. If the contribution to the common ground gets too big or is based on wrong assumptions, misunderstandings occur that might be difficult to detect. Incorrect assumptions about the common ground are often caused by an egocentric bias: “If I know something, I am more likely to expect others to know it too”.

There are three principles that are related to the grounding process. The principle of closure states that agents performing an action require evidence that they have succeeded in performing that action in a way that is sufficient for current purposes. In case of joint actions (e.g., communication acts) the principle of joint closure becomes important: The participants in a joint action try to establish the mutual belief that they have succeeded well enough for current purpose. At last, the principle of least effort states that all things being equal, agents try to minimize their effort in doing what they intend.

Clark’s theory about common ground in communication was applied by (Bromme et al., 2004) to communication scenarios with both sender and receiver having major knowledge differences. This so-called expert-layperson communication is characterized by a low common ground between the actors at the beginning of the communication process. Parts of knowledge in the layperson perspective might be missing and knowledge elements might be embedded in the layperson’s “amateurish” cognitive reference framework that holds the risk of resistance against changes. For example, the expert might receive positive feedback from the layperson, while in fact the layperson did not understand the transferred knowledge in the correct way (Bromme & Jucks, 2001). This may cause an illusion of evidence to the expert. In our case study we observed expert-layperson communication between HO team and the subsidiaries. Both parties share a common ground concerning knowledge related to the banking domain, but only HO team consists of experts regarding the specific requirements of Basel II. Since after the distribution of the software specification the subsidiaries, which are laypersons regarding Basel II requirements, did not make extensive support requests besides minor issues, HO team interpreted subsidiary’s “silence” as prove of a correct understanding of the specification, i.e. HO team suffered from an illusion of evidence.

We now see that a common understanding or common ground respectively is an important success factor of knowledge transfer. But in order to fully understand the communication defects which we have observed in our case study, more in-detailed explanations concerning the process of knowledge transfer are necessary.

For knowledge to be communicable to other persons, the knowledge holder has to represent its mental concepts in a way that they are perceivable for other persons. The result of this representation process is data that can be stored by and shared between persons and machines. Since knowledge is what people have in their minds and data is external perceivable structures, transfer of data is not identical to knowledge transfer. Or, as (Garavelli et al., 2002) put it: “Even when knowledge can be materialized in an object […], the transfer of that object does not necessarily fulfill the knowledge transfer process.” That is because data has to be interpreted by people. Data is inherently meaningless, „It simply exists […] – all waiting to be interpreted, all waiting to have meaning attached – by people.” (Miller, 2002). It is in the moment of interpretation when new knowledge is created in a person’s mind. But one can never be sure that two persons associate the same representation with the same object. „Identical [data] almost invariably provokes (or evokes) different meanings in each of us. […] Although the words we use stand for meaning, we should not assume that those words necessarily provoke the same meaning in others” (Miller, 2002). Therefore, knowledge transfer is not a well defined process.
Problems concerning knowledge transfer occur, if persons speak different languages, whereas language not only refers to natural languages like English, German or Spanish, but also jargons like “mechanical engineer” or “field sales” (Davenport & Prusak, 1998). These persons may have different ways of codifying and communicating their knowledge. A group of people sharing the same representations associated with the same objects is called language community (Kamlah & Lorenzen, 1984). Obviously, for persons to be members of a language community, they must share the same knowledge, i.e., they must have knowledge about the same objects independently of their representations. Therefore, the existence of a common ground is a prerequisite for the existence of a language community.

CLASSIFICATION OF COMMUNICATION DEFECTS

Based on the theory of knowledge transfer, we analyzed and classified communication defects that we observed in our case study. In this section we consider a sender-receiver-situation with the sender (“A”) formulating a software specification (HO team in our case study) and the receiver (“B”) having to understand the specification and transforming the requirements into concrete action or software respectively (the subsidiaries in our case study). The sender codifies a specific requirement, which we call X, using a representation R.

A proposition of the considered situation is that the sender’s aim is an efficient and reasonable communication of X, i.e., the sender pursues a communication strategy that maximizes the probability that the receiver’s understanding of X is correct and minimizes the costs that arise from the codification and transfer of X (see principle of least effort). The communication process is successful, if after the transmission of R both sender and receiver have the same idea of X in mind. We identified six basic conditions that must be fulfilled any communication process to be successful and applied these conditions to the communication defects that occurred in our case study.

