

ORIGINAL ARTICLE

Building a high quality medical data architecture for multiple uses in an integrated health care environment

Frank Boterenbrood, Irene Krediet, William Goossen

Windesheim University of Applied Science, Department ICT innovations in health care, Zwolle, Netherlands

Correspondence: Frank Boterenbrood. Address: Windesheim University of Applied Science, Department ICT innovations in health care, Campus 2-6, 8000 GB Zwolle, the Netherlands. E-mail: f.boterenbrood@windesheim.nl

Received: December 31, 2013
DOI: 10.5430/jha.v3n5p55

Accepted: March 25, 2014
URL: <http://dx.doi.org/10.5430/jha.v3n5p55>

Online Published: April 15, 2014

Abstract

Objective: The aim was to create a reliable information provisioning system in healthcare for both care and research processes, based on existing data standards and standardized electronic messages. The research question is: How can a Clinical Data Ware House (CDWH) be developed for standardized basic patient data, generic nursing data and data about oncology nursing, allowing management of Electronic Health Record data, electronic data exchange and data analytics?

Materials and methods: The main instrument used was the Detailed Clinical Model (DCM) data standardization approach. Further, data communication utilized HealthLevel7v3 (HL7v3) messages, transported by Mirth Connect. In an incremental, design-oriented research project, CDWH-prototypes were constructed using Enterprise Architect, a HL7v3 generator plug-in, SQL Server technology and PostgreSQL-based CDWH-technology.

Results: The project resulted in 16 existing DCMs selected and 6 new DCMs defined. From those DCMs, a HL7v3 message structure was generated and a CDWH created. Implementing specialized Data Marts (DM) a connection between the CDWH and one Electronic Health Record system was built.

Discussion: Detailed Clinical Models improve data quality, yet building them consumes time and resources. Some required data codes could not be identified in time and dummy codes were used instead. The existence of unstructured medical data in legacy systems may prove to be an obstacle in the future.

Conclusion: The project shows that using Detailed Clinical Models as the sole source for system development leads to a sound design for a CDWH and HL7v3 messages, supporting a standards based health information system, suitable for multiple uses.

Key words

Detailed clinical model, Electronic health record, Nursing informatics, Clinical data warehouse, HL7v3 care record

1 Introduction

1.1 The data

For decades, IT promised to solve all business problems once and for all. However, with each wave of technological innovation reality was becoming increasingly unmanageable and recently we have finally admitted that IT actually is

cloudy business. And now, in the post-Cloud-hype era, the focus is where it should have been all along: on the data. And we find ourselves confronted with a lot of it. Indeed, today we recognize that Big Data is our real issue.

In healthcare, the situation is not much different. Many years of innovation have led to a multiform IT portfolio, in which systems of inoperable technologies try to integrate in order to achieve a logical and integrated electronic patient record^[1, 2]. Even worse, healthcare is described by unparalleled complexity in which many medical specializations try to establish a coherent view on a patient's wellbeing^[3, 4]. Medical data reflects this complexity. By nature, medical data is often a complex collection of data elements, requiring special knowledge to be understood fully. It is highly sensitive data too, being specifically addressed as data that requires special attention^[5].

One may conclude that this special type of data is subject to special care. This is not entirely the case. Medical data is stored in many dispersed databases, duplicated in many cases and stored in a variety of formats, of which "free text" is not an uncommon solution^[1]. While Big Data may refer to the immense amount of data, in healthcare it may additionally be assigned to the immense complexity of data, in terms of data definitions, data sensitivity, quality requirements and availability.

1.2 Objective

In 2011, the research group IT-Innovations in Healthcare of Windesheim University, Zwolle, Netherlands joined forces with Isala Clinics Hospital Zwolle, Icare Homecare Meppel, Results4Care Amersfoort and Mgrid Amsterdam in a research project to find a way to create multiple views on reliable, high-quality medical data within the hospital, and using this data for exchange between partners and research, without the need to completely overhaul the current IT infrastructure^[6].

The two healthcare organizations participating in this project are Isala Clinics Hospital and Icare Homecare. Isala Clinics is the largest Top-clinical hospital in the Netherlands, with 5,300 employees and 1,000 beds. Per annum Isala Clinics treats more than 502,000 patients in outpatient visits, 47,000 patients via admissions and 48,000 in day care. Top clinical functions offered include heart surgery, neonatal intensive care, trauma center, neurosurgery and bone and stem cell transplants. At Isala Clinics, research results in 180 to 200 annual peer reviewed publications.

Icare Homecare is one of the brands of the Espria concern for home care, childcare, maternity care and mental health care, among others. Espria has 16,000 employees at full concern level. Espria offers living, welfare, care and social services. Icare Homecare offers home care, mainly via professional nurses. One specialty of Icare Homecare concerns the oncology visits in collaboration with Isala Clinics.

