

Internal Data Market Services: An Ontology-Based Architecture and Its Evaluation

Fons Wijnhoven
*University of Twente, Enschede,
Netherlands*

a.b.j.m.wijnhoven@sms.utwente.nl

Edwin van den Belt
*Ordina Finance,
Netherlands*

Edwin.van.den.belt@ordina.nl

Eddy Verbruggen
*Ordina Finance,
Netherlands*

eddy.verbruggen@ordina.nl

Paul van der Vet
*University of Twente, Enschede,
Netherlands*

p.e.vandervet@cs.utwente.nl

Abstract

On information markets, many suppliers and buyers of information goods exchange values. Some of these goods are data, whose value is created in buyer interactions with data sources. These interactions are enabled by data market services (DMS). DMS give access to one or several data sources. The major problems with the creation of information value in these contexts are (1) the quality of information retrievals and related queries, and (2) the complexity of matching information needs and supplies when different semantics are used by source systems and information buyers. This study reports about a prototype DMS (called CIRBA), which employs an ontology-based information retrieval system to solve semantic problems for a DMS. The DMS quality is tested in an experiment to assess its quality from a user perspective against a traditional data warehouse (with SQL) solution. The CIRBA solution gave substantially higher user satisfaction than the data warehouse alternative.

Introduction

In modern organizations the amount of data stored and the number of data stores are substantial. The potential value of data stored can only be achieved when the data are meaningfully shared among people and IT applications. This is highly problematic, because standardization of data definitions and meaning is complex and often infeasible (Davenport & Prusak, 1997). The idea of creating a corporate database (with one shared data model) as a sound foundation for the integration ambitions of systems often fails, because of the semantic complexities involved (Davenport, 1998; Stamper, 1988), the high maintenance costs resulting from complex database schemas (Mitra et al, 2000), and the limitations of database technologies with respect to scope and size (cf. Broadbent et al, 1996). A possible "solution" for data exchange without standardization, using intersystem interfaces, also is not very successful. On the longer term, the creation of many of these interfaces leads to an explosion of complexity: who will manage and understand the interfaces (cf. Laudon & Laudon, 2001)? Consequently, data stores easily are transferred in huge data sources without successful reuse (Markus, 2001).

Material published as part of these proceedings, either on-line or in print, is copyrighted by Informing Science. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission from the publisher at Publisher@InformingScience.org

The problems mentioned above can be approached as an information market problem (Wijnhoven, 2001). In information markets, one or many actors may want to supply data (defined here as computer-based carriers of information) for one or many potential information buyers (who are able to transform the data to information). Information is defined here semantically as meaningful representations of reality, objects and ideas (Stamper, 1973). When there is only one information buyer, the needs of this actor may be clearly defined and the information suppliers know exactly what to deliver. This happens in so-called hierarchic, oligopsonic or monopsonic information markets (Wijnhoven, 2001). When there are many information buyers, it may be more difficult for suppliers to customize their offers. As a consequence, the suppliers may want to supply a general data source or several data sources (if there are more suppliers), and give the buyer tools to create his own information goods via information retrieval facilities. When there is only one information source - a monopolist information market-, the buyer may focus on understanding the semantics of the source to optimize information goods value. Often, though, an information buyer has a complex information need, and one supplier may not be sufficient to serve the buyer. Additionally, suppliers may have many potential buyers, and thus cannot optimize the resource to the needs of one buyer, implying that choices have to be made in favor of preferred buyers, and that the others just have to understand that they cannot be fully served. The consequence is the need of market segmentation and price differentiation over these segments (Kotler, 2000; Shapiro & Varian, 1999). In these n:m information markets, services may come into existence that can deliver added value as intermediate between the initial suppliers and the final consumers. These added values consist of market aggregating functions (Timmers, 1998) for combining access to different sources and providing buyers an interface to query these sources. Such a data market service (DMS) is generally called an information refinery (Meyer & Zack, 1996). In an information refinery, the information suppliers deliver data to the acquisition activity of the service. Next these data are analyzed, integrated, synthesized, standardized, and consequently stored in one or several source systems and made available to clients to facilitate customized deliveries. At least three different levels of fulfillment customization (Kotler, 2000) can be conceived here: (1) data deliveries in fixed bundles for the mass market, (2) customized delivery on basis of processing client-profiles, and (3) client self-service by giving the client access tools (especially information retrieval tools) to let the client pick the specific data from selected data sources. The distribution process is similar to the retail process and thus also includes the contracting and product presentation. Figure 1 gives a process model of the information refinery and the internal data market activities.

