A Performance Measurement Perspective for Event-Driven Process Chains*

Veronika Stefanov and Beate List
Women's Postgraduate College for Internet Technologies
Vienna University of Technology
{*stefanov, list*}*@wit.tuwien.ac.at*

Abstract

Business processes access Data Warehouse (DWH) information for performance measurement, supporting business decisions. Today, there are no conceptual models available that make the relationship between DWH architectures and business processes transparent. In this paper, we extend the Event-Driven Process Chain, a business process modeling language, with an additional perspective to make this relationship explicit in a model.

1. Introduction

In modern organizations, Data Warehouses (DWH) are utilized for performance measurement [8]. Defined as "a subjectoriented, integrated, time-variant, nonvolatile collection of data in support of management's decision-making process" [6], they are more than just big databases: Standalone DWHs of the early 1990s have evolved into today's large DWH architectures comprising a DWH, several data marts, sometimes an operational data store, as well as Online Analytical Processing (OLAP), data mining, and reporting tools for accessing data.

The components of a DWH architecture provide Key Performance Indicators (KPIs), also called metrics or performance measures in other disciplines, to business processes. There are many examples showing how important KPIs have become for business processes: When a person applies for a loan in a bank, the applicant is scrutinized to find out if she or he has caused a financial loss to the bank previously, or is likely to do so in the future. The DWH supplies the data. The process of designing a new product in a telecommunication company or an airline, or the composition of the product range in a supermarket require information on customer behavior from the DWH.

*This research has been funded by the Austrian Federal Ministry for Education, Science, and Culture, and the European Social Fund (ESF) under grant 31.963/46-VII/9/2002.

Surprisingly, this knowledge - how dynamic business structures access parts of the DWH architecture and how these are being used in every day business life - is not made explicit in existing models. There is a need for an integrated model to make the relationship between the DWH architecture and the business processes more transparent.

To bridge this gap, we extend a business process modeling language with an additional perspective, the Performance Measurement (PM) Perspective, to be able to create models that show

- where and how business processes use KPIs from specific areas of the DWH architecture, and
- which areas of the DWH architecture are required by which parts of the business processes

Such a model provides the following contributions for process managers and DWH designers:

- It shows where the business processes utilize KPIs.
- Business processes enriched with the PM Perspective can support the design phase of a DWH project, making it possible to prioritize the projects according to business needs.
- The PM Perspective can be used to justify the costs of DWH projects by pointing out the unseen relationships between important business decisions and the DWH.
- Such a model provides transparency: If a data mart fails, an integrated model enables better reactions because it is known which business processes are affected.
- It also allows to discover which parts of the DWH architecture are not accessed at all, and decide if those parts should be further maintained.

The business process modeling language we have chosen to extend is the Event-Driven Process Chain (EPC, section 2). We have added the PM perspective (section 3), which allows to show different levels of aggregation and points of view, from accessing the full DWH or a data mart, to single facts and measures, as well as KPIs in reports. We have identified three typical usage scenarios (section 4) for the PM Perspective. The applicability is shown with an example in section 5, followed by related work in section 6.

2. Event-Driven Process Chains (EPCs)

The *Architecture of Integrated Information System* (ARIS) [7] divides complex business process models into separate views to reduce the complexity. The views can be handled independently. There are three views focusing on functions, data, and the organization (see Figure 1), and an additional view focusing on the integration of the other three.

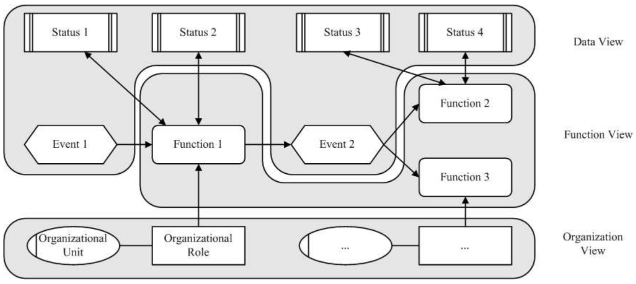

Figure 1. ARIS Views

The *Data View* contains events and statuses. Events such as "customer order received", or "invoice written" are objects that represent data. Statuses such as "customer status" and "article status" are also represented by data. As it was the most widespread design method in the area of data modelling, Chen's Entity-Relationship (E/R) model [3] was adopted into the ARIS framework. Today, the UML class diagram is also used [9].