The aim of this project is to find a way to create a reliable information base for both healthcare processes and research, based on the fit for use idea: quality of medical data should meet requirements set by the medical processes. Furthermore, juridical requirements for data exchange with care partners and scientific research are investigated.

Aiming to provide high-quality medical data for multiple purposes, based on using a Clinical Data Ware House (CDWH) data integration strategy, the research questions for this project are the following:

- In the context of Isala Clinics, how can a CDWH aimed on the provisioning of basic patient data, generic nursing data and data about oncology nursing, allowing management of Electronic Health Record (EHR) data, be developed?
- Can the nursing data be standardized and mapped to specifications for different data uses as electronic exchange, quality indicators and clinical studies?
- Can we use existing standards to facilitate electronic exchange of nursing oncology care data?

1.3 Background

The current situation on data management in healthcare can have negative consequences on business efficiency and patient safety^[3, 7]. The patient is subjected to repeated scans, blood tests or other medical examinations. Misinterpreted or erroneous data may lead to erroneous decision making, putting the patients' health on risk. Therefore, achieving reliable views on data is paramount.

To create multiple views on standardized hospital data, a data integration strategy is required. In general, two distinctive strategies are available:

There are a variety of possible architectures for data integration, but broadly speaking, most systems fall somewhere on the spectrum between warehousing and virtual integration^[8].

Therefore, a choice has to be made. Will data be loaded into a centralized warehouse (or operational data store)^[9] or will it be made available through real-time integration services, directly accessing the data in the source databases^[10]?

Another issue to contemplate is the notion of semantic dissonance^[11]. This term describes a situation where data that appears to be the same may not necessarily mean the same thing^[11]. This again indicates processing issues when massive sets of data are requested from their dispersed source systems for research purposes. It can be argued that data needs to be made centrally available and the semantic dissonance issues solved before data is retrieved for various purposes, such as evaluation. Finally, the use of data for research implies the use of massive queries, which may result in performance conflicts in the source systems if those queries are executed in the source systems directly^[12]. Therefore, for this research, the use of an operational data store or datawarehouse as a central point of information is preferred. Since this datawarehouse is used in a clinical environment, it is referred to as a CDWH.

Semantic Dissonance is related to data quality. Data quality is defined by dimensions, and those dimensions in turn are measured by data quality attributes^[13-15]. Data Quality dimensions recognized in literature are Accessibility, Accountability, Accuracy, Completeness, Confidentiality, Consistency, Currency, Integrity, Reliability, Specification, Timeliness, Uniqueness and Volatility^[13-17].

Those data quality dimensions require meeting a certain threshold. To establish such a threshold the Data Management Association introduces the Data Quality Management function:

“Data Quality Management – Planning, implementation and control activities that apply quality management techniques to measure, assess, improve and ensure the fitness of data for use”^[18].

This definition points the way for the assignment of the right threshold: data should be fit for use. Fit for use relates to the ability of data to support business rules^[19].

In medical information systems, medical data is stored and processed in patient records. This data is personal and processing data about a person is subject to legislation^[5, 20]. Therefore, juridical requirements play an important role in processing medical data. In fact, according to article 16 of the Dutch Personal Data Protection Act, data describing one's health is considered to be special data, of which processing is prohibited except for specific institutions and for purposes identified by article 21^[5]. According to this, additional dimensions of data quality include^[5]:

- Transparency. The patient is to be informed about what data is processed and by whom.
- Purpose-specification. The purpose of data processing activities must match the purpose for data collection.
- Legitimate basis. Data processing activities are in line with business goals, patient approval and (inter)national legislation as supervised by the Dutch Data Protection Authority.
- Data Security. Shielding data from unauthorized use will be in line with available technical solutions.

2 Materials

The main instrument used in this research project is the Detailed Clinical Model (DCM). The DCM concept is initiated by Huff of Intermountain Healthcare^[21] and is currently standardized by ISO TS 13972: 2014^[22]. A DCM defines medical data in its clinical context and, by doing so, turns it into well-defined, fit-for-use information.

“A Detailed Clinical Model is a logical model designed to express one or more clinical concepts and their context in a standardized and reusable manner, specifying the requirements for clinical information as a discrete set of logical clinical data elements”^[22].

Structurally a DCM provides the purpose and evidence of a concept, data elements and attributes of the clinical concept, including the possible values and types of attributes, and relationships needed to convey the clinical reality in a fashion that is understandable to clinical domain experts, modelers and implementers^[23].

Please note that it is observed that many initiatives in numerous countries, such as work done by the National Cancer Institute (NCI) at the National Institute of Health (NIH) of the United States^[24, 25], use synonyms for clinical modelling, for instance clinical elements, (clinical) templates, care information models, clinical content models, archetypes, clinical fragments, general purpose information components, detailed clinical models^[23], continuity of care documents^[26].

