

Semantic Data Modelling - an Established Software Engineering Method?

In literature, it is common knowledge that semantic data modelling is well established. It is the goal of this paper to review this opinion critically. By means of an empirical study which investigated data administration in Styrian industrial enterprises with more than 1000 employees it will be shown that even in companies of that size semantic data modelling is not commonly used. Only around one third of the enterprises model their data. The main reason for this is the increasing use of packaged software.

Key words: semantic data modelling, software engineering method.

1. Introduction

Information systems can be described at different levels. One distinction between these levels is based on their proximity to information technology. Sheer, for instance, differentiates between three basic descriptive levels: requirements definition, design specification, and implementation description.¹ The requirements definition describes the business application to be supported in a formalized language. This description is finally transferred to concrete hardware and software components (implementation description). A similar division is made by Heinrich. This author distinguishes between logical and physical models,² contrary to physical models logical ones are free from any technical implementations.

Semantic data modelling can be simply defined as the design of a "logical"

¹ SCHEER A.W.: Business Process Engineering: Reference Model for Industrial enterprises, pp.13

² HEINRICH L., BURGHOLZER P.: Systemplanung 1, p.23.

model for the data view of an information system. Such models attempt to reflect more structure of the real world, they endeavor to incorporate more meaning.³

In the past, many methods for semantic data modelling were developed. A good overview is given by Ortner⁴, Vinek, Rennert, and Tjoa⁵, as well as by Hars et al⁶. In practice, the entity-relationship method⁷, respectively derivations from it, is most used. Entity-relationship diagrams are supported by many CASE tools.

2. Existing scientific findings

In the academic community, data modelling has been discussed extensively. Well known authors take the view that semantic data modelling is a well established software design method:

According to Oesterle, it is sure that data modelling is used in larger enterprises⁸. Denert points out that data modelling by means of the entity-relationship method has spread out world-wide and won recognition as a standard⁹. Sinz argues that data modelling and especially enterprise-wide data modelling was widely accepted in enterprises during the last years¹⁰.

The only empirical study on this topic gives a similar impression even though its results are less euphoric. This study was performed in Austria, Germany, Switzerland, and German-speaking Italy between March and August 1992¹¹. More than 3000 questionnaires had been sent out to organizations in different lines of business, 200 of which were returned. This reflects a response rate of less than 10%. The size of the responding corporations was quite heterogeneous: 35% had less than 1000, 27% between 1000 and 3000, and 38% more than 3000 employees.

³ DATE C.J. An Introduction to Database Systems, pp.226.

⁴ ORTNER E.: Semantische Datenmodellierung - Datenbankentwurf auf der Ebene der Benutzer.

⁵ VINEK G., RENNERT P.F., TJOA A.M.: Datenmodellierung: Theorie und Praxis des Datenbankentwurf, p.188-208, pp.226-245.

⁶ HARS A. et al.: Concepts of Current Data Modeling Methodologies; A.Survey.

⁷ CHEN P.: The Entity-Relationship Model - Toward a Unified View of Data used. Entity-relationship diagrams are supported by many CASE tools.

⁸ R&O: Datenmodellierung in der Praxis, p.97.

⁹ R&O: Datenmodellierung in der Praxis, p.97.

¹⁰ SINZ E.J.: Editorial zum Schwerpunktthema Untemehmensweite Datenmodellierung, p.267.

¹¹ R&O: Datenmodellierung in der Praxis

The main result of the R&O survey is that half of all corporations use semantic data modelling, 14% have introduced it theoretically, and 19% have planned its introduction. Only for 12% of all enterprises data modelling is not relevant.

3. Semantic data modelling in Styrian industrial enterprises

In May, June and July 1994, an exploratory study was performed dealing with data administration in Styrian industrial enterprises with more than 1000 employees¹². All eleven contacted enterprises (with an average of 2000 employees) were prepared to collaborate in the study. Therefore, the results of the investigated population are representative (response rate: 100%). Data were collected by means of face-to-face interviews.

In addition to data modelling, the study included further fields of data administration like database administration, data use, and organizational aspects. In the following, only the results concerning data modelling will be presented.

The most surprising result of the survey was that there were only four corporations in which semantic data modelling was practised and only one corporation which planned its introduction. On the other hand, semantic data modelling is not relevant to six enterprises.