The *Function View* describes the activities to be performed by the process, the individual subfunctions, and their relationships.

The *Organization View* represents the organizational structure. This includes the relationships between organizational units, between employees and organizational units, and employees and roles.

The *Control View* links functions, organization and data, thus integrating the design results of the different views. The various elements are connected into a common context by the control flow. The resulting model is the EPC.

EPCs are used by many companies for modelling, analyzing, and redesigning business processes. They were developed in 1992 at the Institute for Information Systems of the University of Saarland, Germany, in collaboration with SAP AG. The EPC is based on the concepts of stochastic networks and Petri nets.

A basic EPC consists of the following elements (Figure 2): *Functions* are active elements. They model the activities within the company. *Events* are created by processing functions or by actors outside of the model. An event may act as a pre-condition or post-condition of a function. *Logical operators* (AND, XOR and OR) connect functions and events. The extended EPC adds the following elements: An *Organization Unit* or *Role* is responsible for performing a function. *Information Objects* represent input or output data of a function. They correspond to entities or attributes of the E/R model.

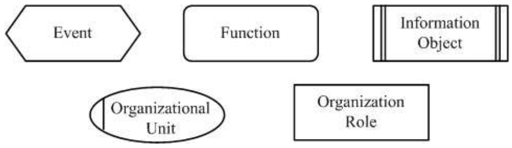

Figure 2. EPC Elements

3. Performance Management Perspective

We extend the EPC with an additional perspective. It integrates the information about where the process makes use of the KPIs provided by the DWH architecture. These models make the hidden knowledge about the relationships between the business processes and specific areas of the DWH architecture explicit.

3.1. The Extended Meta-Model of the EPC

We have chosen the EPC as a basis for our model because of its wide-spread use in many companies for modelling business processes, and because of its flexible view concept, that allows to separate the different aspects of a business process. We can easily add another perspective while keeping the original structure intact.

The EPC meta-model (white) including the *PM Perspective* (dark) is shown in Figure 3. Each EPC consists of one or more *Functions* and two or more *Events*, as an EPC starts and ends with an event and requires at least one function for describing a process.

An *Additional Process Object* may be assigned to one or more functions, for example an *Information Object* or an *Organizational Structure*. All types of additional process objects may be assigned to any function.

We extend the EPC meta-model with the PM perspective and introduce a *PM Information Object* as an additional process object. The detailed meta-model of the PM perspective is shown in Figure 4. All elements of the meta-model are specializations of the PM Information Object. All PM information objects are additional process objects in terms of the EPC, which means that they can be assigned to a function that uses the KPIs supplied by them.

What is a PM information object? It represents the ways in which a business process might access specific areas of the DWH architecture, e.g. decision makers might use reports or analysis tools to obtain KPIs. In a detailed model of a business process, the individual data entities and attributes accessed by a function could be shown. On the larger scale,

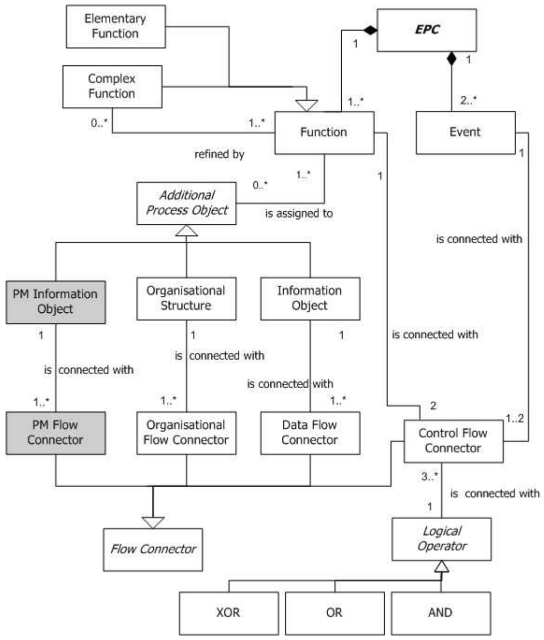

Figure 3. EPC Meta-Model with PM Information Object and PM Flow Connector

a process accesses the whole data warehouse, or individual data marts.

