EJBI 2017 ISSN 1801 - 5603

European Journal for Biomedical Informatics

Volume 13 (2017), Issue 1

Special Topic
Re-Shaping Healthcare System

Editors
Bernd Blobel, Alexander Berler

www.ejbi.org

An Official Journal of the European Federation for Medical Informatics

An Official Journal of:

EFMI EUROPEAN FEDERATION for MEDICAL INFORMATICS

Aims and Scope

European Journal for Biomedical Informatics is an online journal publishing submissions in English language, which is devoted towards the progress in biomedical science. The journal welcomes all types of research communications through its rational open access window to maximize the possibilities of global reach for each article. The frequency of issue release is bi-annual, and strictly supported by peer-review process. The primary aim of the journal is to make all the informations freely available irrespective of different socio-economic barriers, which further will help in the progress of science. Moreover, the journal intends to showcase all the timely research information on biomedical informatics and its related studies for the effective contribution in the field of biomedical methods and technologies.

European Journal for Biomedical Informatics has specific instructions and guidelines for submitting articles. Those instructions and guidelines are readily available on the instruction for authors. Please read and review them carefully. Articles that are not submitted in accordance with our instructions and guidelines are more likely to be rejected.

European Journal for Biomedical Informatics accepts manuscript submissions through a mail attachment: editor@ejbi.org

Author Guidelines

This journal follows the guidelines of the International Committee of Medical Journal Editors (www.icmje.org) and the Committee on Publication Ethics (www.publicationethics.org).

Scope of the Journal

European Journal for Biomedical Informatics is dealing with a range of topics which is included under the scopes of biomedical informatics. The focus of the journal is on the major aspects of implementation of information science in Medical sciences, which includes Biomedicine, Ontology, Genomics, Biometrics, Biostatistics, Health informatics, Medical informatics, Clinical informatics, Semantics Interoperability, Informatics applications in Imaging.

Contents

Re-shaping Healthcare Systems

Bernd Blobel[1,2,3] and Alexander Berler[4,5]

[1] Medical Faculty, University of Regensburg, Germany

[2] eHealth Competence Center Bavaria, Deggendorf Institute of Technology, Germany

[3] First Medical Faculty, Charles University Prague, Czech Republic

[4] HL7 Hellas, Athens, Greece

[5] Gnomon Informatics SA, Thessaloniki, Greece

Correspondence to:

Prof. Dr. habil. Bernd Blobel, FACMI, FACHI, FHL7, FEFMI, MIAHSI
Medical Faculty, University of Regensburg, Germany
E-mail: bernd.blobel@klinik.uni-regensburg.de

EJBI 2017; 13(1):1-2
published: October 10, 2017

This Special Issue of the International Journal on Biomedicine and Healthcare is dedicated to the International HL7 Interoperability Conference (IHIC 2017) "Re-shaping healthcare systems", 22-24 October 2017 in Athens, Greece (http://www.ihic2017.eu). It contains papers selected by an independent peer review process, strictly performed by experts from countries different from the authors' country of residence.

IHIC 2017 is the 17th event of the International HL7 Interoperability Conference series, which has been inaugurated in 2000 by the Board of HL7 Germany and its unforgettable Chair and interoperability pioneer Joachim W. Dudeck. The first event in Dresden, Germany, was entitled "Advanced Healthcare Information Standards". While the first conferences have been characterized by focusing on CDA, over the time, the scope of the conferences has been extended towards all aspects of health information interoperability. The concept of interoperability has dramatically changed from standardized electronic data interchange (EDI) based on data representation at application level, the 7th level of the ISO Open Systems Interconnection stack, having been the name giver for the Health Level 7 standards framework. Meanwhile, the semantics of shared data as well as service level interoperability, but also domain-specific issues and even social aspects are considered, bringing terminologies and ontologies, but also implementation and conformance challenges on board. The relations to IHE and the FHIR success are especially highlighted at IHIC 2017. So it is just consequent to address also in 2017 both technological and non-technological issues of interoperability.

IHIC 2017 has been structured into four sections: a) Paradigm Changes in Healthcare and Resulting Interoperability Challenges, b) Infrastructure and Services for Healthcare Transformation, c) CDA-related Contributions, and d) Other issues. The thematic sections have been introduced by related Keynotes. The papers published in this IJBH Special Issue address the different aspects of the interoperability challenge from a theoretical and methodological perspective, usability requirements, professional groups' preferences, process design, semantical ambiguity, and implementation details..

In his Keynote "A New World for Better Health", Edward Hammond, Duke University (USA), addresses the ongoing healthcare systems' paradigm changes towards accountable care, a patient-centered focus, extensive data sharing, multiple data types, intensive collaboration, open source environments, and mobile technologies to better care an aging society. Thereby, he also highlights related evolving standards such as Fast Healthcare Interoperability Resources (FHIR). Bernd Blobel from the University of Regensburg (Germany), who addressed the aforementioned paradigm changes in health systems over many years already, discusses in his this year's Keynote "Standardization for Mastering Healthcare Transformation - Challenges and Solutions" the interoperability between existing and emerging standards and specifications. Based on a system-theoretical, architecture-centric, ontology-based approach to interoperability, he presents a meanwhile at ISO and CEN standardized Interoperability Reference Architecture Model. He demonstrates the feasibility of that model for cross-domain, i.e. multi-disciplinary, interoperability for integrating so different standards like ISO 13606 "EHR communication", ISO 22600 "Privilege management and access control" and ISO 21298 "Structural and functional roles".

In the first section on Paradigm Changes in Healthcare and Resulting Interoperability Challenges, Fabrizio Pecoraro and colleagues from the National Research Council, Rome (Italy) describe the Health at Home project to provide health and social care at home, using existing standards for enabling a cross-domain integration and continuity of care perspective over distance. The ContSys standard is used to harmonize the presentation of the involved concepts, and FHIR to implement the solution for information exchange. As security and privacy services including the resulting among the stakeholders as well between them and the patient are crucial for successful health services, Alexander Mense and Bernd Blobel discuss standards and component to support the implementation of the European General Data Protection Regulation (GDPR).

In the second section on Infrastructure and Services for Healthcare Transformation, Junquiao Chen et al. report about a U.S. population health partnership, exchanging and reconciling clinical decision support outcome among a cross-functional, distributed care team. CDS output is automatically harmonized between an EHR system and a population health services platform. That way, care gaps could be identified and overcome, so improving quality of care. Roberta Gazzarata et al. present a regional HIE solution for the Veneto region in North Italy to share clinical data. Thereby, a health terminology service based on CTS2 – a specification jointly developed by HL7 and the Object Management Group (OMG) – for correctly managing semantics is deployed to exchange laboratory reports encoded in LOINC and implemented in CDA documents. Dimitrios Katehakis et al. demonstrate the use of a Personal Health Record (PHR) system for managing chronic diseases, home care and wellbeing. For enabling interoperability between all actors involved, the HL7 PHR Functional Model and Meaningful Use criteria are applied. Finally, Giorgia Gazzarata et al. present a smart solution for access control in distributed EHR systems. That way, trustworthy communication and cooperation in a regional cancer care network will be enabled. Thereby, different access control models and different maturity levels of security and privacy services can be harmonized and consistently managed. The solution is based on a system-theoretical, architecture-centric, ontology –driven solution for policy mapping is deployed.