Data is presented to the environment using services. In the case of healthcare, the communication model defined by HealthLevel7 (HL7) is the current standard applied. In a modern XML-based Service Oriented Architecture^[27] the XML-based implementation of HL7v3 seems to fit best. The HL7 working group “Patient Care” developed a HL7v3 message allowing clinical content to be exchanged between medical information systems^[28]. This Care Record message is based on clinical statements and is useful for sending structured nursing data between health care facilities for continuity of care^[28].

To transform a DCM into a HL7v3 message structure, the CASE-tool Enterprise Architect from Sparx systems, equipped with a specialized HL7v3 clinical statement generator plug-in by Results4Care, was used. A first CDWH-prototype was constructed using Microsoft’s SQL Server and Integration Services (SSIS) technologies. Later prototypes were developed using PostgreSQL-based CDWH-technology supplied by Mgrid. In order to transport messages, the open-source HL7 based integration engine Mirth Connect was used.

3 Methods

3.1 Introduction

In the previous paragraphs, a combination of a CDWH, CDMs and the HL7v3 Care Record is argued to be a solution fitting the needs of modern healthcare information integration and provisioning. Indeed, these elements have been used individually, however combining these building blocks into an integrated solution offers new uncertainties. Each building block calls for a specific approach, while combatting the combined uncertainties requires an agile, incremental project design.

3.2 Project design

Therefore, a multi-method approach has been adopted to answer the first research question. This approach is recognized as a valuable approach for understanding complex healthcare information processing issues^[29-31]. The complexity and uncertainty of the subject required an agile approach^[32, 33], combining both PRINCE2^[34] and Scrum^[35] project management methods. Implementing this in a design-oriented research, a functional prototype was constructed. This process covered the first four phases (First hunch, Requirements & assumptions, Structural specifications and Prototype)

of the design-cycle as proposed by Verschuuren and Hartog^[36]. In the end, the prototype was to position a CDWH as an intermediary data store, integrating information systems from both Isala Clinics and Icare using DCM-based HL7v3 messages (see Figure 1) and providing data for research and business process management (data analysis). In the context of this research, the prototype provided information to answer the second and third research question.

3.3 Project execution

The project started in September 2011 by creating an environment in which the research could take place. Over a period of two years, starting in February 2012, four successive teams of Healthcare students, IT and Business IT students and students of Law and Ethics from Windesheim University worked closely together in four incremental cycles, each successive cycle extending the results obtained in the previous cycle. Specialists from Windesheim, Isala Clinics and Icare provided guidance.

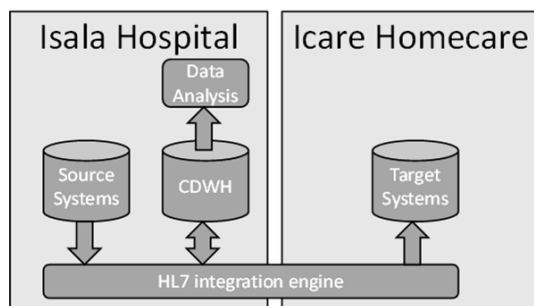


Figure 1. Integration architecture

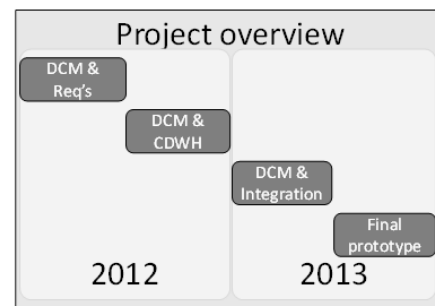


Figure 2. Project phases

As Figure 2 explains, incremental development cycles were used in this project. In the first increment, healthcare students identified and developed initial DCMs, using a list of required data provided by medical personnel.

Starting in September 2012, the second increment resulted in the first prototype being build, based on DCMs and requirements developed in the first increment, and DCMs identified and developed in the initial phase of this second increment. The CDWH exchanged information using DCM-based HL7v3 messages with a second, dummy information system.

The third increment of the project started in February 2013. The goal of this increment was to create a CDWH and, in a research-environment, connect this CDWH to the destination system at Icare Homecare.

The fourth and final increment of the project, connecting Isala sources systems to the CDWH, resulting in a final prototype and completing the proposed integration architecture has started in September 2013 and concluded in February 2014.

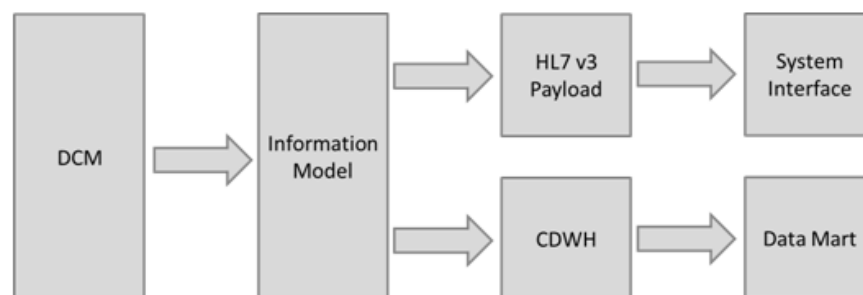


Figure 3. Results per project increment

In each project increment, a set of interdependent results were constructed (see Figure 3). For each result, the methods and techniques used are described in the next paragraphs.

