An Evolutionary Approach to IT Support for Medical Supply Centers

Gregor Endler[1], Michael Langer[2], Jörg Purucker[3], Richard Lenz[1]

[1] Chair for Computer Science 6 (Data Management)
University of Erlangen-Nuremberg
Martensstraße 3
91058 Erlangen, Germany
{Gregor.Endler, Richard.Lenz}@cs.fau.de

[2] Astrum IT
Am Wolfsmantel 46
91058 Erlangen
Michael.Langer@astrum-it.de

[3] Chair of Information Systems II
University of Erlangen-Nuremberg
Lange Gasse 20
90403 Nuremberg, Germany
joerg.purucker@wiso.uni-erlangen.de

Keywords:
data integration, data quality, dataspaces, smooth migration, demand-driven improvement

Abstract:
In this paper, we show that in cooperative environments like medical supply centers, new demands on IT infrastructure are created that cannot be satisfied by traditional information systems. We demonstrate our vision for a system to remedy this situation by offering smooth pay-as-you-go migration instead of forced integration to autonomous data management systems.
We give an evaluation of applicability of traditional techniques in the field, followed by our planned approach for this project, including a panoramic view of our proposed system architecture.

1 Introduction

Cooperative environments are a current trend in health care. To increase their power to compete [HE09], single practitioners often pool their resources in both medical supply centers and networked practices. Both structures create new demands on IT infrastructure,

erschienen im Tagungsband der INFORMATIK 2011
Lecture Notes in Informatics, Band P192
ISBN 978-3-88579-286-4
weitere Artikel online:
http://informatik2011.de/519.html

especially through the new role of clinical practice manager. These administrative officers are responsible for supervision and controlling of the center's financial development, and as such, need a consolidated overview of all processes and events within the network. This information however is largely hidden within the insular patient data management systems of the participating practices, and very often, communication between practices and practice manager is paper-based only. The technical needs of a central point of information, being completely new to the sector, are not adequately supported by current software systems.

The current situation is defined by pre-existing independent systems. As most practices had patient management systems long before joining a medical supply center, the practices' systems are heterogeneous and autonomous in different organizational units. Requiring a replacement of these systems when joining a cooperative environment is not an option, as the necessary expense in time and resources would likely deter most practitioners from joining at all.

This article introduces a BMBF supported joint project by the University of Erlangen-Nuremberg (Chair for Data Management and Chair of Information Systems II) and the software company Astrum IT. The project is aimed at conceptualizing and developing an ERP system for medical supply centers which minimizes the efforts for interconnecting autonomous systems and supports step by step continuous improvement of IT supported cooperation, as well as being capable of delivering data on a day-to-day basis.

2 Related Work

In this section, we examine existing solutions for problems related to our project, and whether they can be integrated with our approach.

2.1 Integration

There are several traditional approaches to integration of heterogeneous database systems. Among them are for example the ETL process in data warehousing [CD97] and building schema mappings in federated databases or multidatabases [SL90]. While safe and well known, these methods lead to a very high upfront effort for semantic integration as well as systems mostly resistant to change. Since for our project, new systems are expected to appear frequently, and current ones may change at any time, a standard ETL process is only usable for the static parts of our project. For the components subject to change, it would take too long to deal with modifications in the patient data management systems, compromising our day-to-day requirement. For this reason, we pursue an evolutionary approach to integration in addition to the traditional one (see sections 2.3 and 3).

Even though there are many automatic techniques for semantic integration [RB01], resolving it still is expensive. For this reason, we investigated how an approach motivated by dataspaces [FHM05], allowing coexistence of heterogeneous data sources, could serve as an integration platform. Data coexistence, for our ends, denotes that little requirements are posed to newly attached patient data management systems. A new system is integrated only as far as automatically possible, and all additional data needs, particularly supporting use cases aimed at cooperation among physicians, are addressed in a pay-as-you-go manner [JFH08], gradually improving with the help of user feedback.