In the third section on CDA-related Contributions, Elisabeth Pantazoglou et al. from the Niederrhein University of Applied Sciences, Krefeld (Germany) describe a new project for communicating genomic-related medical data. Based on a process analysis, the concepts deployed will be represented by semantic annotation in an international terminology and thereafter implemented using HL7 CDA. Thus, the presented GENeALYSE project will support documentation and communication between diagnostics, medical therapy and research in the field of genome analysis for oncological diseases. Abderrazek Boufahja and Eric Poiseau from IHE Europe, Paris (France) investigate richness of CDA R2 documents and validate the reliability of related test and certification programs. Based on the investigation of existing methods and tools, the authors defined relevant test data, analyzed template inheritance and template containment to derive matrixes for template richness template containment. From that, a richness score of CDA R2 documents as the number of present CDA templates and components related to the number of possible templates and elements is calculated. The results have been demonstrated by assessing the richness of existing CDA implementation guides and approved CDA documents. Fabrizio Pecoraro et al. provide a solution for extracting data from CDA documents, transforming them properly and loading them into a Data Warehouse for secondary use. That way, decision-making for patient safety, healthcare quality assessment, clinical and translational research including clinical trials, comparative analysis of therapy pathways and best practices application can be supported. For this purpose, CDA schemas have to be transformed into a data warehouse dimensional model by mapping the CDA schema primitives with concepts of the dimensional model.

In the final section "Other issues", Robert Snelick from the National Institute of Standards and Technology (NIST), (US), presents a platform for developing specifications, test plans and testing tools to improve conformance and interoperability of solutions. In that context, a set of tools has been presented to define and to constrain HL7 v2 specifications including the development of test plans and test tool building blocks. The offered platform allows the automatic creation of conformance testing tools from those building blocks, that way enabling an end-to-end process improvement.

Additionally to the papers presented here, practice reports and implementation experiences have been shared at the conference. IHIC 2017 was completed by two Tutorial Days covering topics such as CDA, FHIR, but also security and privacy issues, SNOMED-CT and its relations to the TermInfo project as well as IHE in practice.

The Editors wish all interested parties an enjoyable reading.

The Guest-Editors are indebted to thank all authors and reviewers for their excellent work. Finally, they thank HL7 International for sponsoring the event, HL7 International HL7 and Germany for financing the Joachim W. Dudeck Award as well as HL7 Germany and HL7 Hellas for monetarily supporting the Tutorials.

A New World for Better Health

William Edward Hammond[1,2]

[1] Duke Center for Health Informatics, Duke Medicine, Durham, North Carolina, USA

[2] Clinical & Translational Science Institute, Duke Medicine, Durham, North Carolina, USA

Abstract

Issue: The healthcare industry is undergoing an exponential change. Our attempts to deal with these changes are inadequate, unimaginative, untimely, and too narrowly focused.

Goal: The goal of this paper is to increase the awareness of the technical community – particularly the standard developers and the implementers – about what is happening.

Methods: The paper identifies a number of changes that are occurring and discusses consequences that may result from these changes. Many of these potential changes are recognized through observations relating to new and developing technologies, through the new roles for the set of involved people, and for recognized problems with current systems. The major funded research initiatives were used to predict changes as were the directions a number of industry giants are moving, such as, Apple, Google, IBM, and a number of start-ups. Certainly the hype for Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing is moving toward reality.

Conclusions: Pick your time frame – from 2020 to 2050 - these changes are very likely to occur. Healthcare, as we know it, will change. Computers and robots will replace humans for many of the tasks now done by humans. Population Health will engage all sectors of the world around us – individuals, families, neighborhoods, communities, regions, and countries. Schools, churches, and social groups will become important in health and well-being. Patients will play a major role in managing their own health and healthcare. Becoming over age 60 will not mean waiting to die. New types of data other than just clinical will be just as important as clinical data. Genetic data will be available about everyone and will be used to determine risks.

Keywords

Healthcare paradigm changes; Big data; Interoperability; Standards; Population health

Correspondence to:

William Edward Hammond, PhD, FACMI, FAIMBE, FIMIA, FHL7
Duke Center for Health Informatics, NC 27705, USA
E-mail: william.hammond@duke.edu

EJBI 2017; 13(1):3-8
received: July 14, 2017
accepted: August 30, 2017
published: October 10, 2017

1 Introduction

The world we live in today changes at an unbelievable pace. Change is the constant in our world today. Over the next few years, health and healthcare will change in almost every aspect. The advent of mobile devices, wearable sensors, patient reported data, augmented reality, virtual home visits, machine learning, artificial intelligence, data sharing, geospatial coding, and aspects of Big Data will totally change the way, the place, and the how of medical care. Roles of healthcare providers – doctors and nurses - will change. The way we collect data will change. The types of data – not just clinical but social, economic, environmental, behavioral, and genetic data – will change how decisions are made and treatments are determined. Aggregation of data not only for individuals but for regions, states and provinces, countries, and globally will change and produce new problems. Initiatives in the United States (US) include Population Health, Precision Medicine, Big Data to Knowledge, Health Information Exchanges, Health Analytics and others reflecting these changes. Translational medicine will increase the use of data for multiple purposes [1].

The amount of data collected per person will be in the petabyte range. Input of new data will trigger thousands of clinical decision making algorithms. The required speed of access and navigation to specific data, for speed in calculations, and for effectively presenting decisions and results will require new architectural infrastructures and new concepts in databases. As a consequence of these changes in health and healthcare, the increase in life expectancy at birth has increased over my lifetime by almost 17 years. At my current age, my expected remaining years of life continues to rise from its current prediction of 8.1 years. Technology

has contributed a significant percentage to that increase. This longer lifespan has raised significant issues in health and healthcare affecting cost, social issues, crowding, and really just what to do with old people. People retiring at the age of 65 now expect to have, on average, an additional 20 years of life. Hence the problem. Factors impacting health include multiple chronic diseases, out-living retirement income, disruptive families, boredom, and overcrowding [2].

These new directions have introduced many new "words" and new ways of thinking about the future. Interoperability and standards that will enable interoperability, global standardization of semantics, cloud computing, integrated care plans, block-chain, quality, provenance, privacy, consumer engagement, socializing, and disruptive innovation are a necessity for success. Modifiers to health now include digital, tele, and virtual.

This paper discusses some of those changes and addresses how HL7 International and the expanded Informatics community needs to address these changes. First, from my observation point, we seem to spend most of our time in working around a problem rather than addressing the problem directly. Second, we seem to address most problems with solutions that are already out of date. Third, we address each problem as a single, isolated problem rather than dealing with the problem in a broader setting. Fourth, when we start to solve a problem, we start with what we know and have available, rather than looking at the best solution then looking to see what exists to solve the problem. Fifth, we ignore the hard problems. Six, they frequently say, "You can't do that because …" or "That is impossible". We need to design systems for tomorrow and not systems for yesterday.