3.4 DCM creation

Creating a DCM proved to be an iterative, discussion-rich activity. The healthcare students executed structured interviews, collecting data from medical specialists in the field of oncology and DCM design. DCMs created were reviewed with these specialists and discussed. In successive iterations, the DCMs were scored on their usability by the Quality Metrics for Detailed Clinical Models ^[37]. This measurement tool scores a DCM on 8 quality domains and 29 quality metrics. The score is an indicator for DCM quality, where a high score indicates high quality. In the end, for each DCM, a textual description was created.

During this first cycle of the project, juridical requirements were of particular interest. For the investigation of the legitimate basis for the two types of processing (exchange and research) a specific consent per type of data use was defined by a team of students of Law and Ethics. This definition was based on both a literature study combined with a limited survey amongst a group of thirty patients, establishing their views on the use of their medical data in research.

3.5 Information modelling

Once a DCM's definition was agreed upon, it became the foundation for the IT development part of the project. For IT development, a more linear approach was executed, transforming the DCM into a working IT solution.

Part of each DCM is an information model, defining the information objects that are included in the DCM. Using Enterprise Architect, medical information detailed by healthcare students is collected by IT students into packages, and transformed into an UML Class Diagram. Each class in this diagram is defined as a specific data type, based on the ISO standard 21090. Value-ranges for specific data types are defined by standards like Snomed-CT. The root of the Class Diagram links the class to the DCM.

3.6 HL7 v3 payload generation

The completed class diagram became the foundation for the creation of the CDWH and the HL7v3 message payload definition.

The payload definition was automatically generated introducing model-driven development into the project. This generation process made extensive use of the data type and value range definition. Figure 4 presents an example information model (DCM Body Weight) and a snippet of the corresponding generated HL7v3 Care Record message definition.

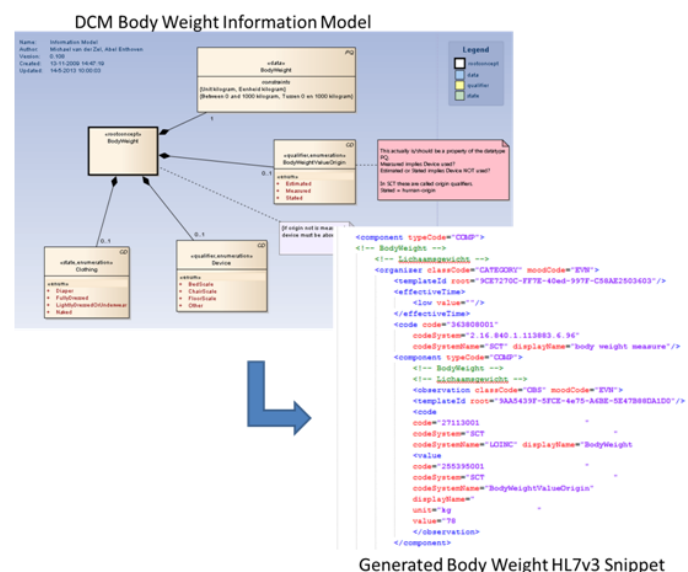


Figure 4. DCM Body Weight information model and Generated HL7v3 Care Record snippet

3.7 System interface construction

Based on the generated HL7v3 payload and the technologies and information systems involved, system interfaces were constructed supporting the ETL (Extract, Transform and Load) process. The construction of these interfaces started with a functional and technical design. It was found that the software has to have three groups of functionalities: it has to validate an incoming HL7v3 message, it has to traverse through it to create a data structure and it has to generate a valid input/output format for the applications involved. Therefore the software was divided into three packages: Conversion, Parsing and Validation. Interfaces were build using the C programming language.

3.8 CDWH and datamart construction

A CDWH was constructed based on the CDWH architecture proposed by Inmon^[38]:

Inmon's philosophy recommends starting with building a large centralized enterprise-wide data warehouse, followed by several satellite databases to serve the analytical needs of departments (later known as "data marts"). Hence, his approach has received the "Top Down" title^[38].

For the CDWH, the database structure was based on the HL7 Reference Information Model^[39]. Using this model as the main database structure, any HL7v3 message can be handled by the CDWH.

For each data use (research, data exchange, management information) a separate datamart was constructed. Each datamart was formatted using a star-scheme datamodel and populated by dedicated SQL-procedures.

4 Results

The first result was a General Patients Questionnaire comprising information required for transferring oncology patients from hospital to homecare.

Table 1. Integrated patient questionnaire

Used separate transfer lists	New designed transfer list
Care Professional data (HL7 CMET)	Integrated Patient Questionnaire
Person data (HL7 CMET)	
Patient administrative data (HL7 CMET)	
e-Transfer list V&VN/Nictiz	
Transfer list Isala / Icare	

As shown in Table 1, the questionnaire was an integrated version of five lists originating from HL7v3 CMETS's person and patient, from data of a V&VN/NICTIZ project e-Transfer and a list designed by Icare and Isala.