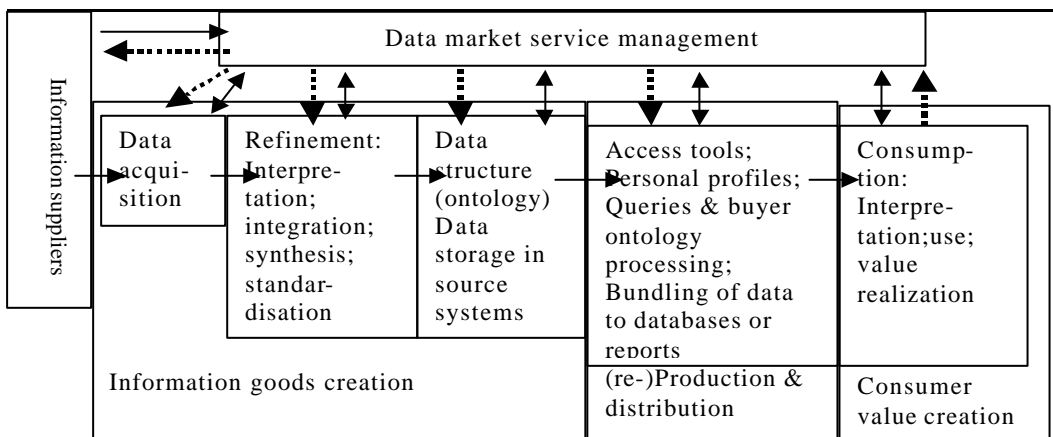


Figure 1: Data market service processes.
 Note: Arrows present information flows. Dashed arrows represent resource transfers

Although information refineries have been conceptually well thought-out, the publications in this field do not clarify how demand and supply of source systems and information buyers meet each other. This

market-matching problem in information markets has two dimensions: (1) the determination of prices, and (2) the creation of joint meaning and values. This article is devoted to the second problem. For the first problem, price mechanisms, the reader is referred to other work (e.g. Picard, 2000; Shapiro & Varian, 1999).

Outline of this Article

This article analyzes the problems that rise when different source systems and different information buyers have different ways of conceptualizing the world, and thus may give different meanings to the same information goods. This is a main problem for information market services, because it equals the problem of misunderstanding between a client and a supplier and thus results in low levels of client satisfaction. This problem is further analyzed in section 3.

Section 4, gives our (ontology-based) solution of matching differences in conceptualizations between data sources and DMS clients. This theoretical exploration is made concrete in a prototype system that is used in an experimental setting to assess the resulting client (user) satisfaction. In this experiment the CIRBA prototype --developed by IT-consulting firm Ordina (www.ordina.nl) in collaboration with the first and last author-- is compared with a data warehouse system that is not based on the ontological principles we developed. The precise description of the experiment is given in section 5. Section 5 also presents the analysis of the data gained.

Finally we evaluate the results, and reflect over the nature of internal data market services.

Ontologies and Ontological Misfits

An ontology is defined here as a specification of a conceptualization (Gruber, 1995). Ontological misfits easily happen in communications because people and information systems may use different conceptualizations of the 'same' phenomena, use different terms for the same concepts (synonyms), use the same terms for different concepts (homonyms), or they use different units, measures and formats, or apply different deduction or reasoning rules (attributes, (sub)classes and semantic implications) (cf. Gruber, 1995; Mitra et al, 200; Shaw & Gaines, 1989). Theoretically these differences may be solved, but practice shows that it is difficult always to be explicit and aware of one's own conceptualizations, recognizing the exact use of others may be even more difficult, and coming to a standardization of all is often attempted but always more difficult as can be envisaged from the outset. Many attempts to standardize ontologies consist of lexicons, taxonomies, database schemes, data libraries, object libraries, symbol tables and domain object models. Standardization of ontologies may be realized by hierarchy or monopoly, so that a buyer or a supplier can dictate a standard. This may be the case with some information services, but is certainly not true for all, and real problems happen when different source systems, owned or managed by different 'dictators' have to be combined. If there are only two source systems that have to be combined, a translation may be developed that can express the ontology of one source to that of the other and vice versa. The problem though quickly becomes worse when the number of source systems increases (if N is the number of source systems, the number of translations that have to be developed and maintained is $N(N-1)/2$; but the problem explodes in complexity given the number of concepts/terms and relations for each system). The problem even comes worse when we know that source systems and users do not always have the same ontologies, and users may each have their own specific ontologies.

The foundation of an ontology consists of a collection of well-defined concepts and meaningful relations among these concepts. Ontologies are specifications of the concepts people use in their communications. People also apply ontologies in the definitions of and communication with their databases. The ontological approach recognizes that users and sources may use different ontologies, and also that different users and different systems may have different ontologies (cf. Gruber, 1995). A source ontology ex-

presses the metadata (data definitions and data models) of a database or information system. People may share an ontology in a certain field, the domain ontology, and still may use their own personal ontology (user ontology). Obviously domain and user ontologies will have connections, and the one may be (partially) expressed in the other. Very often users and domains are not very well aware of their ontologies (cf. Nonaka, 1994; Schreiber et al, 2000), and source ontologies are often poorly documented because of the general lack of interest in project documentation (cf. Laudon & Laudon, 2001). The awareness of these ontologies helps in translating personal information needs to the abilities of source systems to supply what is needed. This requires some semantic integration, though not necessarily standardization. CIRBA provides features to translate (via mediators; König-Ries, 2000) parts of a domain ontology into the ontologies of relevant source systems. This presupposes explicit identification of semantic implications (like homonyms and synonyms), classes and subclasses, attributes of classes, and instances in the ontology (Mitra et al, 2000). A simple example of ontologies for an insurance firm is explained below.