This paper will discuss a number of factors that are consequences of the changing world of health care. The most significant changes are a result of advances in technology. Those changes are reflected in the continued validation of Moore's Law. Computers are faster, smaller, and more powerful than ever. Costs are way down. The advent of smart phones has had one of the biggest impacts on society since the invention of cars and airplanes. Now, mobile devices and wearable sensors have changed our views of data – how we collect it, and how we use it. Real time, global communications have introduced problems in dealing with world-wide epidemics.

2 A Changing World

2.1 Change from Sick Care to Health

One of the most significant changes in the health care system is the reimbursement focus change of fee-for-service to accountable care. With fee for service, healthcare institutions attracted the sickest patients because these resulted in the highest revenues. With accountable care organization, the healthiest patients result in the most profit, and sick patients cost money. Hospitals dominated the health scene. Now, we can view hospitalization as a failure of the system.

2.2 Shift from Provider Focus to Patient Focus

For much of my life, patients should be seen and not heard. Patients were not allowed to see their data because they would not understand. Further, for the most part, patients were not interested in seeing their data. With the increased interest in a healthy life, with marketing for a healthy life style, with social media and internet searches, patients have become much more sophisticated. Patients do understand to a greater extent their diseases and want to have a major role in their treatment. Personal Health Records and Patient Portals now provide access to their EHRs. I recognized that the data contained in my system EHR are mostly baseline data resulting from my annual check-ups. What happens to me in everyday life – in my working and playing environments – are much more reflective of my health state. That data is what I want to share with my provider. This push of data into the healthcare process has taken advantage of technology with mobile devices and wearable sensors that provide significant health status data. Patient Reported Outcomes (PRO) provide critical information about the effectiveness of treatment or a dynamic health status.

2.3 Restricted Access to Shared Data

Largely for financial reasons, access to clinical data has been carefully constrained, protected, restricted and controlled. Health care systems compete, and they protect their client base, clinical outcomes, and other sensitive data. The sharing of personalized data across multiple sites causes new challenges in security and privacy [3]. Physicians treat any data from external sources as being of poor quality and of little value. The only data they trust for decision making is the data for which they are responsible. Systems, standards, knowledge, even terminologies and tools developed by vendors and institutions have perceived value and are shared only for a price. One of the consequences is that each group – at all levels of definition – develop what they need independently. In the U.S., each state has responsibility for the development of systems which they do redundantly and slightly differently. However, patients cross state boundaries to receive their care with a high probability that data would not be shared or used across state lines.

Fortunately, we are moving into a period in which the value of sharing data, knowledge, tools, and systems for little or no cost is being accepted. In the U.S., proposals for research grant funding require a section stating how data will be shared with others. Sharing data is a necessity for creating an aggregated Patient-Centric EHR. Big Data and its potential for new knowledge and more data for clinical research is a product of sharing. Rare diseases are less rare as a result of sharing. Pragmatic clinical trials have the potential

of being a primary method of clinical research. However, if the EHR is the data source for clinical trials, the demands for quality, completeness and consistency is paramount. These requirements become a mandate for interoperability. Unfortunately, we are not yet there.

2.4 New Types of Data

In 2002 McGinnis and others published a paper in Health Affairs [4] in which they concluded that shortfalls in medical care contributed only 10% to the health status of an individual. Behavior patterns are responsible for 40% of the health index, yet we spend only about 5% of healthcare expenditures at a population level to improve health-related behavior. Fortunately, that amount of money in the new world has increased considerably. Genetic predispositions contribute 30%, social circumstances 15%, and environmental exposures 5%. These findings have not only influenced our healthcare expenditures but are influencing the kinds of data we collect. Current research efforts are directed toward how to collect and use these kinds of data in healthcare. What genetic data should be included in the EHR, and how will it be stored? How should we present socioeconomic and environment data to the clinician? Geospatial coding of data is now becoming part of the EHR. Suddenly, the clinician is faced with thousands of data items in the decision making process. Humans find it difficult to use more than 5-7 facts in decision making. Do physicians now have to become geneticists to effectively understand and use genetic data?

2.5 Competitive to Collaborative

Competition drives much of our activities. New economic models are pushing more and more healthcare organizations to create collaborative groups for survival. High-speed transport makes it easier for patients to travel to more distant health care facilities. The increase in cost of overhead such as EHR systems, costly diagnostic equipment, and shortage of trained professionals are forcing small and rural institutions to become part of a larger network. Centers of excellence for specific diseases attract patients to well-known institutions such as Mayo, Cleveland Clinic, Johns Hopkins, Duke, and others. Experts have suggested that healthcare in the future may be delivered by less than 10 healthcare systems in the U.S.

New appreciation for the value of Big Data has encouraged collaborative networks for research. Big Data has the potential to help us to better manage chronic diseases. Federal funding for such collaborative research efforts as the NIH Collaboratory [5] grants and the Patient-Centered Outcomes Research Network (PCORnet) [6] are creating access to over 122 million patients. PCORnet is a large network of networks that harnesses the power of collaboration. This network covers much of the U.S.

2.6 Licensed for Fee to Free and Open Source

The balance between proprietary systems and open source systems has been constantly evolving. Products, including standards, terminologies, decision support algorithms, tools, rules, and even data that have value to the entire global community are being made available to all at no cost. If the goal is to have everybody use the same systems, then making that product freely available to all makes sense.

From its beginning in 1987, HL7 required membership in HL7 to access its standards. In September 2012, HL7 International standards became licensed at no cost. Across the industry, there is still an inconsistency in what products are available without cost and what require payment. The problem is that nothing is free. There is a cost in creating, distributing, educating, and administrative infrastructure to support the product. We need to develop a business model that will support and fund the open source process.

2.7 Site Specific to Mobile

The current model for a person to go to the healthcare facility for both health care and illness care is changing. Getting an appointment in a large medical center, especially an academic medical center, one is faced with long wait times – on the order of months. Urgent Care Centers scattered throughout the region have been created to address this problem. In addition, Walgreens, CVS and others now offer urgent care. Patient satisfaction is further strained by long waits well beyond the scheduled times for the appointments. The cost of care is increased by parking fees that are not insignificant.

With the advent of mobile devices, new models for the delivery of care are appearing. The computational capability, the resolution of the display, and the quality of the incorporated video/camera has offered new methods of communication between individuals. The old house call, long disappeared, is coming back in the form of a smart phone encounter. The time, cost, and inconvenience of traveling to a clinic now is being replaced by a call. Further, the smart phone can be combined with a smart application programming interface (API) to collect data, to provide a reminder, to educate the patient, to answer a question, or a number of other tasks.

Wearable sensors can now monitor about 20 data elements in addition to patient reported data. The challenge to the IT community is how to integrate this data and how to validate its accuracy. Apple [7], for example, is addressing health as part of its business model. Apple introduced the Apple Watch which tracks activity and heart rate. Development continues to incorporate more sensing devices to measure and track other important data items. It is interesting to note that that just the reporting feature of step activity has motivated persons to set and reach exercise goals. Apple's Health app on IOS and the HealthKit framework support the aggregation of health-related data. ResearchKit is an open source framework that allows researchers and developers to create apps for medical research. At 6 months, over 100,000 people were using medical apps, and that number is rapidly growing. An example of these apps include autism (Duke),

melanoma-screening (Oregon Health & Science University), and EpiWatch, a seizure-tracking app. CareKit allows individuals to manage their own medical conditions. Other groups, including Fitbit, are developing sensor functionality.

