JUSTIFICATION OF DATA WAREHOUSING PROJECTS

Reinhard Jung and Robert Winter

University of St.Gallen, Institute of Information Management,, Mueller-Friedberg-Strasse 8, CH-9000 St.Gallen, Switzerland
Phone: +41 71 224 2935, Fax: +41 71 224 2189, {reinhard.jung | robert.winter}@unisg.ch

ABSTRACT.

Project justification is regarded as one of the major methodological deficits in Data Warehousing practice. As reasons for applying inappropriate methods, performing incomplete evaluations, or even entirely omitting justifications, the special nature of Data Warehousing benefits and the large portion of infrastructure-related activities are stated. In this paper, the economic justification of Data Warehousing projects is analyzed, and first results from a large academia-industry collaboration project in the field of non-technical issues of Data Warehousing are presented. As conceptual foundations, the role of the Data Warehouse system in corporate application architectures is analyzed, and the specific properties of Data Warehousing projects are discussed. Based on an applicability analysis of traditional approaches to economic IT project justification, basic steps and responsibilities for the justification of Data Warehousing projects are derived.

1. INTRODUCTION

An empirical analysis [3] of large companies' research needs in the field of Data Warehousing shows that Data Warehouse project justification is regarded as a major issue which requires a considerable research effort. As a consequence, only 53% of the participating companies tried to set up a Data Warehousing business case at all. Of those companies that performed an evaluation of costs and benefits, 58% used multi-attribute utility techniques, 33% used investment techniques, and 17% used various other approaches [3, 15]. Superficial or missing analyses of Data Warehousing projects are usually attributed to the special nature of those projects, e.g. to the 'obvious impossibility' to assess 'intangible benefits' [7]. If estimations are made, the return on investment (ROI) ranges from -1857% to an incredible 16000% with an average of 401% (IDC study cited in [7]).

In analogy to the term database system, the term Data Warehouse system denotes the entire range of applications and databases that is needed to utilize a Data Warehouses for business purposes. Data Warehousing then denotes all activities that are linked to the development, utilization, and operations of the Data Warehouse system.

Several facts contribute to the problems that companies are facing when they try to calculate ROIs for Data Warehousing projects:

- The Data Warehouse system is a complex middleware architecture built up incrementally by several Data Warehouse development projects.
- If a large number of data sources is integrated and a large number of applications is supported by the Data Warehouse, a huge investment is necessary, and lots of internal and external side-effects may be influencing the project(s).
- The Data Warehouse system comprises various components which are utilized by different business units in a different manner while the investment maybe has to be made jointly. While data consuming components (e.g. interfaces to decision support applications or horizontal applications) can be assigned quite easily to 'consuming' business units, no such assignment can be made for infrastructural components (e.g. the core Data Warehouse, interfaces to source applications, meta data management).
- Due to the dynamic nature of many management processes, the Data Warehouse system is subject to frequent changes. From an investment theoretical point of view [5, p.38], diversifica-

tion investments have to be taken into account in addition to initial investments.
- Although less frequently, not only information consuming applications, but also source applications are subject to changes (e.g. migration to standardized software packages).

In this paper, the economic justification of Data Warehousing projects is analyzed, and first, descriptive results from a large academia-industry collaboration project in the field of non-technical issues of Data Warehousing are presented. As conceptual foundations, the role of the Data Warehouse system in corporate application architecture is discussed, and the specific properties of Data Warehousing projects are analyzed in Section 2. In Section 3, the applicability of traditional approaches to economic IT project justification is discussed. Based on that analysis, basic justification elements (i.e. tasks and responsibilities) for Data Warehousing projects are derived in Section 4 and summarized in Section 5.