An insurance firm has a number of products (like fire, life and health insurances), which are retailed by independent insurance agents. These agents sell the products to their clients, receive the clients' payments (called premiums), and pay the clients' claims. For these services, the agents charge the insurance firm. The insurance firm has a system to monitor the performance of products and agents. One of the domain expertises of the insurance firm consists of the account management domain, which support the agents to increase the product performance and give the agents feedback on their work. They also search for new agents, which they perceive as their clients. The account managers split the client group in segments per region, size and specialization. The other domain is the product management domain, whose interest is mainly on the behavior of the products in the market. They emphasize to split the group of products in life, health, fire, social security, and car insurances. The domain ontologies of both groups thus are rather different, and each of them has a different source system. The domain ontologies of both and the source system ontologies are given in figure 2.

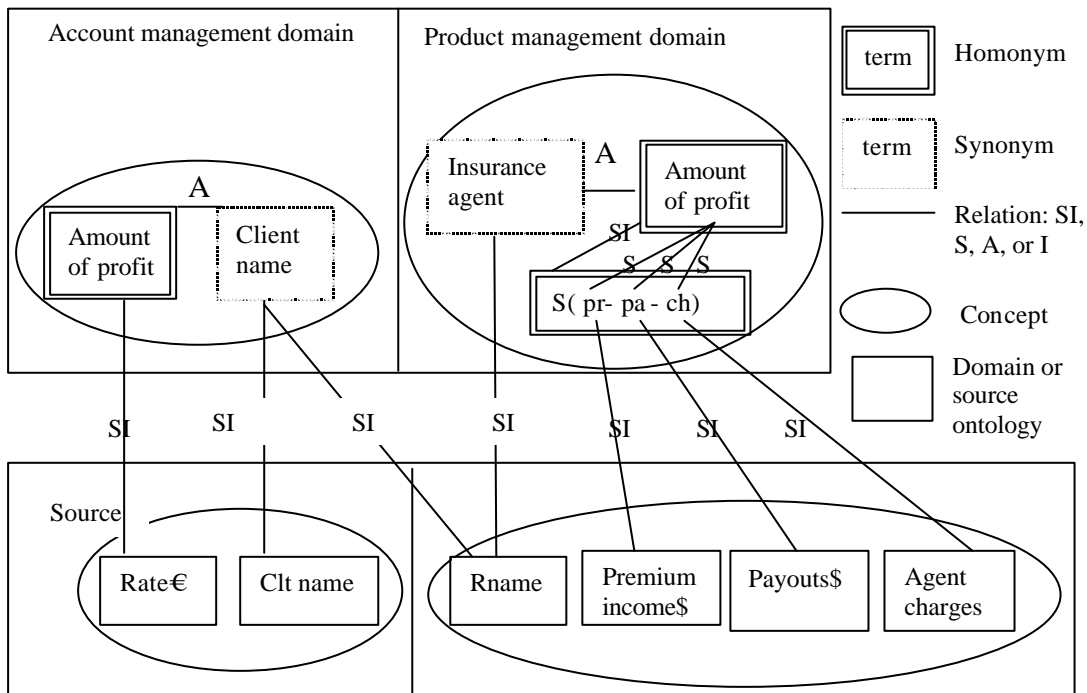


Figure 2: Illustration of an ontology for the profit concept of an insurance firm.

Note: for sake of simplicity we have not presented the user ontology (which will be discussed later), only split out the “product management domain”, though we did not subclass the product concept.

The problem we want to focus on is internal, Intranet, information markets. Internal semantic standardization is not less difficult than external standardization (Boisot, 1998). An internal information market service probably is less able to differentiate with respect to market group, and thus has less opportunities of choosing its clients. The opportunity of a client group choice is important, because this may focus the service to a group with an equivalent ontology, and to a reduced number of suppliers without risking too little turnover (Shapiro & Varian, 1999). An internal market service has to serve all groups and integrate all relevant and available data sources. To cope with this diversity, we have to define an architecture for combining the sources and making them retrievable. The data warehouse (DWH) comes in front as a possible solution, because DWHs, and its datamarts, are particularly developed to select data from different sources in one database for a particular set of problems or a particular group of users (cf. Bischoff & Alexander, 1997; Inmon, 1996). This solution requires many insights in user group information needs, which is feasible for more static and standardized information needs. During information buying, the user's needs and the source's supply must match. This involves that the user's needs have to be expressed in a query, which is able to use the ontology of the source system(s), and is able to combine and present the results. In the DWH approach, this problem is solved by the development of corporate data models (CDM), which is basically a supply-focused (push) solution. The CDM is able to process queries, when these queries are correct according to the CDM's ontology (definitions of data). We state that this will only partially solve the information market problems, because the buyer may not be able to use this ontology fully, and the sources will not always be able to detect ontology misfit, resulting in underutilization of resources.

An alternative, and more flexible approach --from the perspective of information buyers-- is the use of intermediators, which link the user's interface with source systems. This approach desintermediates the need of suppliers to structure information markets, and reintermediates an ontology-based query system that is able to translate a user's ontology to a source system's ontology, and help users to evaluate and improve the demand and supply relation. This should lead to a generic insight in internal DMS and an improved user satisfaction versus the traditional DWH approach. The related research questions are:

1. What architecture is able to match user and supplier ontology in a multi-user, multi-source system internal information market?
2. Does this architecture result in higher user satisfaction compared to a data warehouse approach?
3. What generic insights does this architecture deliver for the design of internal data markets services?