2.2 Data Quality

Data quality ties in closely with integration in this case. In accordance with the dataspaces approach, imperfections in quality of newly attached systems will be improved by degrees as needed [JFH08], similarly to our approach to integration. To support a demand-driven step by step improvement, our system architecture is complemented by a metadata repository aimed at data quality monitoring. This was inspired by the TDQM (Total data quality management) approach [Wan98], which requires data quality measuring and monitoring as a basis for continuous data quality improvement. Many different dimensions of data quality are named in the literature [SMB05]. For a closer look on some of them, [BS06] gives definitions of accuracy, completeness, time related dimensions and consistency. We are currently in the process of identifying the dimensions most important for our chosen area of application.

In addition to preset quality features, the users, being domain experts, should be able to define their own data quality characteristics. To accomplish this, a system for defining and verifying rules [BLL10] will be integrated in our application to enable demand-driven evolution of data quality monitoring.

2.3 Evolutionary Information Systems

Detecting integration and data quality needs is only one part of the project. Another important concern is the system's capability to evolve over time with minimal effort [Len09]. This includes extending the database by generic database tables as well as mechanisms that allow handling of inconsistent, inaccurate or incomplete data within the central database. The rules system for data quality mentioned above also creates evolutionary capabilities.

This is important because the healthcare system, as well as the technical details of patient management systems, are subject to frequent changes, some unforeseeable. To keep this in mind is essential, especially at the beginning of a project: Called the principle of ”design for change” in [Par94], a system's architecture lays the foundation for the system's ability to evolve. The effects of lacking evolutionary power are described in [Leh96] through the eight ”Laws of software evolution”.

Relying on medical standards like xDT or HL7 may prove a crutch to the software's evolvability. To begin with, relevant data may not be covered. Neither are future data needs guaranteed to be integrated into the standards. Secondly, many standards like HL7 are tied almost exclusively to the hospital sector and do not appear at the patient management systems of resident doctors. Lastly, we need to keep in mind the possibility that a system

erschienen im Tagungsband der INFORMATIK 2011
Lecture Notes in Informatics, Band P192
ISBN 978-3-88579-286-4
weitere Artikel online:
http://informatik2011.de/519.html

we want to integrate doesn't support any standards at all. This is why we are planning software fully capable of making use of established standards without relying on their existence.

2.4 Existing Implementations

Eleven existing implementations of cooperative medical environments were evaluated and compared to our approach in over thirty categories like support for controlling and accounting, data security, and flexibility. As it turned out, none follow a holistic approach: Only singular use cases are supported, and newly emerging requirements like integrated controlling or treatment coordination are implemented in new autonomous applications, even though the data sets are similar. This leads to ever increasing heterogeneity and aggravates the data quality problem, especially for use cases requiring comprehensive data. In contrast, we strive to build a system capable of supporting newly emerging use cases on the already existing database.

3 Approach

The goal of our project is building an ERP-like system offering a consolidated view on appropriate patient data, enabling on-the-fly integration and improvement of patient data management systems through smooth migration. In the following, we use the term ERP loosely, since we focus on uncovering data hidden in insular sources, and do not enable direct write access to them. A medical supply center or practice network does not have the same structure as a regular enterprise but is rather more loosely organized. Still, central supervision is desired, and future cooperative needs may be unforeseeable. Thus, we can not use off-the-shelf ERP solutions. In the following, we sum up our approach to this end, including a basic system architecture.