We have thus identified three main categories of PM information objects: *PM Data Repositories* (the different databases of the DWH architecture), *PM Data Objects* (the elements of a data model), and *PM Information Presentation Objects* (the data presentation to the user). They are shown in Figure 4.

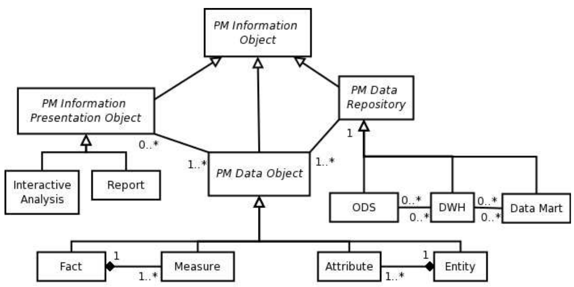

Figure 4. Meta-Model of PM Perspective

3.2. PM Information Presentation Objects

In a DWH architecture, there are usually tools and applications providing users with prearranged information. We call these collections of information *PM Information Presentation Objects*, and have identified two different types: *Report* and *Interactive Analysis*. A report is a kind of document containing the KPIs related to a certain area, for example a report on sales in the south region for the 4th quarter of 2004. The values contained in a report do not change over time. An interactive analysis on the other hand is closer to a tool. It provides its users with regularly updated values and can be used for continuous performance monitoring. In the EPC we can show e.g. a certain report that a function accesses. The notation for PM information presentation objects is shown in Figure 5.

Figure 5. Notation of PM Information Presentation Objects

3.3. PM Data Objects

In order to provide a more detailed view of the data accessed by the functions of an EPC, we also want to model the individual data entities contained in the PM data repositories. These *PM Data Objects* are generally represented by conceptual data models. If a function needs data on the revenue of a certain product range, it can be modeled to access the corresponding PM data object directly.

Depending on the type of repository, the overall architecture, and the preferences of the designers, different kinds of data models can be used. The two main types relevant to DWH applications are entity-relationship modeling and multidimensional modeling [4]. In the first case, we use the Entity-Relationship model (E/R) [3] and in the latter, the Multidimensional Entity-Relationship (ME/R) model [10].

We have chosen the ME/R as a conceptual model for multidimensional modeling because of its simplicity and expressiveness. The ME/R extends the E/R model by adding three elements that are specializations of existing E/R elements. In terms of a multidimensional model, a *fact* table contains the subject of analysis (e.g. sales fact, transactions fact, etc.). Its attributes (e.g. amount, duration, etc.) are called *measures*.

The PM data objects of an E/R model are either *Entities* or individual *Attributes*. In the case of the ME/R, they are *Facts* or *Measures*. Their notation is shown in Figure 6. Whether entities/facts or attributes/measures are to be used as additional process objects in the EPC depends on the granularity of the EPC functions.

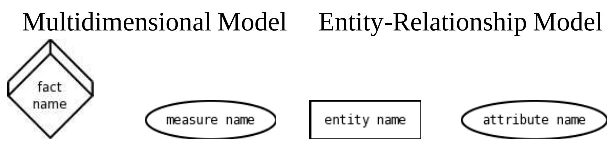

Figure 6. Notation of PM Data Objects

3.4. PM Data Repositories

PM Data Repositories represent different types of databases as used in DWH settings. Our approach is not limited to any specific DWH architecture, but can be applied to a wide selection of architecture types.

Different combinations of PM data repositories may occur in an organization. In large multinational organizations it is not uncommon to have more than one DWH. A large DWH often co-exists with smaller Data Marts (DM), departmental subsets of a DWH focused on selected subjects [2]. The data marts might obtain their data from the data warehouse, meaning that a data mart acts as a kind of materialized view on the DWH. Or the data marts may be created individually, and then later be integrated into an organization-wide DWH, making operations that span several data marts possible. Also, there may be none, one or more Operational Data Stores (ODS), located between the operational systems and the DWH [5]. Depending on the architecture, end user applications may query individual data marts and/or the DWH, or even access the data in the ODS directly.