Ontologies Matching

A Review of Solutions for Ontologies Matching

Organizations mostly have several data source systems. These source systems may consist of legacy systems, operational data stores, DWH, and data marts. Legacy systems have a database developed for a single application or program. This makes it difficult to combine data of different systems for answering questions. Therefore operational data stores (ODS) are created to organize data around a specific topic (cf. Inmon, 1996). This requires the transformation and integration of data from these different systems. A more ambitious integration of data is the data warehouse, which is: "...a collection of integrated, subject-oriented databases designed to support the decision support function, where each unit of data is relevant to some moment in time" (Inmon, 1996). The DWH is useful as a means of integrating data from different internal or external sources. The integration results in a consistently defined data collection, and is useful for grounding broader scope tactical and strategic decisions. The DWH is able to respond much quicker to specific information needs than the ODS or legacy system approach. The problem though is that the DWH requires shared and standardized data definitions, which are mostly infeasible.

ble in an organization. Practitioners have mentioned that the maintenance of the DWH is a huge problem as well (cf. Inmon, 1996). Consequently, the practice shows more moderate ambitions, called data marts, which contain more condensed and more synthesized and standardized data for a smaller user group. Because the data of data marts are less standardized between groups, it is more difficult to meaningfully exchange their data among groups.

Mediators are developed to integrate data from diverse primary (legacy or ODS) sources, and as such replace DWH and data marts (cf. Austin et al, 1996; König-Ries, 2000). From the supplier perspective, mediators access wrappers, which are able to make the functionality of the source systems available to mediators. At the information buyer perspective mediators may link to intelligent querying facilities, which are able to understand the ontologies of the source systems and are able to match the user's ontology with the source systems ontologies. As such, the mediators are the link between the source system wrappers and the user interfaces, but presuppose an ontology explication to make relevant translations between the ontologies of the users, domains and source systems. Table 1 gives an overview of the advantages and disadvantages of the mentioned ODS, DWH and data marts, mediator architecture, and the mediator architecture with a smart querying --ontology-based-- architecture.

Solution	Advantages	Disadvantages
Operational data stores	<ul style="list-style-type: none"> ▪ Useful for short term, real-time decisions. ▪ Technically easy to implement. ▪ Reduces workload on operational databases 	<ul style="list-style-type: none"> ▪ Disintegrated and temporal nature of data disables the use for longer-term decisions. ▪ Source data need transformation, integration and cleaning up. ▪ Low flexibility in querying
Data warehouse & data marts	<ul style="list-style-type: none"> ▪ Useful for middle and longer term decisions. ▪ Relatively easy to implement. ▪ Reduces the load on operational databases considerably ▪ Useful for storing historic data and data analysis (e.g. data mining) 	<ul style="list-style-type: none"> ▪ Not useful for short-term (real time) decisions; because of time lag and high aggregation level of data. ▪ Needs cleaning, transforming and integration of data. ▪ Needs corporate data model, which constrains the user's language. ▪ Data marts have less scope and reintroduces data inconsistencies among databases. ▪ Different users may have different objectives with the data warehouse; these objectives may be difficult to combine. ▪ Connecting new source systems is complicated.
Mediator	<ul style="list-style-type: none"> ▪ Useful for short, middle and longer term decisions. ▪ No need for transforming source systems data. ▪ Is able to create real-time and ad hoc data selections from heterogeneous sources. Connects different domains. ▪ The distributed architecture is a foundation for incremental development and scalability. 	<ul style="list-style-type: none"> ▪ Technically rather complex to realize because of complex architecture. ▪ Executing queries on operational databases leads to considerable workload on these systems, and systems performance problems.
Mediators with ontology	<ul style="list-style-type: none"> ▪ Same advantages as without an ontology plus... ▪ The user is not constrained by a given language, and is able to specify requests in his own language. ▪ Each group of data buyers may be served by their own language and worldview. New sources and users are easily connected via the ontology. 	<ul style="list-style-type: none"> ▪ Same problems appear as with mediators without ontology, but.... ▪ Additionally, this option requires the creation and maintenance of ontologies for source systems.

Table 1: Advantages and disadvantages of different ontology matching solutions

Matching Supply and Demand in the CIRBA Prototype

To realize the match of the information needs and information supply, we thus propose that the user ontology must be translated in terms of the source ontology, and that several intermediating services help to actually realize the match. These service activities are the following:

1. It should be possible to store the ontologies in a repository.
2. The repository should be retrievable to select the proper ontologies. This implies that if a specific user submits a query, the system should be able to know what ontology will be used.
3. The system should be able to integrate the data that are relevant from different sources. Homonyms, synonyms and other data relations (subclasses, attributes, instances, semantic implications) applicable to the different source systems should be identified and used for translation to an answer for the user.
4. The system should be able to handle a dialogue with the user, so as to specify the question and to communicate to the user if an answer can be given.
5. The user ontology should be transferable to a structured query language.
6. The data used for an answer, should be presented in the user's own ontology (including the formats and units applied).
7. It should be possible to maintain ontologies, because people and domains will learn or develop their ontologies, and new sources may be connected (thus need their ontology description).
8. Wrappers make source systems data accessible.
9. Mediators divide the user query to sub activities with respect to each separate source system by using the ontologies.