Next, DCMs that are linked to the required data in the Integrated Patient Questionnaire were selected (see Table 2). Available DCMs were selected from earlier projects. Placed in the context of the transfer of patients and their data from hospital to homecare, they were adjusted using a change request where required. A review by students, oncology specialists and homecare nurses resulted in 16 DCMs available for incorporation into the CDWH.

In addition, the healthcare students, oncology specialists and homecare nurses, created six new DCMs. These DCMs were designed according to the national DCM guideline for DCM projects in the Netherlands^[40]. DCMs for the TNM classification of oesophagus carcinoma and of stomach carcinoma were designed. Following the two DCMs on cancer classification, a new DCM was created for tube feeding at home for patients with a jejunostomy after an oesophageal resection.

Table 2. List of available and new oncology related DCMs

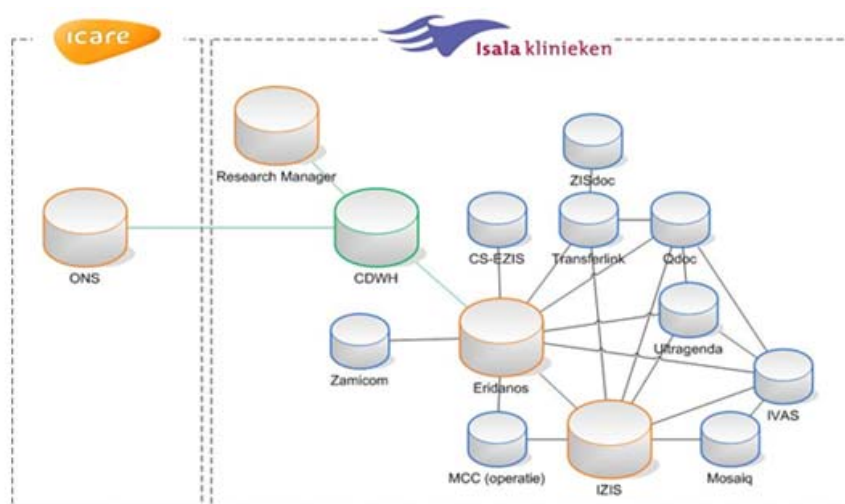
DCMs related to the General Questionnaire		New DCMs first and second increment
Personal information*	Will statement	Tumor Node Metastasis (TNM) esophagus carcinoma*
Social information*	Nutrition*	Tumor Node Metastasis (TNM) stomach carcinoma*
Pain*	Allergic reactions	Tube feeding for jejunostomy after esophageal reconstruction*
		<u>New DCMs third increment</u>
Barthel-index	Decubitus treatment	Malnutrition after oesophageal Resection with cardiogastric tube reconstruction
Usage of Appliances	Length	Postoperative wound infection after oesophageal Resection with cardiogastric tube reconstruction
Defecation	Sensitiveness	Repeated pain measurement
Diabetes Diagnose	Decubitus description	
Weight*	Referrals*	

Note. DCMs labeled with an * were scored on their usability by the Quality Metrics for Detailed Clinical Models ^[37]

DCMs are not used for clinical purposes only, they are also used for quality reporting and clinical research. Therefore, three new additional DCMs were developed which focus in particular on the quality of care of patients: malnutrition after oesophageal resection with cardiogastric reconstruction, postoperative wound infection after oesophageal resection with cardiogastric reconstruction and a DCM on repeated pain measurement.

The DCM Quality Measures revealed that only two TNM classification DCMs can be used immediately, obtaining 26 points out of 29. Based on the quality score, all the other DCMs related to this project scored between 16-24 points, and therefore required some adjustments before becoming usable.

As part of this research the analysis of legal options was carried out. Acquiring approval of patients for both data exchange and research on CDWH data should include an explanation of the (implications of the) data processed. Medical data is so sensitive that in addition to creating a legitimate basis, data used for scientific research should be made anonymous by removing all personal identifications. However, data transferred to the hospital's care partners explicitly require those identifications, in order to be able to identify the right person in the homecare processes. This conflict resulted in the decision being made to develop the CDWH along the lines set by Inmon ^[38]. This architecture for CDWH presents scalability and at the same time allows for development of specific data marts (DMs) for specific stakeholders, meeting also the requirements defined by the juridical and ethical aspects. For each purpose of data-use one specialized data mart was created, meeting specific data and legal requirements, in some cases stripping patient identifiers from the dataset.