In order to allow the greatest possible flexibility while still providing meaningful content in the models, we have identified three basic types of PM data repositories: the *DWH*, the *DM*, and the *ODS*. The notation for these elements is shown in Figure 7.

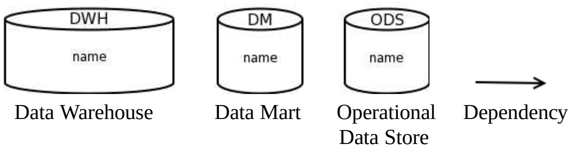

Figure 7. Notation of PM Data Repositories

To illustrate the relationships between the PM data repositories we propose a simple repository dependency diagram. In the example shown in Figure 8, the DWH depends on two independent ODS systems, and in turn supplies four data marts with PM data.

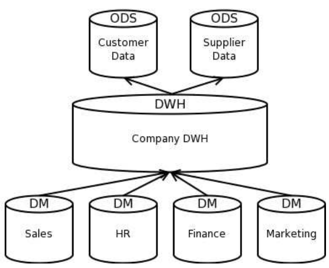

Figure 8. Data Repository Dependency Diagram

4. Usage Scenarios

We have identified three main scenarios regarding the usage of the PM perspective. They depend on the target user group, and offer modeling solutions for typical every-day requirements.

DWH managers are looking for the big picture, an overview of what is going on. They will use EPCs showing business processes with the PM data repositories. This allows them to find answers to questions such as ”Which processes use this data mart?”, ”Which business processes require direct access to the DWH?”, or ”If this data mart fails, which processes are in danger?”.

Business users are interested in business decisions. An EPC model showing an individual business process or subprocess in connection with reports and interactive analyzes will support questions such as ”Which reports are accessed where?” and ”Which important business decisions are supported by the DWH environment?”

DWH designers and *developers* need to understand the details of how the DWH environment is used. They will use a fine-grained model of a business process or subprocess with facts and entities, attributes or measures.

5. Example: EPCs and the PM Perspective

We have extended the EPC with the PM perspective. The PM information objects can be used as additional process objects, indicating that a function accesses the information provided by the object.

The example process in Figure 9 does not correspond directly to one of the usage scenarios introduced in the previous section, but it is better suited as a light-weight example. It illustrates the use of the PM data object ”Fact” and the PM data repository ”DWH”.

An insurance company checks every claim it receives for potential fraud. The three-step fraud detection process is aimed at reducing the overall cost for the company by minimizing the number of claims that have to be checked by human specialists.

Therefore, each claim is first subjected to an automated check that sorts out the large amount of genuine claims. Beyond identity verification and checks against lists of known offenders, this step involves various analysis techniques (significant outliers, conflicting or duplicate transactions, etc.). The function *automated check* does not need access to the DWH. As a result of the function, the claim is either *suspicious* or *not fraud*. The *suspicious* claims are forwarded to the fraud detection department for review and further analysis.

The function *review* lies in the responsibility of the fraud detection department. The details of the previous step are