The related CIRBA architecture is given in figure 3.

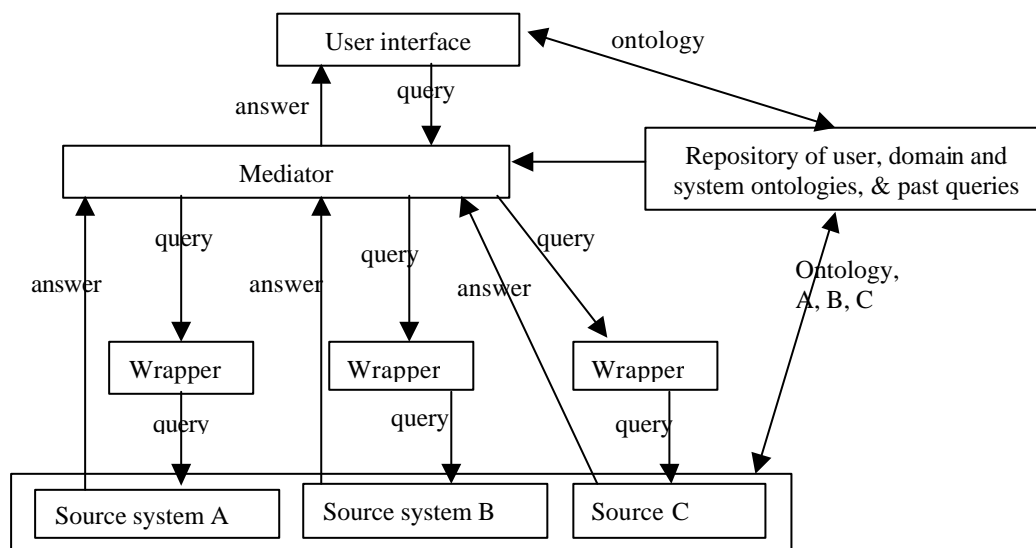


Figure 3: A simplified model of the CIRBA architecture.

Following figure 3, a user A at an insurance firm may for instance ask how much agent Y earns with product Z. In composing the query, the mediator consults the repository. If the query is not new for user A, the query can be retrieved and the answer can be quickly produced. If the query is new, the mediator will check the repositories, and may find that user A means by “earn” the same concept as “amount of

profit”. Because user A may be a product manager, the ontology of “amount of profit” for the account managers is retrieved (including subclasses of products), and the same may respond by delivering a standard report, or asking the user to stipulate for which products he want the answer. The user may respond by selecting two products, e.g. fire and life insurances. The query may activate the wrappers of systems A and C, which hold profit data created by the product management department for the selected products. The mediator aggregates the data of the two systems to one table, adds a calculation of profit applying the formula listed in the (ontologies) repository, and transfers the data terms to the set used by the user.

Testing Client Satisfaction

Variables and Experimental Treatment

An ultimate goal of the CIRBA architecture is that it should improve the user satisfaction with existing databases and information retrievals. More precisely it was suggested that the DMS should result in higher user satisfaction than the traditional DWH solution. This section presents empirical evidence for this proposition.

The literature on information systems user satisfaction distinguishes user satisfaction and end-user satisfaction (cf. Doll et al, 1994). User satisfaction includes many aspects (i.e. the satisfaction with user support and infrastructure reliability) at which CIRBA’s ontology system does not aim. We therefore want to specify the argument to end-user satisfaction, which equates with the information buyers’ evaluation of the information goods received. Doll et al (1994) have developed measurements for end-user satisfaction and concluded that this construct has five dimensions: content, accuracy, format, ease of use, and timeliness. Because we want to evaluate the value of CIRBA in an experiment, it is difficult to evaluate the dimension timeliness. The main objectives of CIRBA were on the ease of communication of information needs to source systems for information supply (ease-of-use), and the ability of the system to answer questions (content). Accuracy (e.g. the precision of a prediction or estimation) may be an important topic in specific user settings, though this value requires complex use settings, which are very hard to simulate. Format, the shape in which answers are presented (i.e. text, tables, pictures and graphs) has not been subject for the development of CIRBA, and thus CIRBA could not be evaluated on this respect. It also does not relate with the ontological issue focused on in the CIRBA design.

We asked users of the systems (the DWH and CIRBA) to state their agreement with the following statements on a 5-point scale (adopted from Doll et al, 1994):

With respect to content the following statements were used

- C1. The results of the system supply sufficient data to answer the question.
- C2. The results only contain relevant data.
- C3. The results contain all relevant data, not more nor less.

With respect to ease-of-use the following statements were used

- E1. I was able to gain the results quickly.
- E2. I was able to find the results easily.
- E3. The system gave useful assistance.
- E4. The system gave assistance while trying to find the results.