2. CONCEPTUAL FOUNDATIONS

2.1. The Role of the Data Warehouse System in Corporate Application Architecture

Data Warehouse systems are widely accepted as a new middleware layer between operational applications and decision support applications, thereby decoupling systems focussed on efficient handling of business transactions from systems focussed on efficient support of business decisions. Such a middleware layer is necessary because the direct, individual access of decision support applications to data of operational, transaction oriented applications has proved to be technically or economically infeasible: Data quality problems and complex integration requirements usually make it impossible to supply consistent, integrated data realtime to various decision support applications. Even if technically feasible, the development and maintenance of mn interfaces between m Decision Support applications and n transactional applications cannot be economically useful. As an intermediate systems layer, the Data Warehouse system is decoupling decision support applications and operational applications, thereby reusing integration mechanisms and derived data for various decision support applications and allowing maintenance to be focussed on few, well-defined interfaces.

In [8] a high-level application architecture model is proposed that locates applications along the dimensions 'function', 'product (group)', and 'process'. The dimension 'function' lines up the

various functional areas of the corporation (e.g. order processing, materials management, financials). The dimension 'product (group)' lines up the various divisions or product groups of the corporation (e.g. loans, cash deposits, custody). The dimension 'process' represents the course of business processes (e.g. information request, negotiation, contract, fulfillment/clearing, archiving). Most operational applications comprise modules that cover all functional aspects of a (more or less) complete business process for a specific product, product group, or division [4, p.2-3]. Hence, the application architecture comprises a relatively small number of components that can be designated 'vertical' applications due to their optical appearance in the model.

While the transfer of functions like customer data management or product configuration and pricing from vertical applications into dedicated cross-product applications has started in the 1990ies, channel-specific functions have not been transferred into dedicated channel-specific applications until recently. This is due to the fact that a strong demand for multi-channel (i.e. face-to-face, letter-based, phone-based, and electronic) access to corporate applications is associated with increased mobility and the recent advent of electronic business and widespread access to the Internet. Channel-specific applications integrate access and/or distribution functions which are specific for a certain channel, but may be implemented identically for different products or product groups. If access and/or distribution channels have to be flexibly assigned to products or services, channel-specific functions, product-specific functions, and cross-product functions should be implemented in separate applications.

The clustering of channel-specific functions into applications is determined by the respective access media. As a consequence, vertical applications have to be complemented by alternative, channel-specific application add-ons like Call Center support, WWW portal, WAP portal, Letter Center/Document Management support, ATM support, and traditional, inhouse transaction applications. These applications may differ not only by supporting a different access and/or distribution channel, but also by different security mechanisms. In the application architecture model, channel-specific applications are represented by cubes that comprise various products for selected functions and for a certain portion of the underlying business processes. Due to their optical appearance, channel-specific applications can be designated as 'horizontal' applications.

The positioning of vertical applications, cross-functional applications, horizontal applications, the Data Warehouse system, and finally decision support applications in corporate application architecture is illustrated in Figure 1.

Figure 1. ***The Role of the Data Warehouse System in Corporate Application Architecture***
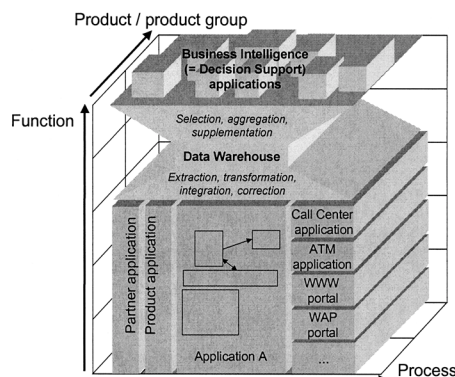
2.2. Data Warehousing Development Projects

In contrast to development projects for mature corporate IT platforms that are often 20 to 30 years old, Data Warehousing development projects are still characterized by an immature, work-in-progress IT platform. One the one hand, this immaturity creates opportunities that do not exist for traditional IT projects, examples being the recent inclusion of Operational Data Stores and horizontal operational applications into corporate information architecture. On the other hand, 'early' development projects using a novel IT platform have to pay a significantly higher portion of infrastructure costs. This phenomenon can be observed for every novel IT platform, e.g. real-time applications in the 70ies/early 80ies or intranets in the early 90ies. Figure 2 illustrates a typical development of a new IT platform: With every additional IT project, usually a lesser amount has to be spent for the development of the IT platform (light gray cost curve) because synergies can be used and significant IT platform development efforts have been paid for by earlier projects. When using a mature IT platform (i.e. IT development project in t_1, left part of Figure 2), the infrastructure costs I_1 usually are very small compared to the actual application development costs A_1. When building a new IT platform (i.e. IT development project in t_2, right part of Figure 2), the ratio of infrastructure costs I_2 and actual application development costs A_2 is more unfavorable.