3.1 Identification of Requirements

In cooperation with several physicians, a broad spectrum of use cases was identified (see Figure 1 for an example). The core data needs for the administrative use cases were identified. Methods for continuous data quality management were analyzed and adapted to the particular data quality needs in this project. Based on the data needs for the administrative use cases we developed a core database schema for a central database, combining a fixed schema, satisfying the known data needs, with the EAV paradigm [NMC+99] to support future data requirements. This database serves as a basis for multiple applications providing a spectrum of enterprise resource planning functionality for medical supply centers.

erschienen im Tagungsband der INFORMATIK 2011
Lecture Notes in Informatics, Band P192
ISBN 978-3-88579-286-4
weitere Artikel online:
http://informatik2011.de/519.html

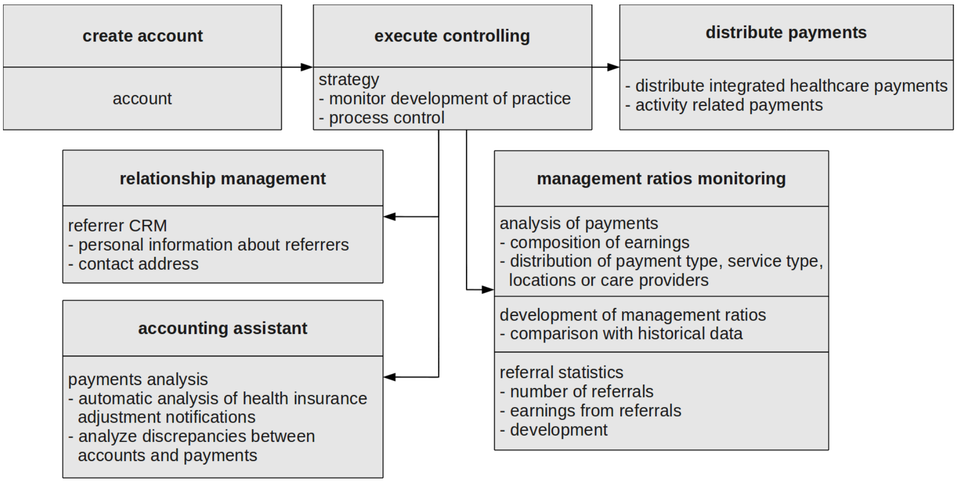

Figure 1: Use Case: Remuneration analysis

3.2 System Architecture

Figure 2 shows a panoramic view of the proposed architecture of our project. To increase clarity, the metadata monitoring components are omitted. Please note that the arrows connecting the components denote dataflow, not direct connections. Communication will take place via messages, to ensure the flexibility of the system via loose coupling. In contrast to a classic hub-and-spoke architecture, we do not aim to enable communication between the attached systems, but to show a consolidated image of the whole data pool. The autonomous patient management systems are each fitted with a staging database. These exist physically and independently on site of the practice. This way, direct read-only access to the patient management systems is possible, barring system outages. The connection to the integration engine is made via Internet, taking care to keep the systems only loosely coupled, for example by adapting the REST architectural style. The integration engine then performs what would be the transform and load steps in traditional ETL. The staging databases are necessary since data changed during a possible outage of remote connectivity must not be lost. Additionally, data privacy is a concern. Since the protection of patient data is a highly delicate matter, all data must be automatically anonymized by the time they reach the staging databases, except if the patient in question has explicitly agreed to an exchange of data. If this is not the case, a de-anonymization must not be possible.

The integration engine is responsible for coordinating the data into one consolidated view. This includes schema and identity matching, duplicate elimination, and semantic integration issues. For our current use cases, it is enough to deliver data current to the day. If future use cases require data transfers more often or even on demand, the frequency can be increased accordingly.

The integration engine then relays the adjusted data into the central database. Special application databases, for example for controlling, load the needed data directly from the

erschienen im Tagungsband der INFORMATIK 2011 weitere Artikel online:
Lecture Notes in Informatics, Band P192 http://informatik2011.de/519.html
ISBN 978-3-88579-286-4

central database. Introducing separate databases for different applications has several advantages. For one, giving the applications only an excerpt of the available data will net an increase in the speed of any tasks performed on the smaller data pool.