<table>
<thead>
<tr>
<th>Condition</th>
<th>General description</th>
<th>Case study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The sender must know at least one representation R of X in order to communicate X.</td>
<td>Due to agreed specification language a fulfillment of condition 1 was always observed.</td>
</tr>
<tr>
<td>2</td>
<td>Both sender and receiver must have knowledge of X independently of any representation.</td>
<td>Some communication defects based on non-fulfillment of condition 2 have been observed.</td>
</tr>
<tr>
<td>3</td>
<td>R must not be manipulated on its way from the sender to the receiver.</td>
<td>By using textual specification and electronic data transfer a fulfillment of condition 3 was always observed.</td>
</tr>
<tr>
<td>4</td>
<td>The receiver must see R as a part of a communication process and must feel addressed by R.</td>
<td>Due to a clearly defined communication process a fulfillment of condition 4 was always observed.</td>
</tr>
<tr>
<td>5</td>
<td>The receiver must know at least one interpretation of R and one of these interpretations must correspond to X.</td>
<td>Some communication defects based on non-fulfillment of condition 5 have been observed.</td>
</tr>
<tr>
<td>6</td>
<td>If the receiver knows more than one interpretations of R, he or she must choose the one that corresponds to X.</td>
<td>Some communication defects based on non-fulfillment of condition 6 have been observed.</td>
</tr>
</tbody>
</table>

Table 1: Conditions for successful communication

Condition 1

At the beginning of a communication process the sender must know a set of representations with each element representing X in a way that is appropriate and communicable to the receiver (condition 1). This set is called \( P_A (X) \). Since it contains predications (P) that a sender (A) knows for a certain object (X). \( P_A (X) \) must contain at least one element or representation respectively. What elements are in \( P_A (X) \) depends on the sender’s capability of expression and on external circumstances, e.g., language guidelines in an organization. In the case of \( \left| P_A (X) \right| > 1 \) the sender knows several representations of X of which he or she must choose one. According to the principle of least effort, the sender will choose the most efficient representation that the receiver is assumed to understand. The representation that the sender chooses from \( P_A (X) \) and communicates to the receiver is called R. In our case study we did not find situations, where the sender was not able to
choose an appropriate representation for X. Therefore, the observed communication defects did not result from non-fulfillment of condition 1.

**Condition 2**

The second condition demands, that the receiver must know X independently of its representations, i.e., the receiver must know the object that is represented by R. If the sender knows for sure or assumes, that the receiver has knowledge about X, R can be used without further definitions of X. Additionally, conditions 3 to 6 have to be fulfilled as well. Alternatively, if the sender is not sure about the receiver’s knowledge about X, it must be explained to ensure successful communication. Explaining or defining an object X means to combine several other representations that the receiver is assumed to understand and to equate these combined representations with a shorter representation R that will be used further on to achieve a more efficient communication (Seiffert, 1996). In our case study, definitions were only conducted if new and project-specific ideas and concepts were introduced, that the receiver was unlikely to know. For example, the specification contained an attribute group named “multi source, multi entity fields” that was represented with the term “MSME fields”. Since the sender, i.e., HO team, did assume that the receiver neither knows the underlying concept nor the used representation, a detailed definition of “MSME fields” was given. Due to the elaborate definition, no communication defects occurred regarding “MSME fields” during the project duration. In another situation, the word “limit”, which was included in the specification, was misunderstood by the subsidiaries. The reason for this misunderstanding was that the subsidiaries did not know the concept that was represented by the word “limit”, which is in a Basel II context an external committed credit line. “Limit”, however, was known by the subsidiaries only for internal risk limitation. Due to an egocentric bias HO team did not define “limit” with explicit focus on the Basel II context. As a consequence subsidiaries used their own interpretation of “limit” without knowledge of the object that the sender actually wanted to represent and developed program code that was not compatible with the specification. Therefore, some communication defects in our case study are based on non-fulfillment of condition 2.

**Condition 3**

According to the classical sender-receiver-model published by Shannon and Weaver, communication comprises the transmission of a signal from sender to receiver over a communication channel (Shannon and Weaver, 1963). During transmission a signal can be manipulated by an external noise source, e.g. background noise during a face-to-face conversation or technical problems in case of e-mail communication. In our case study we did not observe any communication defects resulting from noise that changed the transmitted representations (condition 3).

**Condition 4**

For the communication process to be successful, the receiver must realize the intention of communication and feel addressed if receiving R (condition 4). We assume that this is always the case if commonly known symbols like letters, numbers, or elements of a well known modeling language are used, since these symbols do not exist naturally but are created by men with intend to communicate. But even so, the intended receiver might, for example, see R as „internal notes“ of the sender that are not addressed to him or her. In our case study we did not find any situations where the receiver did not feel addressed or did not see R as part of a communication process. Therefore, the observed communication defects did not result from a non-fulfillment of condition 4.

The receiver must be able to interpret R so that X is among the possible interpretations (condition 5). Furthermore, if more than one interpretation is available, the receiver must choose the correct interpretation (condition 6). Given a fulfillment of condition 2, the sender must make an assumption about the objects that the receiver associates with R. The interpretations of R that the receiver is assumed to know are called $I_{A \rightarrow B}(R)$ (Interpretations).