Figure 2. IT Platform Maturity vs. IT Development Costs
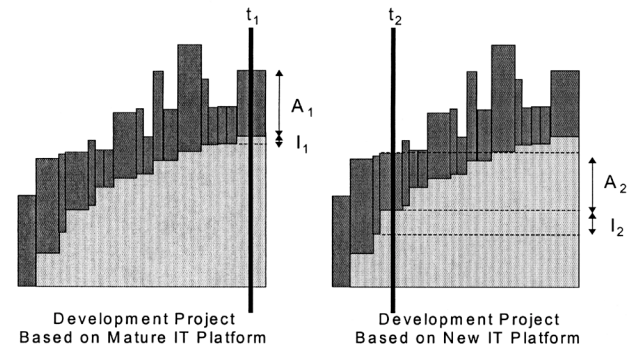
From these considerations follows an important consequence for the economic justification of Data Warehousing projects: High infrastructure costs resulting from the immaturity of the IT platform must be carefully differentiated from application-related development costs. Based on a decision-oriented paradigm, only actual application development costs should be at the expense of business-driven Data Warehousing projects, while infrastructure-related costs should be allocated to corporate projects like 'central IT infrastructure'. If a cost split is not possible, development projects could be credited platform grants to avoid an unfair competition for financial and other resources between Data Warehousing projects and other IT projects.

3. APPLICABILITY OF TRADITIONAL JUSTIFICATION APPROACHES

The justification of IT development projects usually comprises the estimation of expected costs and the estimation of expected benefits. The term 'benefit' is used instead of 'revenue' because IT applications often create results that cannot be measured directly. This is particularly true for Data Warehousing projects because most applications do not directly affect value-creating business processes and, therefore, cannot directly be as-

signed to profit centers. In the following, we analyze traditional approaches to cost and benefit estimation with regard to their applicability for Data Warehousing projects.

Estimation of Costs

Due to the volume and significance of many IT development projects, a large number of methods have been proposed (cf. overviews in [1, 167ff.] and [5, 61ff.]). However, the applicability of traditional cost estimation approaches to Data Warehousing projects is limited because (1) a function or process centric phase model of application development is assumed and (2) specific requirements are regarded as given and elicitable from business users. Moreover, often only initial development costs are analyzed although total costs of ownership could be significantly higher due to the long-term utilization of the Data Warehousing infrastructure [10].

1) In contrast to most traditional IT development projects, we have not observed that Data Warehousing projects follow a certain sequence of distinct phases. Since nearly the entire development process is centered around data / information structures, modularization and development strategies like top-down or bottom-up have much less significance. Instead, activities like architectural planning, meta data management, end-user involvement, and project marketing become key issues for successful projects.

2) Due to the nature of many management processes, objective information requirements of managers are often not explicitly known when Data Warehousing projects start. In contrast to most traditional IT development projects, therefore, requirements have to be elicited incrementally from business users.

Estimation of Benefits

A very comprehensive analysis of benefit estimation for IT projects can be found in Nagel's work [8]. Two fundamental types of estimation approaches are differentiated:

- Single and two dimensional methods use one or two types of input variables (e.g. revenues and costs) to derive an output variable (e.g. profitability). A well-known example of this type of methods is the net present value method by which (input) net revenues at different points in time are aggregated to one (output) net present value using some discounting rate.
- Multi-dimensional methods are used to evaluate choice options regarding a set of quantitative as well as qualitative attributes. A well-known example is the multi-attribute utility analysis that is e.g. used to evaluate creditworthiness or standardized software packages.