The challenge for the IT community is how to integrate these devices into the EHR and into the healthcare process. It is likely that apps will be developed independently. We need to develop methods to integrate and aggregate the different data items into a cohesive whole. For example, combining activity measurements with heart rate and blood pressure measurements will provide much more information than either measurement alone. Adding one's calendar events would add additional information. The potential of these individual real-time measurements will significantly impact the health status of individuals. Heart attacks and strokes could be detected before they happen, and measures taken to prevent or minimize the event before it happens. These disruptive innovations will impact the sites of care, the delivery of care, communication, and interventions.

2.8 The Aging Population

The aging of the population is one of the most significant trends of the 21st century [8]. One in eight people (12.5%) in the world are age 60 or over. As long as fertility rates continue to decline and life expectancy continues to rise, older people will steadily increase as a proportion of the population. By 2050, the percentage of persons over 60 increase to 22%. The problem for society is what to do with the aged population and how can we afford to support this group. Population Health addresses this problem with incentives including appropriate use of this population and independent aging in the home. Many of the technological developments already discussed provide solutions to many of these problems. Sensors, videos with real time analysis, medication assists, and fall-proofing the homes are examples. The use of robotics in the home have the potential of providing acceptable communications with humans. Real-time data with real-time analytics can predict impending health-negative events. Prearranged connections to a healthcare facility will provide direct triaging and timely interventions.

3 What these Changes Mean to Better Health

Health and healthcare is undergoing remarkable changes. Over the next few years healthcare will flip almost 180 degrees. First, the emphases on health and behavior modification will keep most people out of the hospital. Technology and the tools produced will become strong motivators for healthy lifestyles including nutrition, exercise, and better behavioral habits. The increased use of mobile devices, wearable sensors, and patient reported data shift the focus of healthcare outside of the hospital. A consequence of this pattern shift will mean the people who use the current system will have to pay more for illness care. More importantly, large institutions will have to scale down both in size and staffing. Healthcare will shift from brick and mortar to care where the patient is.

The use of machine learning and artificial intelligence used with Big Data will create new knowledge that will rapidly be applied to patient care. With the recognition of new data types and the integration of that data into the EHR and into decision making, humans will be outclassed, and computers will become the decision makers in healthcare, and the role of clinicians will change. Rich sets of phenotypes and care plans – whether inpatient or home-based will become the norm for applying knowledge to care. Parallel processers and cloud computing will accommodate the demands for faster computing. With every input of new data, a person's health condition will be completely reevaluated. This reevaluation may trigger a thousand decision support algorithms. Ultimately there may be a processor chip per person. Many of the tasks currently carried out will be performed by robots. Virtual and augmented reality will be used for viewing data, for education, and for entertainment. All of these changes will have a professional impact, an economic impact, and a social impact.

This new world will be dependent on the liquidity of data. Interoperability becomes even more critical. Transport standards, a universal common set of complete data elements, open source APIs, phenotype development, new workflows, new data structures, and new data architectures will be only a few of the new requirements facing the technical community.

One of the challenges that must be overcome is how can we keep our systems up to date with new technology, new policies, new models and new demands. Most of the current commercial EHR systems, for example, are based on a database architecture that is over 40 years old. With large customer bases and billions of invested dollars in current systems, how can we accommodate change? A suggestion is to move all functionality from the EHR system itself. The EHR should become simply a container to input and output data (a push and pull operation) supporting a global common data element database. All functions supporting the capture of data, the analysis of data, and the presentation of data will be external to the EHR. These functions can evolve with new technologies and new requirements can be more quickly created and built into work flows.

4 The World as it is

Today's world is largely siloed. We tend to deal with new problems or requirements one at a time. We live with many less than ideal solutions that may only partially solve the problem. We do not have an effective way of bringing all the interested parties together in a comprehensive way to mutually find tomorrow's solution. The section below discusses some case examples of this issue.

Interoperability begins with being able to uniquely identify a person from any and every source. The only logical solution to this problem is a Universal Person Identifier (UPI). The U.S. is the last developed country that does not have a UPI. We attempt to identify a person with between 12-

15 different data elements that may or may not be correct or present. The consequence is an error rate that exceeds 10%. That error rate prevents the level of accuracy to aggregate data across multiple sites. For example, one cannot create a medications list across multiple institutions with sufficient accuracy to make the list reliable and hence usable.

The second problem that must be solved is a common language across all healthcare settings and inclusive of all clinical domains. The current standard sets of preferred controlled terminologies is inadequate to support an information loss-free exchange and aggregation of data. There are too many choices to represent a concept consistently within a single institution much less across an enterprise, region, country, or the global community. The set of attributes such as units, value sets, data types and other characteristics are not standardized. The granularity of data items is not defined at the atomic level required to drive many decision support algorithms. The problem is compounded by the fact that many different groups are attempting to define a common data element set that differs from other groups.

Most of controlled terminologies deal only with a specific classification of terminologies such as diagnoses, procedures, laboratory tests, or medications. Administrative terms are frequently not defined. For example, what is an inpatient? Does an ambulatory visit in which the patient is kept overnight become an inpatient visit? There is no consistent answer to this question. What is the time of discharge of a patient? Is it when the discharge order is written? Is it when the patient leaves the hospital room? Is it when the patient checks out administratively? Analytics require specific definitions of these and similar terms.

A common method for transporting data to other locations is in the form of a document or container. HL7's Clinical Document Architecture (CDA) [9] has been used to define a family of the Continuity of Care Document (CCD) based on the CDA as an electronic document exchange standard for sharing patient summary information. Another Standards Developing Organization (SDO) - Integrating the Health Enterprise (IHE) - has developed an Implementation Guide called Structured Document Capture (SDC) [10] based on the Retrieve Form for Data Capture Profile (RFD). RFD supports the retrieval of a form from a form source, the display and completion of the form, and the return of instance data from the display application to a receiving application. The European for Standardization (CEN) standard, CEN/ISO EN13606 [11], transports data in folders. Most systems store these documents as part of the EHR making it more difficult to extract dynamically the level of granularity required for analytics and decision support. Most data retrieval tools do not support the extraction of atomic data elements from these stored documents.

Specific data elements are generally stored within the contents of the lab or diagnostic test or source of the data. In many cases, a specific data element might be a component of several different tests but is, in fact, the same data element. If the data is stored only within the larger component, a retrieval program must have the knowledge to look at all possible sources to see if the test exists. Finding that a specific data element has never been collected is often a challenging and time consuming task. A new approach to database architecture needs to be approached. Rather than a relational database, we need to consider a noSQL [12] approach in which the storage architecture needs to be driven by operational requirements rather than a table structure. Epic uses two types of databases in that a Mumps/M database is used for a single patient's record and a standard relational database (Clarity) is use when searching across patients.