**Figure 5.** Integrating information systems

The first cycle concluded with the identification of the application and communication architecture involved. At the application side, it is found that there is not one single system that can act as a source system for the data defined by the DCMs. Please note that Figure 5, representing a subset of Isala Clinics' application landscape, pinpoints a system called Eridanos as a central hub system. Having a central information provisioning system already in place might help greatly in disclosing the data required. However, it was discovered that the data in Eridanos was stored in unstructured documents mainly, rendering it useless for extracting useful data automatically. Eridanos cannot fulfill requirements for Accuracy, Completeness, Consistency and Reliability. The big data approach often is envisioned as promising, mending these data dimensions, but well-structured data will always carry the precision for continuity of care and medical research without the need for an additional repair effort. Therefore, data is to be acquired from the source systems directly and the intended use of Eridanos as a central source for data was dropped.

During consecutive cycles of the project, it was found that, after DCMs were scored on their usability by the Quality Metrics ^[37] and adjusted, using an automated generation process turning DCMs into Information Models and generating a HL7v3 payload from this model, improved the quality of the DCMs even further. The Results4care EA plug-in contains multiple checks and validations, rejecting DCMs that do not meet fine-grained criteria. However, these checks threatened the project to grind to a halt. It was found that in cases, where HL7v3 requires unique codes to be in place, these codes could not be identified in time and (temporary) dummy codes were used instead.

In order to interface with Icare's main information system (ONS), a bespoke interface was programmed. It was found that this target system was not able to store DCM data directly. To overcome this obstacle, free fields extending the target systems' database functionality were utilized.

It is observed that DCMs tend to evolve continuously. Therefore, creating a coherent set of a CDWH, datamarts, message structures and external interfaces requires release management, binding versions of DCMs, data stored and software together.

Based on these results, the research questions can be answered:

In the context of Isala Clinics, how can a CDWH aimed on the provisioning of basic patient data, generic nursing data and data about oncology nursing, allowing management of EHR data, be developed?

In this project, an agile approach on combatting uncertainty proved valuable. Detailed Clinical Models were successfully transformed to UML specifications. This transformation resulted in an increase in quality of the DCMs involved. Based on these specifications, HL7v3 Care Record payload segments and CDWH database definitions can be developed. It was found that Inmon's top down philosophy on data warehouses suited the hospital's complexity best. In this process, healthcare students developing DCMs and IT student creating an information provisioning infrastructure proved that business and IT specialists, working together, are able to create a rich and future-proof infrastructure.

Can the nursing data be standardized and mapped to specifications for different data uses as electronic exchange, quality indicators and clinical studies?

During this project, it was found that data can be made available for different purposes and audiences by creating a datamart for each purpose. Each datamart can be adapted to the needs and dynamics of its specific audience, whilst incorporating anonymizing technologies where privacy law required protection of sensitive data. However, processing sensitive data may require extra measures, like obtaining a formal agreement from the patients concerned.

Can we use existing standards to facilitate electronic exchange of nursing oncology care data?

The research showed that newly designed elements like DCMs, CDWH, datamarts and HLv3 Care Record messages can be tightly bound together, in some steps using generating technologies. Strict version management is required to keep track of the volatility in DCM definitions. In some cases, discussions on DCM definitions required dummy-codes to be used in order to enable the development process to proceed. Integrating these elements into existing information systems proved to be a challenge. In the target system at Icare, free format fields were used to store the information, and at Isala, an important source system showed a habit of storing data in an unstructured manner, rendering it useless for automated processing. The problem of unstructured data in source systems may prove to be an obstacle for the effective use of medical data in the future.

5 Discussion

The use of DCM in itself may not be a new concept^[26]. Yet, although this project is based on existing methods, the combination of those, using DCM as a single starting point for the development of a CDWH and HL7v3 based data exchange, results in a new approach. There are alternatives with respect to the precise formalism used, *e.g.*^[41]. However, on the methodology level, it is the similarity in the combination of methods which results in valuable cross validation.

The juridical requirement that data used for research should be made anonymous has implications for the design of the data mart as part of the CDWH. The CDWH is designed as an extension to the EHR, including patient identifications. For each purpose of data-use one specialized data mart is created, meeting specific data and legal requirements, in some cases stripping patient identifiers from the dataset. Additional privacy enhancing technologies may be required and added.

The creation and validation of DCM consumes time and resources. However, being international standards for the definition of medical data on a semantic level, a DCM contributes greatly to data quality dimensions like Accountability, Accuracy, Completeness, Consistency, Reliability, and Specification. Using HL7-based real-time messaging supports both Currency and Timeliness of healthcare data. It was noticed that in cases where HL7v3 requires codes to be in place, these codes could not be identified in time and dummy codes were used instead, potentially leading to errors in the future. Code requests will be submitted to relevant standards bodies.

During the execution of the project, it was found that a central information system stored data in an unstructured format. Data stored in unstructured format seems to be common practice in healthcare, and may suggest that the data quality dimension Specification is missing for many healthcare business objects. Unstructured data may prove to be an obstacle for the effective use of medical data in the future.