While traditional cost estimation approaches in general do not seem to be applicable to Data Warehousing projects, most benefit estimation methods are successfully applied by companies to Data Warehousing projects. In the following section, methodological and organizational aspects of an appropriate justification approach for Data Warehousing projects are discussed, and open issues are identified.

4. METHODOLOGICAL AND ORGANIZATIONAL ASPECTS OF DATA WAREHOUSING PROJECT EVALUATION

Whatever the method is that is used to estimate project costs or project benefits, it is important that those persons and/or organizational units are involved in the project justification that own the management processes or operational processes that have to be supported. The involvement of process owners is not only important because application benefits can be estimated most pre-

cisely by business experts, but also because project commitment is much higher when being involved in early project stages.

In addition, it has proven to be useful to document results from a first, rough-cut requirements analysis together with cost and benefit estimations by means of a business case [2]. Business cases are not only used as a basis for investment decisions, but can also be used for project controlling purposes.

4.1. Estimation of Costs

Regardless of the type of application supported by the Data Warehouse, cost estimations for Data Warehousing projects should by guided by the following rules:

- Like the development project itself, cost estimations should be decomposed into clear modules (‚components' or ‚increments', cf. [6, 15 ff.]).
- If analogies to similar projects in the respective company cannot be used to derive cost estimations, external consultants with large practice databases should be brought in.

The following Data Warehousing project component types can be differentiated:

1) Decision support applications (e.g. a Controlling Data Mart) and horizontal operational applications (e.g. a Customer Relationship Management application)
2) Core Data Warehouse including interfaces to vertical operational applications
3) Meta data management system

Since components of type (1) are linked to management and operational processes, respectively, they create both costs and added-value. In contrast, components of type (2) and (3) have to be considered as infrastructural components which create costs, but no direct added-value. Costs for infrastructural components should either be regarded as (IT) overhead costs or should be apportioned and assigned to components of type (1) corresponding to their respective infrastructure utilization (which can be measured e.g. by total data consumption).

4.1. stimation of Benefits

In his general analysis of IT project benefits, Nagel [8] differentiates 'cost reduction', 'productivity / quality gain', and 'competitive advantage' as basic benefit categories. While cost reductions can be calculated, quantitative effects of productivity and quality gains can only be estimated. Quantitative effects of competitive advantages can neither be calculated nor estimated so that respective decisions have to be based on personal judgement. Benefits of Data Warehousing projects can be structured in a similar way: While direct operational benefits by supporting horizontal operational applications can be calculated (e.g. reduced amount of parallel mailings to the same customer), estimated (e.g. increased response rates to individualized marketing campaigns), or decided (e.g. higher customer intimacy), benefits by supporting decision support applications may only be estimated (e.g. higher customer retention by churn management) or decided (e.g. consistent corporate reporting).

Figure 3. Application Type vs. Evaluation Procedure

	Calcula- tions	Estima- tions	Decisions
Decision Support Applications			
Horizontal Operational Applications			

5. CONCLUSIONS

While their actual applicability depends on the type of supported applications, benefit estimation approaches are applicable to Data Warehousing projects in principle. Large-scale cost estimation approaches proposed for traditional IT projects, however, should not be applied to Data Warehousing projects. Our recommendation is to decompose Data Warehousing projects and use adapted cost estimation methods for estimating project component costs instead.

After all, the organization of Data Warehousing project justification seems to be even more important than the particular methods in use: Table 1 assigns the most important steps of project justification to the dominant organizational roles ‚IT' and ‚business sponsor'. The division of responsibilities between IT and business sponsors has not only led to a more precise benefit estimation by business experts, but has also created a higher commitment by the business side for project critical tasks like requirements analysis, meta data elicitation, and systems utilization.