If EHR data is to be used for pragmatic clinical trials, then that data needs to have the highest quality possible. Data must be precise, complete, accurate and consistent. To whatever extent possible, data quality must be checked as part of the data capture. For example, males cannot be pregnant; women do not have prostate cancer. Simple rules applied at the time of data capture can be applied. These rules may be gender or conditional specific. They may be driven by limits. They may be checked by algorithms such as delta-check. The same approach works for consistency.

5 Evolving Standards

HL7 has been developing standards for over 30 years. The first HL7 standard, v2, was developed specifically to enable the creation of a Hospital Information System (HIS) using a Best of Breed approach. The competition was a single integrated system that did everything. Data needed to be passed among the various components of an HIS. The requirements were well defined, and the v2 standard evolved around the concept of segments based on classification of data – demographic patient data, test ordering and result reporting, encounter data, financial data, etc. The Z segment was added to permit implementers to add any new data element not specifically defined within the standard. Version 2 standards are still in wide use today, especially in the U.S. It is not however an interoperable standard.

Version 3 was based on a Reference Information Model and was more complex and challenging to understand and implement. The CDA resulted from the integration of another group into HL7 and had excellent international implementation. The combining of an ASTM effort, the Continuity of Care Record (CCR) into the CDA standard to create the CCD was a substantial step forward toward interoperability. One of the frustrations to the CCD was that you had to send everything defined within the CCD and there was no freedom to add additional data elements.

The user community pushed back strongly against v3, and there were data exchange needs not met by CDA. At an HL7 Board Retreat, the idea was voiced to create of a Fresh Look Task Force to investigate looking to develop a new standard based on the experience of the past 25 years of creating standards. A task force was created, and in July 2011,

Graham Grieve came up with the concept on a new standard named FHIR ® (Fast Healthcare Interoperability Resources®). FHIR captures the simplicity and flexibility of v2 and is rapidly becoming the standard for interoperability. FHIR is based on a set of modules called Resources. Resources are common basic building blocks by which all data exchanges occur. Resources permit the transporting of data at the lowest levels of granularity or at any level of packaged data. Examples of Resources are Patient, Practitioner, Allergy, Family History, and Care Plan. FHIR uses a web services technology (RESTful APIs) to exchange Resources between groups. FHIR is a service-driven application. Resources are combined in Profiles to support clinical, administrative, or other functionalities. Profiles also may constrain the content of Resources, so only the data required by the service is exchanged. Regardless of the paradigm, the content is the same.

Clearly most of the changes that are taking place depend on interoperable data sharing, enabled by standards, a common set of data elements, and high data quality. The Senate and Congress of the U.S. has passed the 21st Century Cures Act [13] that defines a trusted interoperable framework and common agreement for EHR transparency, usability, security, and functionality. This law requires EHR vendors to make patient data accessible without special effort, through the use of application programming interfaces. These APIs will be a common technology used by everyone to exchange data bidirectionally with EHRs, independent of the vendor. In other words, the APIs support both push data into the EHR or to pull/query data from any EHR. A team at Harvard has defined a set of rules for substitutable medical apps & reusable technology (SMART) [14]. The combination of these specifications along with FHIR (called SMART on FHIR) appears to be the tool of choice for meeting the requirements of the 21st Century Cures Act.

6 Better Health

Better health means getting the right data to the right place to the right person who uses the right knowledge to make the right diagnosis and starts the right treatment. No diagnosis or the wrong diagnosis is the norm for the present system. Only 4% of people with rare disease are correctly diagnosed. The reason is frequently a lack of access to a complete set of quality data. Clinicians have trouble finding the data they need in a timely fashion. Medication errors are the most common type of medical error. These errors are a result of the wrong dosage, interactions among drugs, incorrect medication, illegible handwriting of prescriptions, and allergies unknown or missed.

What is the role of HL7 International in making sure that we help build the New World to ensure Better Health? When HL7 first started, it had a focused view about what was required. Over time, that view began to spread as new requirements were identified and new groups came into the organization. HL7 reached out to the international community, to other SDOs and technical groups (leading to CDA), to the clinical decision support community, to the clinical community, to governments, and now to the genomic community. The changes in health, health care, and technology will lead to new requirements. HL7 needs to be out front in defining what is needed, and making sure someone is accepting responsibility.

It is my belief that computers and AI will become much more dominant in bringing Better Health forward. Computers have the ability to consistently follow a set of rules. Humans frequently miss those opportunities. An example to illustrate that point is the case of harvesting donor organs. Duke, unfortunately, harvests only about 1 in 4 of available donor organs. If a computer program could be triggered by data that suggested an organ might become available for harvest and notify appropriate persons, then track that process throughout the encounter, that number should become 4 out of 4. Having processes start as a result of certain data entry, i.e., a trigger to manage data flow or work flow would make a significant contribution to better health. CDS Hooks is a step in this direction.

References

[1] Blobel B. Translational Medicine Meets New Technologies for Enabling Personalized Care. Stud Health Technol Inform. 2013; 189: 8-23.

[2] Ortman JM, Velkoff VA, Hogan H. An Aging Nation: The Older Population in the United States. U.S. Census Bureau, 2014.

[3] Blobel B. Paradigm Changes of Health Systems Towards Ubiquitous, Personalized Health Leads to Paradigm Changes of Security and Privacy Ecosystems. Int J Biomedicine and Healthcare. 2015; 3: 6-11.

[4] McGinnis JM, Williams-Russo P, Knickman JR. The Case for More Active Policy Attention to Health Promotion. Health Affairs. 2012; 21: 78-93.

[5] www.nihcollaboratory.gov

[6] www.pcornet.org

[7] www.apple.com

[8] https://www.census.gov/prod/2014pubs/p25-1140.pdf

[9] www.hl7.org

[10] www.healthit.gov/policy-researchers-implementers/structured-data-capture-sdc.

[11] Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis TJ. An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes. J. Biomed. Inform. 2010; 43: 736-746.

[12] Sánchez-de-Madariaga R, Muñoz A, Lozano-Rubí R, Serrano-Balazote P, Castro AL, Moreno O, Pascual M. Examining database persistence of ISO/EN 13606 standardized electronic health record extracts: relational vs. NoSQL approaches. BMC Med Inform Decis Mak. 2017; 17: 123.

[13] www.congress.gov/bill/114th-congress/house-bill/6

[14] https://smarthealthit.org/

Standardization for Mastering Healthcare Transformation - Challenges and Solutions

Bernd Blobel[1,2,3,4]

[1] Medical Faculty, University of Regensburg, Germany

[2] eHealth Competence Center Bavaria, Deggendorf Institute of Technology, Germany

[3] HL7 Germany, Cologne, Germany

[4] First Medical Faculty, Charles University Prague, Czech Republic

Abstract

Background: The ongoing healthcare transformation results in systems medicine and personalized care. This requires advanced interoperability scenarios.
Objectives: The paper aims at introducing a reference architecture model to enable the harmonization of existing and future specifications and standards.
Methods: For enabling the harmonization of standards and specifications, they have to be correctly and formally represented using a system-theoretical, architecture-centric, ontology-based, policy-driven approach.
Results: An Interoperability Reference Architecture Model has been developed and standardized at ISO TC 215 and CEN TC 251. It is used for re-engineering specifications and standards, that way enabling harmonization with any other specification. The proposed modeling methodology has also been introduced in several HL7 specifications. It is demonstrated by re-engineering the ISO 13606 reference model.
Conclusion: The approach presented in the paper offers a solution for comprehensive interoperability at any interoperability level also including user domains and the individual user. That way, never-ending revision cycles and negotiation rounds for integrating different specifications in complex eHealth solutions become obsolete.