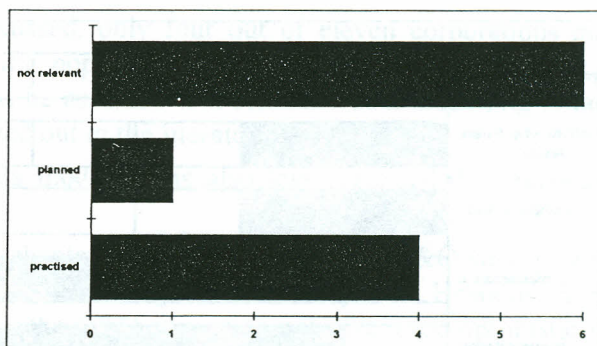


Fig. 1: Adoption of semantic data modelling

This result is much worse than the one presented by the R&O survey. For this reason, the viewpoints explained above must be questioned at least for the underlying population.

¹² SCHLOEGL, Ch.: Datenmanagement auf dem Prüfstand.

The corporations which do not practise data modelling gave the following reasons: use of packaged software, experienced employees, and outsourced computing center.

The corporations which model their data estimate the expenditure at eleven days per month. If the enterprise which works on an enterprise-wide data model is not considered, the mean is only four days.

Referring to the acceptance of semantic data modelling, the results are heterogeneous. Data modelling is fully accepted in one, mostly accepted in two, and nearly accepted in one enterprise. The R&O study gives a similar impression¹³. It seems that semantic data modelling is, if practised, accepted for the most part.

All corporations which model their data use a certain method. In nearly all cases, this is the entity-relationship method. A similar result is reported by the R&O survey according to which this method is used by 75% of the enterprises. Especially for the development of extensive data models, it is necessary to use tools. These can be CASE tools and data dictionaries. They are applied by three at the most five enterprises. It seems that even in enterprises of this size CASE tools and data dictionaries are far from being wide-spread. The corresponding figures of the R&O survey are worse: Only 40% of all corporations which use semantic data modelling have a data dictionary, only 50% a CASE tool¹⁴.

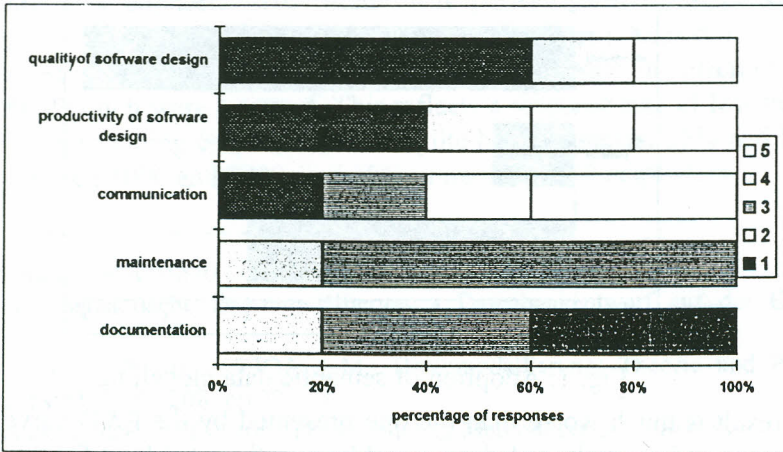


Fig. 2. Ranking of benefits resulting from data modelling

¹³ R&O: Datenmodellierung in der Praxis, p. 66

¹⁴ R&O: Datenmodellierung in der Praxis, p. 63.

The rating of the benefits resulting from semantic data modelling was quite interesting

Three interviewees (60%) estimated the highest benefit in "increase of the quality of software design". "Higher productivity of software design" also got a better ranking. "Improved communication in the company", "less maintenance", and "better documentation" has less importance. A similar outcome is reported by the R&O survey. However, "improved communication" was ranked second followed by "higher productivity of software design".¹⁵

Those corporations which practise data modelling (or plan to introduce it) think that it is economically efficient. Three corporations estimate the increase of productivity of software design to more than 25%, one corporation values it between 11% and 25% and the following between 1% and 10%.

4. Conclusions

The following conclusions can be drawn from the results presented above:

1. Semantic data modelling is much less applied than it is argued in the scientific circles.

As already discussed, only four out of eleven corporations model their data. Because of its specific population, the results of this study cannot be generalized. Nevertheless, it must be assumed that semantic data modelling is far from being as wide-spread as pointed out in the literature.

2. Semantic data modelling is also less practiced than reported by the R&O survey.

In the previous chapter, it was shown that the R&O survey reports on a much higher adoption of data modelling than this study here. However, it must be doubted whether the results of the R&O survey do not give a too optimistic view about data modelling in practice. It is highly probable that the questionnaires were only returned by enterprises which have introduced data modelling. At a response rate of less than 10%, then representativeness of the R&O survey must be questioned¹⁶.

¹⁵ R&O: Datenmodellierung in der Praxis, p. 72.

¹⁶ EICHNER K., HABERMEHL W.: Ergebnis einer empirischen Untersuchung zur Repräsentativität postalischer Befragungen, p. 125.