Using DCM as a single source for the design of CDWH and HL7v3 messages results in a coherent information provisioning system for multiple stakeholders. However, existing information systems may not align with DCM definitions, forcing less elegant solutions to be implemented at the interfaces between legacy and CDWH, whilst time and resource related issues might lure projects into taking shortcuts causing inconsistencies again. The solution applied is that DCMs form the core of the CDWH and datamarts, keeping the source of truth consistent.

6 Conclusion

We conclude this work as follows:

The idea of the project to standardize oncology data for transfer by means of DCM has been successful as 6 new DCMs have been designed, and an additional 16 adjusted DCMs applied. Also the Quality Metrics tool was used to improve the DCMs. Relevant DCMs, related to the Integrated Patient Questionnaire, are useful for the transfer of oncology patients from hospital to homecare and vice versa.

As an instrument to provide high quality medical data, a CDWH proves to be an invaluable instrument for various kinds of data use and reuse. For a CDWH the architecture promoted by Inmon^[38] fits best. This architecture presents scalability and at the same time allows for development of specific datamarts for specific goals, stakeholders, satisfying confidentiality requirements, whilst supporting Integrity, Reliability and Uniqueness of the data-elements included. The project shows that the approach, where DCMs are the sole source for system development, leads to a reliable and sound design for an information provision system in healthcare, and allows incremental development via adding DCMs. However, it is found too that the dynamics of creating a DCM ask for a strict release management, making sure that versions of DCM, data, and software are firmly bound together. Integration of this approach in an existing IT landscape presents challenges in the field of quality of available data and application compatibility.

Competing interests

The authors declare that they have no competing interests.

References

- [1] Dick RS, Steen EB. The Computer-Based Patient Record, Washington D.C.: National Academy Press. 1991.
- [2] Shortliffe EH, Cimino JJ. Biomedical Informatics: Computer Applications in Health Care and Biomedicine, New York: Springer. 2013.
- [3] Knaus W, Draper E, Wagner D, *et al.* An evaluation of outcome from intensive care in major medical centers. *Annals of Internal Medicine*. 1986; 104: 410-418. PMID: 3946981. <http://dx.doi.org/10.7326/0003-4819-104-3-410>
- [4] Winter A, Ammenwerth E, Bott O, *et al.* Strategic information management plans: the basis for systematic information management in hospitals *International Journal of Medical Informatics*. 2001; 64: 99-109. [http://dx.doi.org/10.1016/S1386-5056\(01\)00219-2](http://dx.doi.org/10.1016/S1386-5056(01)00219-2)
- [5] Data Protection Law [Wet Bescherming Persoonsgegevens WBP], 2000. [Internet]. Available from: http://wetten.overheid.nl/BWBR0011468/geldigheidsdatum_04-10-2012. [Accessed 05 17 2013].
- [6] Goossen W, Boterenbrood F, Krediet I. eTelemed conference, in Exchanging Nursing Oncology Care Data With Use of a Clinical Data Ware House, Nice, France. February 2013.
- [7] Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human, Building a Safer Health System, Washington D.C.: National Academy Press. 1999.
- [8] Doan A, Halevy A, Ives Z. Principles of Data Integration, Waltham, MA, USA: Elsevier. 2012.
- [9] Bobak AR. Connecting the data, Westfield, NJ 07090 U.S.A.: Technics Publications. 2012.
- [10] Rotem-Gal-Oz A. SOA Patterns, Shelter Island: Manning Publications. 2012.
- [11] Microsoft Corporation. Integration Patterns, Microsoft Press. 2004.
- [12] Bellika J, Sue H, Bird I, *et al.* Properties of a federated epidemiology query system, *International Journal of Medical Informatics*. 2007; 664-676. PMID: 16949338. <http://dx.doi.org/10.1016/j.ijmedinf.2006.05.040>
- [13] Loshin D. Master Data Management, Burlington: Morgan Kaufmann OMG Press. 2008.
- [14] Batini C, Scannapieco M. Data Quality, Concepts, Methodologies and Techniques, New York: Springer Berlin Heidelberg. 1998.
- [15] McGilvray D. Executing Data Quality Projects, Burlington, MA: Elsevier, Inc. 2008.
- [16] Lee YW, Pipino LL, Funk JD, *et al.* Journey to Data Quality, Cambridge, Massachusetts: The MIT Press. 2006.
- [17] Schumacher M, Fernandez-Buglioni E, Hybertson D, *et al.* Security Patterns, Integrating Security and System Engineering, Chichester: John Wiley & Sons Ltd. 2006.
- [18] Mosley M. DMBOK: Data Management Body of Knowledge. 2008. [Internet]. Available from: www.dama.org. [Accessed 07 05 2013].
- [19] Besouw F. Samenhang tussen bedrijfsregels, bedrijfsprocessen en gegevenskwaliteit, *EGO Magazine voor Informatiemanagement*. June 2008; vol. 7, third edition, 59-62.
- [20] European Data Protection Directive, EUROPEAN DATA PROTECTION DIRECTIVE, 1995. [Internet]. Available from: http://www.dutchdpa.nl/Pages/en_ind_wetten_eu.aspx. [Accessed 05 17 2013].
- [21] Rocha R, Coyle J, *et al.* Integrating detailed clinical models into application development tools, *Medinfo*. 2004; 11(Pt 2): 1058-62.
- [22] ISO TS 13972: 2014. Health informatics - Detailed Clinical Models, characteristics and processes. Geneva, ISO.
- [23] Goossen W, Goossen-Baremans A, van der Zel M. Detailed clinical Models: a review, *Health Inform Res*. 2010; 16(4): 201-214. PMID: 21818440. <http://dx.doi.org/10.4258/hir.2010.16.4.201>