**Condition 5**

Regarding $I_{A \rightarrow B}(R)$, we must at first consider if the sender assumes X to be in $I_{A \rightarrow B}(R)$ or not (condition 5). In the latter case, i.e., $X \notin I_{A \rightarrow B}(R)$, the sender must choose a more appropriate representation, given that $|P_{A}(X)| > 1$. In our case study, the specification team tried to avoid such terms or phrases that the receiver was likely to misinterpret by choosing more appropriate representations if $|P_{A}(X)| > 1$. Given that $|P_{A}(X)| = 1$ and $X \notin I_{A \rightarrow B}(R)$, the sender is not able to choose a more appropriate representation of X that the receiver is less likely to misinterpret. For example, in our case study the specification contained a data field named “original_principal_amount” that had to store the current principal amount, which is calculated in a different way than the original principal amount. That is, the name of the data field suggested another purpose of use as the specification actually required. Due to technical reasons, HO team (sender) was not able to choose
another representation (i.e., another name) of the data field. Unfortunately, the specification did not contain adequate explanations, so that some subsidiaries (receiver) used this data field in a wrong context, resulting in faulty implementations. As opposed to the example given for condition 2, here the receiver had knowledge about both original and current principle amount so that the misunderstanding was only based on the usage of an inappropriate representation. Therefore, communication defects occurred due to a non-fulfillment of condition 5. In order to increase the chance for successful communication, the sender could additionally have added examples or analogies, since the receiver was likely to interpret R in a not intended way.

Since the sender supposes that the receiver knows X independently of any representation (see condition 2), support for the correct interpretation of R should be given, if \( |I_{A\rightarrow B}(R)| \geq 1 \) is assumed. For example, the sender could give a synonym or a list of synonyms of R. The sender could also list and exclude other possible interpretations in order to draw the receiver’s attention to a possible misunderstanding. Neither of these strategies was observed in our case study.

**Condition 6**

If the sender assumes \( X \in I_{A\rightarrow B}(R) \), i.e., if the sender assumes that both sender and receiver are members of the same language community, depending on \( |I_{A\rightarrow B}(R)| \) different communication strategies can be pursued: Given that \( |I_{A\rightarrow B}(R)| = 1 \), the sender assumes that the receiver interprets R exactly as intended and is therefore likely to use R without any further explanation. If \( |I_{A\rightarrow B}(R)| > 1 \) is assumed, R can be used without further explanation, if the sender trusts in the receiver’s ability to choose the interpretation that corresponds to X among several others (condition 6). This trust might lead to communication defects, if the receiver is uncertain about the correct interpretation and chooses the wrong interpretation of R. In our case study, for example, the specification postulates that each subsidiary has to deliver book values according to a specific data format. But in the financial world different definitions of the term “book value” exist, each having different calculation instructions, e.g. according to the German Handelsgesetzbuch (HGB) or the International Financial Reporting Standards (IFRS). Although HO team (sender) knew about the different definitions in principle, it unconsciously assumed that the subsidiaries (receiver) know all of these definitions and would interpret “book value” in the intended way. In fact, the subsidiaries were uncertain about what interpretation was intended by HO. As a consequence, HO had to answer further inquiries from the subsidiaries, leading to higher communication costs. Therefore, some communication defects in our case study occurred due to a non-fulfillment of condition 6. If the receiver is not assumed to choose the correct interpretation, the sender can list and explicitly exclude other possible interpretations in \( I_{A\rightarrow B}(R) \). This communication strategy is, of course, only possible for interpretations of R that are known to the sender, i.e. that are elements of \( I_A(R) \). For example, in our case study a recapitulatory entity for all banking transactions was introduced and called “facility”. But the term “facility” was usually also used in other contexts, e.g. credit lines. In order to avoid wrong interpretations of the term, other possible interpretations were listed and excluded. Because \( |I_{A\rightarrow B}(R)| \) can be greater than \( |I_A(R)| \), the sender can not be sure to have excluded all other possible interpretations. Therefore, additional interpretation support through the use of examples or analogies should be given.
CONCLUSION

Aim of our research was to classify and give reasons for communication defects that we observed in our case study. We therefore identified six conditions that must be fulfilled for communication processes to be successful. These conditions are based upon findings from the areas of knowledge transfer and communication theory. When applying these conditions to our case study, we found that fulfillment of three out of six conditions for successful communications can easily be ensured by basic communication setup and that most misunderstandings occurred due to inappropriate phrasing in the specification and the sender’s unconscious trust in the receiver’s ability to choose the intended interpretation among several possible. As a consequence, we recommend accepting higher costs at setting up the specification either by stronger interaction or a more detailed specification in order to reduce follow-up costs resulting from misunderstandings. The proposed way of structuring communication defects allowed us to precisely classify the problems we observed in our case study and therefore enables us to develop more efficient communication strategies in software development projects. Further research will include appliance of the classification schema to other domains and the extension of the schema to communication transactions in total.

REFERENCES