Keywords

Healthcare transformation; Interoperability; Reference architecture model; Systems theory; Ontology; Policy

Correspondence to:

Prof. Dr. habil. Bernd Blobel, FACMI, FACHI, FHL7, FEFMI, MIAHSI
Medical Faculty, University of Regensburg, Germany
E-mail: bernd.blobel@klinik.uni-regensburg.de

EJBI 2017; 13(1):09-15
received: July 24, 2017
accepted: August 15, 2017
published: October 10, 2017

1 Introduction

Healthcare transformation, translational medicine – what's behind those terms and related moves? Health systems around the world are faced with the demographic change towards aging, multi-diseased societies, the related development of human resources, a demanding attitude regarding health and social services, and medical and biomedical progress, altogether bound to exploding costs for health-related R&D as well as health services delivery.

The first approach for managing the aforementioned challenges by healthcare transformation goes back to Kaiser Permanente's pioneering efforts of "inventing" managed care in the early twenties of the last century. In the seventies this approach of controlling health services delivery and financing in a system under medical and economic aspects with a special focus on process efficiency and efficacy for cost containment has been generalized in the US and implemented in so-called Health Maintenance Organizations (HMOs). A specialty in the context of chronic and/or multi-morbid patients is disease management or case management. Meanwhile, many different managed care concepts have been implemented in different regions around the world [1].

The next level of healthcare transformation focuses stronger on quality and outcome of the care process, resulting in value-based medicine. When measuring health outcome as part of care delivery, clinical practice and life sciences research get aligned [2]. Developing and implementing better prevention, new diagnostic tools and treatments by using a multi-disciplinary, highly collaborative, 'from bench-to-bedside' approach is called translational medicine [3]. Population health supports spreading those achievements throughout the community, so completing

the three pillars 'bench-side', 'bed-side', and 'community' of translational medicine. The convergence of healthcare and life sciences ecosystems leads to personalized healthcare. The transformation of health systems is combined with a changing definition of health from a status free of diseases towards physical, psychological and social wellbeing. The aforementioned demographic, social and economic challenges lead to health systems paradigm changes. Following, the health systems evolution from the perspective of interoperability will be considered in some more details.

2 The Transformation of the Healthcare Eco-System – A System-Theoretical Approach

As discussed in several papers already [4, 5, 6], the health systems paradigm changes we are facing can be grouped in three dimensions: organizational, methodological, and technological. Organizationally, healthcare delivery turns from organization-centric through process controlled (DMP) to person-centered systems. In methodological respect, health services delivery was traditionally performed as an empirical, phenomenological approach to general care addressing health problems by one solution fits all. At next level, population was stratified according to specific clinical conditions for an evidence-based approach to dedicated care. Finally, we move towards personalized, preventive, predictive and participative care considering individual health state, conditions and contexts, i.e. a systems medicine approach from art to multi-disciplinary science. Technologically, a move to fully distributed systems, mobile technologies, nano-, molecular and bio-technologies, knowledge representation & -management, artificial intelligence, big data & business analytics, cloud computing, and social business happens.

For properly modeling the described health systems evolution, we will deploy systems theory for representing, analyzing and designing the systems in question [7].

A system is an ordered composition of interrelated elements, which have properties in common distinguishing them from those elements defining the system's environment. A system interacts with its environment. A system represents constructive or structural aspects on the one side and functional aspects on the other side. Considering just phenomena observed, the system's function as a black box can be described by the transformation of inputs into specific outputs under special environmental impacts. The quality of inputs and outputs can belong to the following categories: material, energy, and information. A system's architecture describes the system's elements (components), their functions and their interrelations. A system can be considered from different perspectives – the system's domains – usually performed by domain experts using their methodologies, terminologies and underlying ontologies. Examples are the medical, legal, or administrative perspective on health systems. The representation of specific perspectives on a real world system – the system's domains – has to be based on those domains' ontologies. An ontology describes an ordering system of entities of a domain and their relations. After Gruber, an ontology is a formal explicit specification of a shared conceptualization of a domain of interest [8]. Rules for selecting components and functions as well as constraints of the relations according to a business case are called policies. Policies define the intended behavior of a system.

The organizational and the methodological paradigm changes health systems are faced with result in higher structural complexity of those systems. Transformative healthcare systems are represented by the subject of care and the processes analyzing and managing his/her health comprises all levels of granularity from atoms through molecules, cell components, cells, tissues, organs, bodies, communities, up to population. Regarding the functional, or in general inter-relational, aspects of that system, the number of perspectives or domains to be included increases, comprising quantum-mechanical effects in the nano-world, biochemical processes, interrelations based on classical physics, and finally social interrelations in the macro-world. We have to have always in mind: As we can consistently model and compute only systems of reasonable complexity, the system analysis or design has to address partial systems when considering higher granularity levels of the system in question.

Health systems transformation leads regarding its structure to fine grained, more complex systems, and regarding its functions and interrelations to dynamic, multi-domain (multidisciplinary) systems.

3 The Transformative Healthcare Interoperability Challenge

Interoperability is traditionally considered according to the IEEE definition as "ability of two or more systems or components to exchange information and to use the information that has been exchanged" [9]. It has been implemented as communication at the level of data/information exchange such as structured messaging (e.g. EDI (Electronic Data Interchange), HL7 messaging) or sharing concepts (e.g. openEHR Archetypes [10], ISO 13940 ContSys concepts [11]), and as cooperation at application level (e.g. Web services, such as HL7 FHIR Resources (Fast Healthcare Interoperability Resources) [12]).

The described interoperability levels are gray shadowed in Figure 1 [13], representing different levels of data/information sharing or application services provision. For such interoperability approaches, the quality of inputs and outputs of the systems considered is limited to information, and the perspective and representation style is just ICT and its

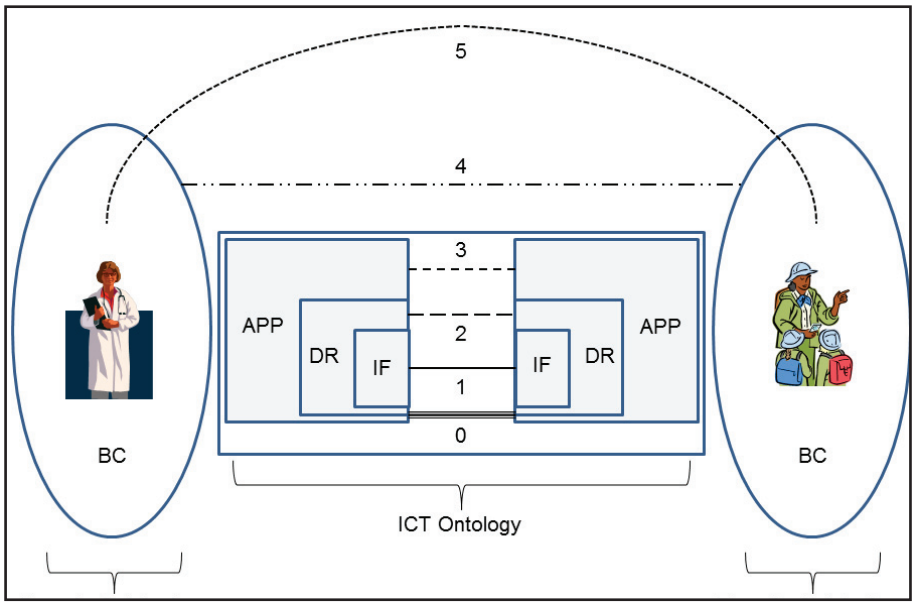


Figure 1: Interoperability schema [13].

ontologies, more or less ignoring the domains (disciplines) to be supported, their knowledge and methodologies expressed in their concepts as elements of their ontologies.