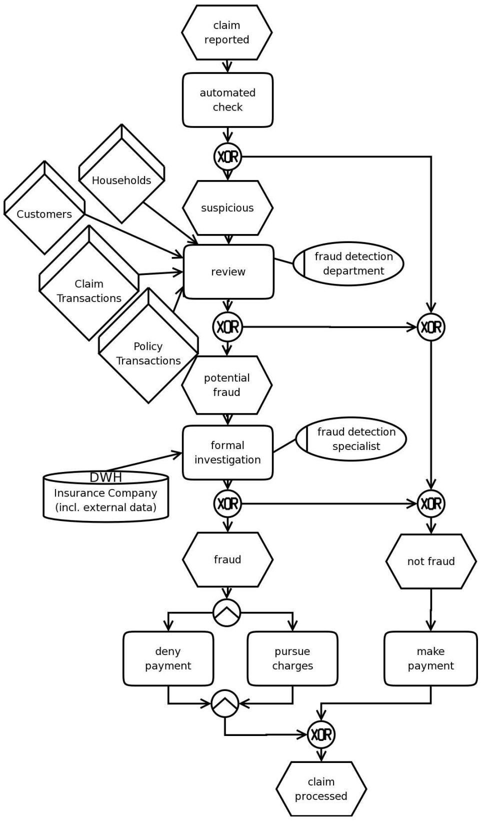

Figure 9. Example of EPC with PM Elements

reviewed and compared to similar cases. The function accesses the information contained in the facts *Households*, *Customers*, *Claim Transactions* and *Policy Transactions*. The output events of this function indicate whether the claim has been identified as *potential fraud* or as *not fraud*. For the false positives generated by the automated check, the required payments will be made.

In the final step, the claims likely to be fraudulent are analyzed in depth by the function *formal investigation*. This requires access to the whole data warehouse, which, in addition to the usual wealth of company internal data, also contains external data. This is typical for a process step that cannot be pre-defined in detail. If the insurance claim has been successfully identified as fraudulent, the payment is *denied* and charges are *pursued* against the claimant.

6. Related Work

There are a lot of conceptual models available for business processes, as well as for databases or DWHs. But there are no models available that focus on the relationship between the DWH and the business processes. EPCs [7] incorporate a data view targeting operational data bases. EPC functions perform read or write operations on the databases and their entities (different to the DWH). A similar concept is included in our PM perspective, but goes much further.

In UML Activity Diagrams [9], data is represented by data store nodes. A UML Action can perform read or write operations, like the EPC function. The data store node is not necessarily linked to a UML class or database.

The Business Process Modeling Notation (BPMN) [1] provides data objects, which are used and updated during the process. The data object can be used to represent many different types of object, both electronic or physical.

7. Conclusion

We have extended a business process modeling language, the Event-Driven Process Chain (EPC), with an additional perspective, the Performance Management (PM) Perspective. The new extended conceptual model bridges the gap between the static structures of the DWH architecture and the dynamic structures of business processes. It enables us to show where and how business processes use KPIs from specific areas of the DWH architecture, and which areas of the DWH architecture are required by which parts of the business processes. The PM perspective offers modeling elements representing the means of data presentation, different types of repositories, and the data model.

References

[1] Business Process Modeling Notation (BPMN). Specification BPMN 1.0 May 3, 2004, 5 2004.

[2] S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. *SIGMOD Rec.*, 26(1), 1997.

[3] P. Chen. The Entity-Relational model: Toward a unified view of data. *ACM Transactions on Database Systems*, 1(1), 1976.

[4] M. Golfarelli, D. Maio, and S. Rizzi. Conceptual Design of Data Warehouses from E/R Schemes. In *Proc. Hawaii Int. Conf. on System Sciences, Kona, Hawaii*, 1998.

[5] M. Golfarelli, S. Rizzi, and I. Cella. Beyond Data Warehousing: What's Next in Business Intelligence? In *Proceedings of DOLAP '04*. ACM Press, 2004.

[6] W. Inmon. *Using the Data Warehouse*. John Wiley & Sons, New York, 1994.

[7] G. Keller, M. Nüttgens, and A.-W. Scheer. Semantische Prozeßmodellierung auf der Grundlage "Ereignisgesteuerter Prozeßketten (EPK)". *Veröffentlichungen des Instituts für Wirtschaftsinformatik*, (89), 1992.

[8] B. List and K. Machaczek. Towards a Corporate Performance Measurement System. In *Proceedings of the 19th ACM Symposium on Applied Computing (SAC'04)*. ACM Press.

[9] Object Management Group, Inc. UML 2.0 Superstructure. http://www.omg.org/cgibin doc?ptc/2004-10-02, 3 2005.

[10] C. Sapia, M. Blaschka, G. Hoefling, and B. Dinter. Extending the E/R Model for the Multidimensional Paradigm. In *Proceedings of ER '98*. Springer-Verlag, 1999.