We asked 240 students, who passed the first year of their study in Computer Science to participate in the study for an hour by trying both systems via a Web version. 31 students responded positively on this re-

quest, completed the full experiment and were given a small refund (25 Dutch guilders) for their efforts. To control for learning bias in the results, half of the group was first given DWH and next CIRBA, and half of the group was first given CIRBA and next DWH. No statistically significant deviation on the dependent variables between both groups appeared. To gain some more information on learning biases, the respondents were given two extra statements to evaluate at the end of the experiment:

C7. The results of the first tool were better than that of the second tool.

E9. The ease of use of the first tool was better than that of the second tool.

For both systems all had the same assignment to accomplish, and had the same database available. The respondents had to perform the following (though seemingly simplistic often difficult to perform in practice) assignments by DWH and CIRBA:

1. Find first name, family name, and social security number of all persons of less than 40 years.
2. Find first name, family name, and social security number of all men of less than 40 years. Only select those who live in a rental house, which is built before 1985, with a rent larger than 900 guilders.
3. Find first name and family name of all persons with a rental house. The house has to be built before 1985 and the net rent should be less than 900 guilders. The landlord must have a renter's insurance.
4. Find first name, family name, and social security number of all persons with a weight less than 80 kilograms.
5. Find first name, family name, and social security number of all persons with a weight less than 80 kilogram, and larger than 165 cm. They should have a rental house built before 1985 and a rental insurance negotiated before 1999.

The only difference between DWH and CIRBA was that CIRBA has an ontology repository tool so that the user can explicit (part) of his ontology (i.e. data definitions) and change other parts of the ontology (i.e. the given data model), and has the availability of an interaction means (called the information broker) between user and the ontology. The database (source system) that was given had the simple structure of a "person" class connected to five attributes (loan, owner-occupied house, insurance, car, mortgage, rental house), and one object (bank) indirectly associated with "person" as presented in figure 4. This database was filled with fictitious data. The user interface, allowed the user to create his own ontology, by merging and subclassing (e.g. one could create a class "house" and identify the subclasses owner-occupied and rental house), renaming classes (e.g. rename "owner-occupied house" as "own-house"), and make specific selection to identify a new class (e.g. "special car" as all cars with a sales price of over euro 60.000).

Internal Data Market Services

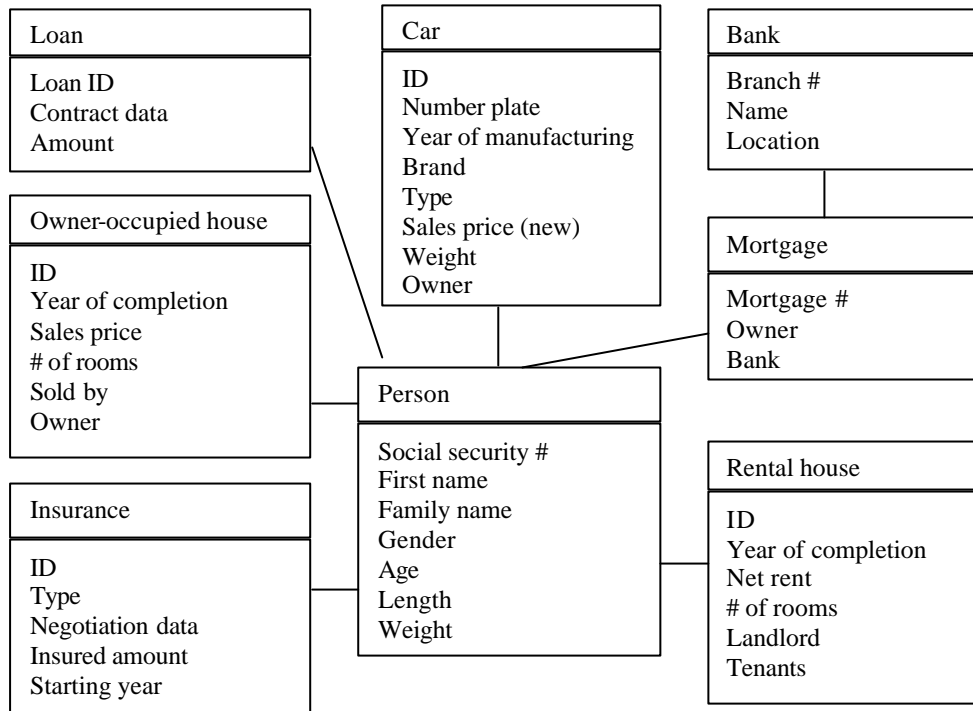


Figure 4: Data model of the assignment database

Observations and Analysis

For ease of analysis, we computed the average for the three content questions and the average for the four ease-of-use questions as a score for each respondent. Next we analyzed if the data show a significant difference between the DWH and the CIRBA. The conclusion is that indeed the CIRBA system delivers more satisfaction in terms of content (α is .002) and ease of use (α is .003). The data for this analysis are given in table 2 below.