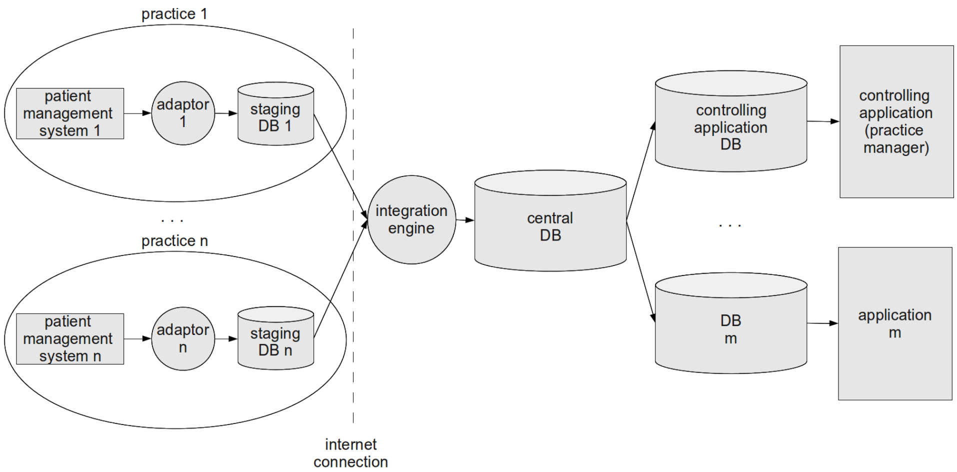

Figure 2: Panoramic view of system architecture

More important though is the increase in data privacy gained by physically disconnecting application data from the central database.
These two reasons hold true especially in the case of business controlling, where mainly aggregated financial data is needed. Without a separate database to hold the pre-aggregated data [CD97], the aggregation either would have to be performed on-the-fly, leading to poor performance, or be stored in the central database, leading to a cluttered central database.
Lastly, if all applications share one data pool, the central database becomes a bottleneck possibly blocked by a resource hungry application.
That said, not all possible applications arguably need their own data pool, and some may conceivably work directly on the central database. Special care has to be taken to avoid the mentioned deficiencies.

3.3 Pay-As-You-Go Improvement

Through the whole process, metadata is read at appropriate points and stored in a metadata repository. This forms the basis of our data quality monitoring system. Here, possible steps to improve data quality are identified, catalogued and ordered by usefulness and urgency [JFH08]. The most important ones are then propagated to the users to ask for feedback on or revision of the detected problems. In this way, new systems are able to be attached and deliver data at once, even if they lack a certain grade of data quality. The most pressing of these concerns will be addressed promptly, while less important ones will improve over time. The necessary metrics and mechanisms are a major part of our ongoing work.

erschienen im Tagungsband der INFORMATIK 2011
Lecture Notes in Informatics, Band P192
ISBN 978-3-88579-286-4
weitere Artikel online:
http://informatik2011.de/519.html

3.4 A User Story

As an example, consider the following user story. An medical supply center (MSC) manager in the process of reviewing treatment data for disease management programs receives a warning from the DQ monitoring component that the data he is currently accessing is not complete as measured by appropriate metrics. Looking up the provenance of the data, the manager realizes that there is one practice location that has not delivered any data at all on diagnoses and treatments.

He contacts the corresponding site manager to find the cause for this problem. It turns out the practice stores diagnostic and treatment reports in record cards rather than its patient management system. When their information system was attached, all treatment entries for the patients of this location were left blank. While this was no a problem at the time, it nevertheless was detected as an insufficient degree of the quality dimension completeness, and was presented as a warning when the relevant data was accessed the first time.

The MSC manager urges the site manager to use their patient management system rather than the record cards. After migration of the data to the patient management system, the information displayed to the MSC manager shows up as complete.

New problems may arise after improvements, of course. In this example, practitioners may be unfamiliar with the new way of storing diagnostic reports, potentially leading to incorrect entries or delays. However, these problems will be recognized, and addressed if necessary, in the next iteration of the data quality management cycle.

4 Conclusion

We presented preliminary results of an ongoing project organized within the "Spitzencluster Medical Valley EMN". We demonstrated the initial situation we are building on, as well as the need for new ways of approaching integration and improvement of information systems used in medical supply centers. We gave an overview of the approach we envision for this project.