A possible sign is the survey performed by Hildebrand in 1991¹⁷. In this survey, electronic data processing (EDP) in German enterprises of different branches was investigated. Some questions also dealt with data modelling. The results concerning these questions are quite similar to those in Styrian enterprises. 32% of the responding enterprises use the entity-relationship method, 25% a CASE tool, and 49% a data dictionary. The corresponding figures in Styrian enterprises are 36%, 27% at the most 46%. Because of the general topic (EDP), it is unlikely that there was a correlation between collaboration in the study and use of data modelling.

3. The main reason for the low adoption of semantic data modelling is due to the increasing use of packaged software.

Four out of six enterprises which neither practise semantic data modelling nor plan its introduction in future explained this explicitly with the use of packaged software. In these six corporations, an average of 65% of the fields of activity are covered by packaged software. In the remaining five firms this share is only 31%. In future, packaged software will gain further importance. Seven companies want to raise its share, only one firm plans to reduce it.

4. Packaged software gets more important compared to the design of custom software (together with data modelling) because the advantages of the use of packaged software seem higher than the related disadvantages.

Ten out of eleven interviewees said that the advantages of packaged software use exceeded the related disadvantages. Removal of development risk, cheaper and faster software procurement, and in packaged software included know-how, might be the main reasons. In comparison with semantic data modelling, the effort to maintain all this is not only reduced, it is even totally transferred to the software producer.

¹⁷ HILDEBRAND K.: Informationsmanagement - Status quo und Perspektiven; Ergebnisse einer empirischen Untersuchung.

References

- [1] CHEN P.: The Entity-Relationship Model - Toward a Unified View of Data. In: ACM TODS, Vol. 1, 1976, Iss. 1, pp. 9-36.
- [2] DATE C.J.: An Introduction to Database Systems. Volume II. Addison-Wesley: Reading et al 1985.
- [3] EICHNER K., HABERMEHL W.: Ergebnis einer empirischen Untersuchung zur Repraesentativitaet postalischer Befragungen. In: Koelner Zeitschrift fuer Soziologie und Sozialpsychologie, Vol. 34, 1982, pp. 117-126.
- [4] HARS A. et al.: Concepts of Current Data Modellin Methodologies: A Survey, Universitaet des saarlandes, Institut fuer Wirtschaftsinformatik: Saarbruecken 1991.
- [5] HEINRICH L., BURGHOLZER P.: Systemplanung I: Die Planung von Informations- und Kommunikationssystemen. Vol. 1, 3rd Edition. Oldenbourg; Muenchen 1987.
- [6] HILDEBRAND K.: Informationsmanagement - Status quo und Perspektiven, In: OeZS, Vol.12, 1987, Iss.1, pp.80-84.
- [7] ORTNER E.: Semantische Datenmodellierung-Datenbankentwurf auf der Ebene der Benutzer, In: Informatik-Spektrum, Vol.8, 1985, Iss. 1, pp.20-28.
- [8] R&O: Datenmodellierung in der Praxis: Eine Marktanalyse ueber die Anwendung einer Methodik, R&O Software-Technik GmbH: Planegger Strasse 16-18, Germering, Deutschland 1992.
- [9] SCHER A.-W.: Business Process Engineering: Reference Models for Industrial Enterprises. 2nd Edition. Springer: Heidelberg et al 1994.
- [10] SCHLOEGL Ch.: Datenmanagement auf dem Pruefstand: Am Beispiel der steirischen Grossindustrie. Dissertation. Universitaet Graz, Institut fuer Informationswissenschaft: Graz 1995.
- [11] SINZ E.J.: Editorial zum Schwerpunktthema Untermehmensweite Datenmodellierung: Probleme und Loesungsansaezte. In: Wirtschaftsinformatik, Vol. 33, 1991, Iss. 4, pp.267-268.
- [12] VINEK G., RENNERT P. F., TJOA A. M.: Datenmodellierung: Theorie und Praxis des Datenbankentwurfs. Physica: Vuerzburg 1982.

Received: 1995-09-16

Schlögl, C. Semantičko modeliranje podataka - afirmirana metoda programskog inženjerstva?

Sažetak

Opće je poznata stvar u literaturi da je semantičko modeliranje podataka dobro afirmirano. Cilj je ovog rada dati jedan kritički pregled takvog mišljenja.

Empirijskom studijom, koja se bavila urpavljanjem podacima u indutrijskim poduzećima Styrian, koja imaju više od tisuću zaposlenih, pokazat ćemo da se čak i u poduzećima te veličine semantičko modeliranje podataka ne koristi često. Samo oko jedne trećine tih poduzeća modelira svoje podatke. Glavni razlog je povećano korištenje programskih paketa.