Table 1.   Organizational Responsibilities for Data Warehouse Project Justification Steps

		Responsibility	
Step		**IT**	**Business Sponsor**
1.	Identification of business sponsor, creation of commitment of business sponsor	✓	
2.	Analysis of decision processes, demonstration of Data Warehousing potentials	✓	
3.	Definition of application scope		✓
4.	Benefit estimation		✓
5.	Cost estimation	✓	
6.	Putting together the Business Case (including results from steps 3 through 5	✓	✓
7.	Aggregation of benefit and cost estimations, final decision		✓

REFERENCES

[1] Baumoel, U., *The Application of Target Costing to Software Development*, Vahlen: Munich 1999 (in German).

[2] Ewen, E.F., Levan-Shultz, K., Medsker, C.E., Smith, J.L., Dusterhoff, L.E., Gottschall, M.A., Data Warehousing in an Integrated Health System; Building the Business Case, in: Proc. *ACM First International Workshop on Data Warehousing and OLAP* (DOLAP'98), 1998.

[3] Helfert, M., *An empirical investigation of data warehousing research issues from the industrie's perspective*, Research Report BE HSG/CC DWS/05, University of St.Gallen, 2000; available at http://datawarehouse.iwi.unisg.ch/arbeitsberichte.htm (in German).

[4] Imhoff, C., The Corporate Information Factory, *DM Review*, December 1999, http://www.datawarehouse.com/dmreview, downloaded 2000-03-29.

[5] Jung, R., *Reverse Engineering of Conceptual Data Schemas*, Deutscher Universitaets Verlag: Wiesbaden 1998 (in German).

[6] Jung, R., Winter, R., Data Warehousing – Application, Reference Architecture and Process Model, in: Jung, R., Winter, R. (eds.): *Data Warehousing Strategie*; Springer, Berlin et al. 2000, pp. 3-21 (in German).

[7] McKnight, W., Data Warehouse Justification and ROI, *DM Review*, November 1999, http://www.datawarehouse.com/dmreview, downloaded 1999-12-08

[8] Nagel, K., *Benefits of Information Processing*, 2nd edition, Oldenbourg: Munich 1990 (in German).

[9] O'Donnell, J.H., Sanders G.L., *Is Data Warehousing Increasing Your Firm's Value?*, *Journal of Data Warehousing*, 5(1), 2000, pp. 24-29.

[10] Thomason, P., *ERP and Data Warehousing: The Cost of Ownership debate*, One Concept White Paper, October 1998; http://www.one-concept.com, downloaded 1999-12-08.

[11] Watson, H., Haley, B., Managerial Considerations, *Communications of the ACM*, 41(9), 1998, pp. 32-37.

[12] Winter, R., *The Current and Future Role of Data Warehousing in Corporate Application Architecture*, Research Report BE HSG/CC DWS, University of St.Gallen, 2000; available at http://www.iwi.unisg.ch/publications.

Related Content

Formal Verification Methods
Osman Hasanand Sofiène Tahar (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 7162-7170).*
www.irma-international.org/chapter/formal-verification-methods/112414

Forecasting Techniques for Energy Optimization in Buildings
Fernando Mateo, Juan José Carrasco, Abderrahim Sellami, Mónica Millán-Giraldo, Manuel Domínguez, Ignacio Díazand Emilio Soria-Olivas (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 967-977).*
www.irma-international.org/chapter/forecasting-techniques-for-energy-optimization-in-buildings/112490

Encouraging Digital Literacy and ICT Competency in the Information Age
Kijpokin Kasemsap (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 2253-2263).*
www.irma-international.org/chapter/encouraging-digital-literacy-and-ict-competency-in-the-information-age/183938

Hybrid Air Route Network Simulation Based on Improved RW-Bucket Algorithm
Lai Xin, Zhao De Cun, Huang Long Yangand Wu D. Ti (2022). *International Journal of Information Technologies and Systems Approach (pp. 1-19).*
www.irma-international.org/article/hybrid-air-route-network-simulation-based-on-improved-rw-bucket-algorithm/304808

An Adaptive CU Split Method for VVC Intra Encoding
Lulu Liuand Jing Yang (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-17).*
www.irma-international.org/article/an-adaptive-cu-split-method-for-vvc-intra-encoding/322433