	CIRBA	DWH		CIRBA	DWH
Score content (c1+c2+c3)/3	Frequency	Frequency	Score ease of use (e1+e2+e3+e4)/4	Frequency	Frequency
2.25	0	0	2.00	0	0
2.50	0	1	2.33	0	1
2.75	0	4	2.67	0	2
3.00	2	5	3.00	0	8
3.25	4	3	3.33	4	1
3.50	3	5	3.67	6	7
3.75	4	4	4.00	8	4
4.00	5	3	4.33	3	2
4.25	3	3	4.67	0	1
4.50	5	0	5.00	7	2
4.75	2	0			
5.00	0	0			
Average	3.90	3.41	Average	4.06	3.57
Mann-Whitney U = 216.000 Asymp. Sig. (2-tailed) is .002			Mann-Whitney U = 203.500 Asymp. Sig. (2-tailed) is .003		
Table 2: Data and analysis.					
For details on the statistical tests applied cf. Blalock, 1976.					

Conclusions

This study started with three questions. The answers found are reviewed and discussed below.

Review of question 1: *What architecture is able to match user and supplier ontology in a multi-user, multi-source internal information market?*

We explained the problems of matching ontologies of users and source systems, and explained how we think that the CIRBA architecture is a solution, which seems to considerably outperform a traditional DWH approach on the content and ease-of-use aspects of end-user satisfaction. The experiment though, has not been able to realize and test the problem of linking multiple source systems via wrappers and mediators. Wrappers and mediators are already on the market. We are not much worried about the feasibility of connecting different source systems to wrappers and mediators, and the CIRBA architecture gives a useful structure to this. The solutions to the ontological problems, though, are dependent on the capabilities of eliciting the ontologies of source systems, which may be complex in case of poorly documented source systems, and the possibilities of integrating these with the domain ontology. We rejected the idea of direct translations between different systems because of the huge complexity this may involve. We proposed domain ontologies as a necessarily intermediating structure. To realize optimal effectiveness of these ontologies we agree with Gruber (1995) that they should conform to the principles of clarity, coherence, extendibility, minimal encoding bias, and minimal ontological commitment. Users also have to be able to express their ontologies and present them to the data sources. The ease-of-use of the CIRBA interface indicates that this can be done highly effectively. Finally domain ontologies have to be developed and maintained, and an organizational task has to be allocated to this job.

Review of question 2: *Does this architecture result in higher user satisfaction compared to a data warehouse approach?*

In our experiment, the user satisfaction seems to be much better in the CIRBA context than by the use of DWH. In practical situations where short-term decisions have to be made in a standardized form, though, the need for CIRBA is unclear for end-user satisfaction. Explicit ontology management is required for the management of the corporate data models and for keeping a high quality interaction between source systems and user needs. Data model maintenance in large systems (i.e. corporate data bases of ERP systems) requires strong support from an easy, well-structured software-support. The developments in the field of ontologies are promising, though in practice many people doubt its value and these ontology-based systems require higher technical performance of the systems. This study showed that data model management is far from abstract and of absolute importance for the successful exploitation of data sources in internal data markets (eventually also for external and commercial uses).

Review of question 3: *What generic insights does this architecture deliver for the design of internal data markets services?*

The CIRBA architecture gives important operational ideas for analyzing the interactions between supply and demand in internal data markets. The refinery model mentioned in the beginning of the article analyzes the internal data market from the process activities needed for internal data markets. Cirba gives an example of how the refinery process model can be elaborated to an information systems architecture. Additionally, though, any market service (being activities and systems to match supply and demand on a market) also requires a business and organizational model (Picard, 2000; Wijnhoven, 2002). The business model defines what services and products have to be produced and delivered for what conditions (e.g. prices and costs) to what market groups. The organizational model defines the ownership rights and the responsibilities of people and departments with respect to the data sources' ontologies and service activities. The organizational model also needs a definition of required skills and organizational means. This is an important issue for realizing CIRBA as an organizational success. The literature on DWH has shown that insufficient attention has been paid to organizational issues around them, resulting in substantial failures (Finnegan & Sammon, 1999). This may not be repeated. A data market service is only complete when the business model, the organizational model, the process model and the systems architecture have been realized.

We believe that we have been able to lay down useful concepts and rather simple concepts to create the IT support for DMS. We have the perception that much of DMS development currently is based on trial and error efforts. Because this work is developed in the context of Ordina's work as IT consultants, we think that its practicality could be high, but only systematic practical experience can say if this is really true. We would appreciate any attempt of sharing experiences on this topic. From these experiences, the development of a methodology for DMS design may be feasible and an important contribution for DMS theory and practice.

REFERENCES

- Austin, A.M. et al. (1996). *Mediator Project. Idaho National Engineering & Environmental Laboratory*.
[Http://id.inel.gov/idim/mediator_summary.html](http://id.inel.gov/idim/mediator_summary.html) Visited December 2000.
- Bischoff, J & Alexander, T. (1997), *Data warehouse: Practical advice from the experts*. Upper Saddle River, NJ: Prentice-Hall.
- Blalock, H.M. (1972). *Social Statistics* (2nd ed.). Tokyo: McGraw-Hill.
- Boisot, M. H. (1998). *Knowledge assets: Securing competitive advantage in the information economy*. Oxford: Oxford University Press.
- Borst, P. & Akkermans, H. (1997). Engineering ontologies. *International Journal of Human-Computer Studies*, 46: 365-406.