However, this is not the only limitation of the traditional interoperability approach. The definition and representation of concepts depends on the involved domains, the concrete business case, the business objectives, the business as well as the individual context, the selected business process for meeting the business objectives. For establishing advanced interoperability, all those conditions defining communication and cooperation between entities must be mutually considered. In detail, this implies in a white box approach the definition of the system to be considered for the business case and the intended business objectives, i.e. the components, functions and interrelations as system's architecture needed, but also related environmental and contextual conditions. The implicit knowledge on the defined system and its behavior provided by domain experts must be represented explicitly for enabling the performance of cooperative decisions on the systems required architecture and processes. So, advanced interoperability requires a system-theoretical, architecture-centric, ontology-based approach. The aforementioned interoperability decisions on the system in question can only be provided at that real-world level, but not at the level of its informational representation necessary for its implementation. In other words, the non-ICT interoperability is not just defining the ICT interoperability solution, it is the real challenge in the game.

As in detail discussed in [14] already, it is not just the complexity caused by multiple disciplines, but also the involvement of differently qualified actors merging different professions, but also professionals and laymen. The subject of care, frequently a layman in the healthcare business, is in the center and dominates or even controls the process. For that purpose, all interoperability levels presented in Table 1 must be provided by all participating systems cooperating on this level.

Organizational, methodological and technological paradigm changes in health systems require a multi-disciplinary, international and even global approach in legal, cultural, language, terminology and ontology respect. This requires the management of multiple domains including multiple policy domains as presented in Figure 2 [14].

Table 1: Interoperability levels of the comprehensive interoperability schema.

Information Perspective		**Organizational Perspective**	
Interoperability Level	**Instances**	**Interoperability Level**	
Technical interoperability	Technical plug & play, signal- & protocol compatibility	Light-weight interactions	Simple
Structural interoperability	Simple EDI, envelopes	Information sharing	T
Syntactic interoperability	Messages and clinical documents with agreed upon vocabulary		
Semantic interoperability	Advanced messaging with common information models and terminologies	Coordination	⇕
Organizations/ Service interoperability	Common business process	Agreed Cooperation	
Knowledge-based interoperability	Multi-domain processes	Cross-domain Cooperation	
Skills-based interoperability	Multi-domain individual engagement	Moderated end-user collaboration	Advanced

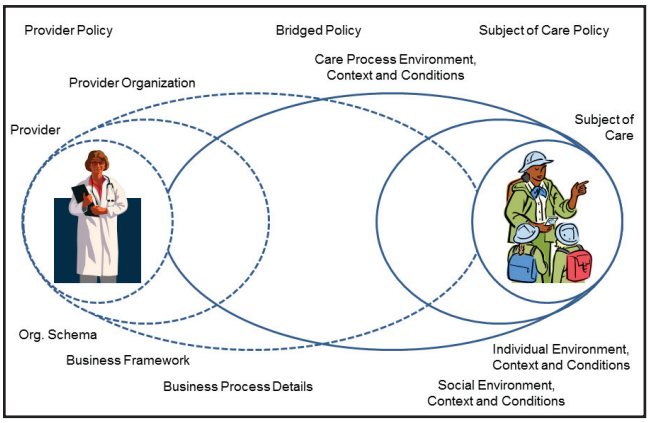


Figure 2: pHealth interoperability schema [14].

4 The Interoperability Reference Architecture Model Standard

Standards Development Organizations (SDOs) claim to enable open and flexible cooperation between actors in health business cases while usually addressing a specific aspect of the business from a specific domain's perspective. At best, they thereby deploy (somehow) that domain's ontology. In real world business systems and especially in the context of advanced healthcare paradigms however, multiple domains are involved.

For meeting the multi-disciplinary interoperability challenges, an abstract domain-independent representation of systems is deployed, based on a system-theoretical, architecture-centric, ontology-driven approach explained above [4, 6]. The mathematical concept representation using the universal type theory in combination with systems engineering methodologies allows representing any system architecturally (i.e. the system's components, their functions and internal as well as external relations) by generically describing its composition/decomposition as well as the aspects (domains) of the system relevant in a specific context (e.g. business case), instantiated using those

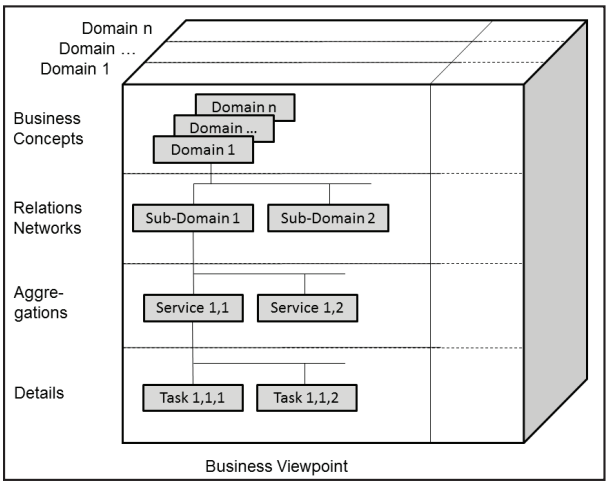


Figure 3: Granularity Levels of the Interoperability Reference Architecture Model.

domains' ontologies. The reference architecture model - here focusing on the business domain - can be used recursively, so representing, e.g., the real-world systems' continuum from elementary particles to the universe (Figure 3).

Additionally to agreeing on one or more, and at best standards-based, ICT ontologies, the agreed domains' knowledge, but also individual (language, education, skills, experiences, social and psychological aspects, etc.) and environmental context must be represented, harmonized and communicated by instantiating the system's architectural components and behavior through the domain-specific ontologies and policies.

By combining that model with ISO/IEC 10746 RM-ODP, the Interoperability Reference Architecture Model (introduced in the nineties as Generic Component Model - GCM) as well as the applicable rules - the Interoperability Reference Architecture Model Framework - (also known as GCM Framework) is completed (Figure 4) [6, 15].