- [24] National Cancer Institute. Cancer Biomedical Informatics Grid. Approved caBIG-NCI Data Standards. 01 05 2013. [Internet]. Available from:
<https://wiki.nci.nih.gov/display/caDSR/caBIG-NCI+Data+Standards#caBIG-NCIDataStandards-ApprovedcaBIG%C2%AE-NCI+DataStandards>. [Accessed 15 05 2013].
- [25] National Cancer Institute. Cancer Data Standards Registry (caDSR), [Internet]. Available from:
<https://wiki.nci.nih.gov/display/caDSR/caDSR+Wiki>. [Accessed 15 05 2013].
- [26] Ferranti JM, Musser RC, Kawamoto K, *et al*. The Clinical Document Architecture and the Continuity of Care Record: A Critical Analysis, *Journal of the American Medical Informatics Association*. May-Jun 2006; 13(3): 245-252. PMID: 16501180.
<http://dx.doi.org/10.1197/jamia.M1963>
- [27] Josuttis NM. SOA in Practice. The Art of Distributed System Design, Sebastopol: O'Reilly Media, Inc. 2007.
- [28] Goossen W. Sending electronic nursing discharge messages using the HL7 v3 Care Provision standard, *Stud Health Technol Inform*. 2009; 146: 269-275. PMID: 19592847.
- [29] Woolery L. Expert Nurses and Expert Systems. Research and Development issues, *Computerts in Nursing*. 1990; 8(1): 23-28. PMID: 2302590.
- [30] Goossen W. Nursing Information Management and Processing: a framework and definition for system analysis, design and evaluation. *International Journal of Biomedical Computing*. 1996; 40: 187-195. [http://dx.doi.org/10.1016/0020-7101\(95\)01144-7](http://dx.doi.org/10.1016/0020-7101(95)01144-7)
- [31] Westbrook J.I., Braithwaite, Georgiou A., Ampt A., Creswick N., Coiera E., *et al*. Multimethod Evaluation of Information and Communication Technologies in Health in the Context of Wicked Problems and Sociotechnical Theory. *J Am Med Inform Assoc*. Nov-Dec 2007; 14(6): 746-755. PMID: 17712083. <http://dx.doi.org/10.1197/jamia.M2462>
- [32] Highsmith J., Highsmith J. A., Eastlake D. Agile Software Development Ecosystems, Boston: Pearson Education. 2002.
- [33] Howson C. Successful Business Intelligence: Unlock the Value of BI & Big Data, Second Edition, New York: McGraw-Hill Osborne Media. 2013.
- [34] Hedeman B., Fredriksz H., v. Heemst G.V. Project Management based on PRINCE 2, Zaltbommel: Haren Publishing. 2006.
- [35] Rubin K. S. Essential Scrum, Boston: Pearson Education. 2012.
- [36] Verschuuren P., Hartog R. Evaluation in Design-Oriented Research, *Quality & Quantity*. 2005; 39(6): 733-762.
<http://dx.doi.org/10.1007/s11135-005-3150-6>
- [37] Anupindi NV. Inmon vs. Kimball - An Analysis, 25 08 2005. [Online]. Available from:
<http://www.nagesh.com/publications/89-information-technology-articles/173-inmon-vs-kimball-an-analysis.html>. [Accessed 17 05 2013].
- [38] Ahn S, Huff S, Kim Y, *et al*. Quality metrics for detailed clinical models, *International Journal of Medical Informatics*. 2013 May; 82(5): 408-17. PMID: 23089521. <http://dx.doi.org/10.1016/j.ijmedinf.2012.09.006>
- [39] Health Level Seven International. HL7 Version 3: Reference Information Model (RIM). Health Level Seven International. 2014. [Online]. Available from: <http://www.hl7.org>. [Accessed 18 03 2014].
- [40] Goossen-Baremans A, v. d. Zel M, Goossen W, *et al*. Guideline Detailed Clinical Model [Richtlijn Detailed Clinical Model], 02 2011. [Online]. Available from: <http://www.parelsnoer.org/>. [Accessed 29 05 2013].
- [41] Tao C, Jiang G, Oniki T, *et al*. A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data, *Journal of the American Medical Informatics Association*. 2013; 20(3): 554-62. PMID: 23268487.
<http://dx.doi.org/10.1136/amiajnl-2012-001326>