- Broadbent, M. Weill, P., O'Brein, T. & Siong Neo, B. (1996). Firm Context and Patterns of IT Infrastructure Capability. In: De Gross, J.I. *Proceedings of the Seventeenth International Conference on Information Systems*: 174-194, New York: ACM.
- Davenport, Th. H. (1998). Putting the enterprise into the enterprise system. *Harvard Business Review*, 76 (4): 121-133.
- Davenport, Th. A. & Prusak, L. (1997). *Working knowledge: How organizations manage what they know*. Cambridge (MA): Harvard Business School Press.
- Doll, W.J., Xia, W. & Torkzadeh, G. (1994). A Confirmatory factor analysis of the end-user computing satisfaction instrument. *Management Information Systems Quarterly*, 18 (4): 453-461.
- Finnegan, P. & Sammon, A. (1999). Foundations of an organisational prerequisites model for datawarehouses. *ECIS 1999 proceedings*, pp. 182-193. Copenhagen: Copenhagen business School.
- Gruber, T.T. (1995). Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43 (5-6): 907-928.
- Inmon, W. (1996). The datawarehouse and datamining. *Communications of the ACM*, 39 (11): 49-50.
- König-Ries, B. (2000). An approach to the semi-automatic generation of mediator specifications. *Lecture notes of computer science*, 1777: 101-117.
- Kotler, P.J. (2000). *Marketing Management*. Upper Saddle River, N.J.: Prentice-hall.
- Laudon, K.C. & Laudon, J.P. (2001). *Management Information Systems: Organization and Technology*. Upper Saddle River (N.J.), Prentice Hall, seventh edition.
- Markus, M.L. (2001). Toward a theory of knowledge reuse. *Journal of Management Information Systems*, 18 (1): 57-94.
- Meyer, M. H. & Zack, M.H. (1996). The Design and Development of Information Products, *Sloan Management Review*, 37 (3): 43-59.
- Mitra, P., Wiederhold, G. & Kersten, M. (2000). Mediators and semantic integration: A graph oriented model for articulation of ontology dependencies. *Lecture notes in computer science*, 1777: 86-100.
- Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation, *Organizational Science*, 5 (1): 14-37.
- Picard, R.G. (2000). Changing Business Models of Online Content Services: Their implications for multimedia and other content producers. *Journal of Media Management*, 2 (2): 60-68.
- Schreiber, A.Th., Akkermans, H., Anjewierden, A., De Hoog, R., Shahbolt, N, Van de Velle, W. & Wielenga, B. (2000). *Knowledge Engineering and Management: The commonKADS*. Cambridge MA: MIT Press.
- Shaw, M.L.G. & Gaines, B.R. (1989). *Comparing conceptual structures: consensus, conflict, correspondence and contrast*. Calgary (CA): University of Calgary.
- Shapiro, C. & Varian, H. (1999). *Information Rules: a strategic guide to the network economy*. Boston (MA): Harvard business school press.
- Stamper, R.K. (1973). *Information in business and administrative systems*. New York: John Wiley.
- Stamper, (1988), Semantics. In: Boland, R.J., Hirschheim, R.A. (eds.), *Critical issues in information science*, pp. 43-78. New York: John Wiley.
- Timmers, P.J. (1998). *Electronic Commerce: Strategies and models for business-to-business trading*. Chichester: Wiley.
- Wijnhoven, F. (2001). Models of Information Markets: Analysis of markets, Identification of Services, and Design Models. *Informing Science*, 4 (4): 117-128.
- Wijnhoven, F. (2002). Design components for information market services: A framework for research and practice. In J. Monteiro, P. Swatman & L. Valadares Tavares (eds.), *Towards the knowledge society: eCommerce, eBusiness and eGovernment*. Boston (MA): Kluwer Academic Publishers, pp. 49-65.

Biographies

Fons Wijnhoven has a BA in international relations, a MA in research methodology, and a PhD in management. Fons is an associate professor of knowledge management and business information systems. Fons researches the development of electronic intermediaries for intra organizational and inter or-

Internal Data Market Services

ganizational knowledge exchanges. He has articles on information market services in journals like Journal of MIS, Knowledge & Process Management, Journal of Electronic Commerce Research, and Journal of the OR Society. He is the author of "Managing dynamic organizational memories" (Twente University Press, 1999).

Eddy Verbruggen has a M.Sc. in Business Information Technology, and he is currently working for Ordina Finance Business Solutions. He works on the ontology-based mediator architecture, which is one of Ordina's innovative concepts.

Edwin van den Belt has a M.Sc. in computer science. Edwin is the leader of Ordina's R&D in the application of ontologies and mediators for information retrieval purposes.

Paul van der Vet has a B.Sc. in chemistry, a M.Sc. in philosophy, and a Ph.D. in chemistry. He is assistant professor in the Department of Computer Science. He lectures on artificial intelligence, knowledge representation, and information retrieval. He is involved in the M.Sc. program "Information engineering" that is offered by the Universities of Twente and Osnabrueck. He has written on the history and philosophy of science, ontology building, information retrieval, and natural-language processing. He is currently applying content-engineering techniques to bio-informatics problems.