Since we are still at the beginning of the project, many open issues remain. The focus is how to enable smooth migration, starting from very little requirements on the autonomous systems and moving gradually and gracefully to full integration. Especially measuring and weighting of data quality concerns are of interest here. Identifying the necessary dimensions and developing metrics and quantifying strategies, as well as policies on how to deal with imperfections, will be our main short term goal.

The system is currently in development at Astrum IT in conjunction with the University of Erlangen-Nuremberg. Programmers, software architects and team leaders work closely with partners in the medical sector to identify requirements and verify concepts. At the end of the project runtime, the medical partners will test the usability of the system in a six month evaluation phase incorporating an on-road trial.

erschienen im Tagungsband der INFORMATIK 2011
Lecture Notes in Informatics, Band P192
ISBN 978-3-88579-286-4
weitere Artikel online:
http://informatik2011.de/519.html

References

[BLL10] Juliane Blechinger, Frank Lauterwald, and Richard Lenz. Supporting the Production of High-Quality Data in Concurrent Plant Engineering Using a MetaDataRepository. In *AMCIS 2010 Proceedings*, 2010.

[BS06] Carlo Batini and Monica Scannapieco. *Data Quality: Concepts, Methodologies and Techniques*. Springer, 2006.

[CD97] Surajit Chaudhuri and Umeshwar Dayal. An overview of data warehousing and OLAP technology. *SIGMOD Rec.*, 26:65–74, March 1997.

[FHM05] Michael Franklin, Alon Halevy, and David Maier. From databases to dataspaces: a new abstraction for information management. *SIGMOD Rec.*, 34:27–33, December 2005.

[HE09] W. Hellmann and S. Eble. *Gesundheitsnetzwerke managen - Kooperation erfolgreich steuern*. Medizinisch Wissenschaftliche Verlagsgesellschaft, 2009.

[JFH08] Shawn R. Jeffery, Michael J. Franklin, and Alon Y. Halevy. Pay-as-you-go user feedback for dataspace systems. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, SIGMOD '08, pages 847–860, New York, NY, USA, 2008. ACM.

[Leh96] M. Lehman. Laws of software evolution revisited. In Carlo Montangero, editor, *Software Process Technology*, volume 1149 of *Lecture Notes in Computer Science*, pages 108–124. Springer Berlin / Heidelberg, 1996. 10.1007/BFb0017737.

[Len09] Richard Lenz. Information Systems in Healthcare - State and Steps towards Sustainability. *IMIA Yearbook 2009*, 1:63–70, 2009.

[NMC+99] Prakash M. Nadkarni, Luis Marenco, Roland Chen, Emmanouil Skoufos, Gordon Shepherd, and Perry Miller. Organization of Heterogeneous Scientific Data Using the EAV/CR Representation. *Journal of the American Medical Informatics Association*, 6:478–493, 1999.

[Par94] David Lorge Parnas. Software aging. In *Proceedings of the 16th international conference on Software engineering*, ICSE '94, pages 279–287, Los Alamitos, CA, USA, 1994. IEEE Computer Society Press.

[RB01] Erhard Rahm and Philip A. Bernstein. A survey of approaches to automatic schema matching. *The VLDB Journal*, 10:334–350, 2001. 10.1007/s007780100057.

[SL90] Amit P. Sheth and James A. Larson. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Comput. Surv.*, 22:183–236, September 1990.

[SMB05] Monica Scannapieco, Paolo Missier, and Carlo Batini. Data Quality at a Glance. *Datenbank-Spektrum*, 14:6–14, 2005.

[Wan98] Richard Y. Wang. A product perspective on total data quality management. *Commun. ACM*, 41:58–65, February 1998.

erschienen im Tagungsband der INFORMATIK 2011
Lecture Notes in Informatics, Band P192
ISBN 978-3-88579-286-4
weitere Artikel online:
http://informatik2011.de/519.html