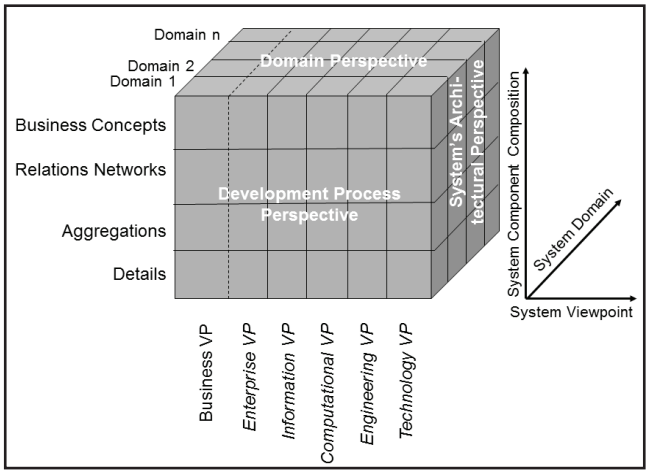


Figure 4: The interoperability reference architecture model.

This Interoperability Reference Architecture Model allows consistently transforming and interrelating any domain-specific subsystem's structure and behavior (e.g. domain-specific standards and specifications) by ontologically representing its concepts and relationships at the real world system component's level of granularity. In other words, the domain-specific subsystem (e.g. a domain-specific standard or specification) is re-engineered using the Interoperability Reference Architecture Model, by that way providing a standardized interface to that specification (Figure 5).

According to the GCM Framework, inter-domain relationships can only be established at the same level of granularity [16]. Therefore, intra-domain specializations/generalizations have to be performed first.

ISO TC 215 "Health informatics" and CEN TC 251 "Health informatics" decided at the Joint Working Group Meeting 2015 in Bern, Switzerland, already to follow the presented approach for all interoperability specifications. The solution has meanwhile been standardized at ISO

TC 215. The Interoperability Reference Architecture Model supports ontology harmonization or knowledge harmonization to enable interoperability between existing systems, standards and solutions of any level of complexity without the demand for continuously adapting/revising those specifications.

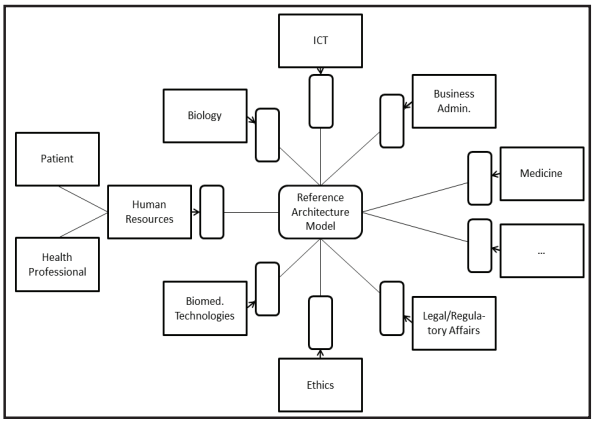


Figure 5: Interoperability mediated by the interoperability reference architecture model.

5 Example for Re-Engineering an ISO Specification Using the Interoperability Reference Architecture Model

A basic ISO standard, endlessly revised for adopting to, and interconnecting with, many related specifications, is ISO 13606 "Health informatics – EHR communication" [10]. Therefore, its conceptual model (ISO 13606 Reference Architecture) has been re-engineered to enable advanced interoperability to any other specification. Figure 6 presents the ISO 13606 Reference Architecture Model.

As an example, Figure 7 presents the outcome of re-engineering the ISO 13606 "EHR communication" Reference Model (colored components) into the GCM, thereby just considering the domains addressed in that model. On that basis, the harmonization of ISO 13606 with other specs such as ISO 22600 Health informatics - Privilege management and access control [17], ISO 21298 Health informatics - Functional and structural roles [18], – a permanent challenge Standards Development Organizations are faced with – can be easily performed. The Figure 7 shows a mixture of different viewpoints requiring advanced transformations not considered in the standard. Furthermore, there is a vast amount of explicit knowledge missing in ISO 13606 but necessary for harmonization as demonstrated by the non-colored components, which complete the architectural model. Just the ISO 13606 components presented in three dimensions represent valid architectural components in the reference architecture model. The others (colored rectangles) have to be transferred into valid components.

The described process can be automated. The same holds for transforming the cross-domain, harmonized, consistent informational representation of the complex business system into the different ISO/IEC 10746 views [19] for analyzing,

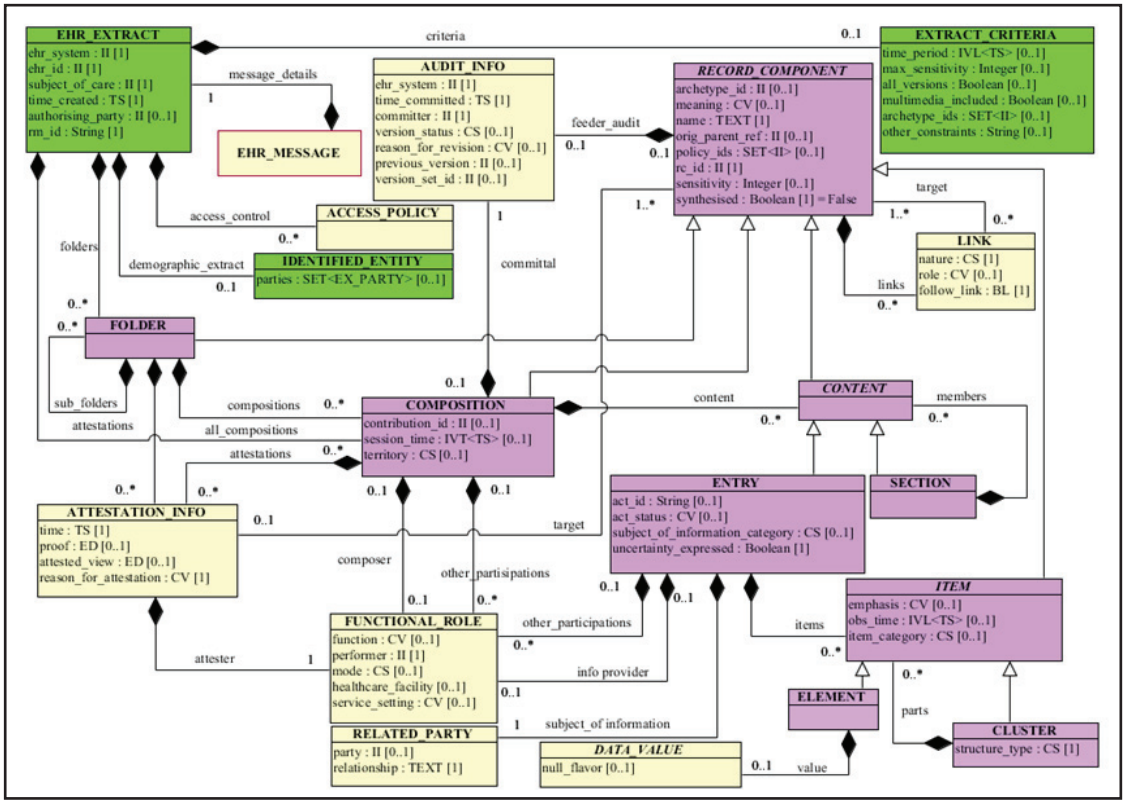


Figure 6: ISO 13606 reference Model.

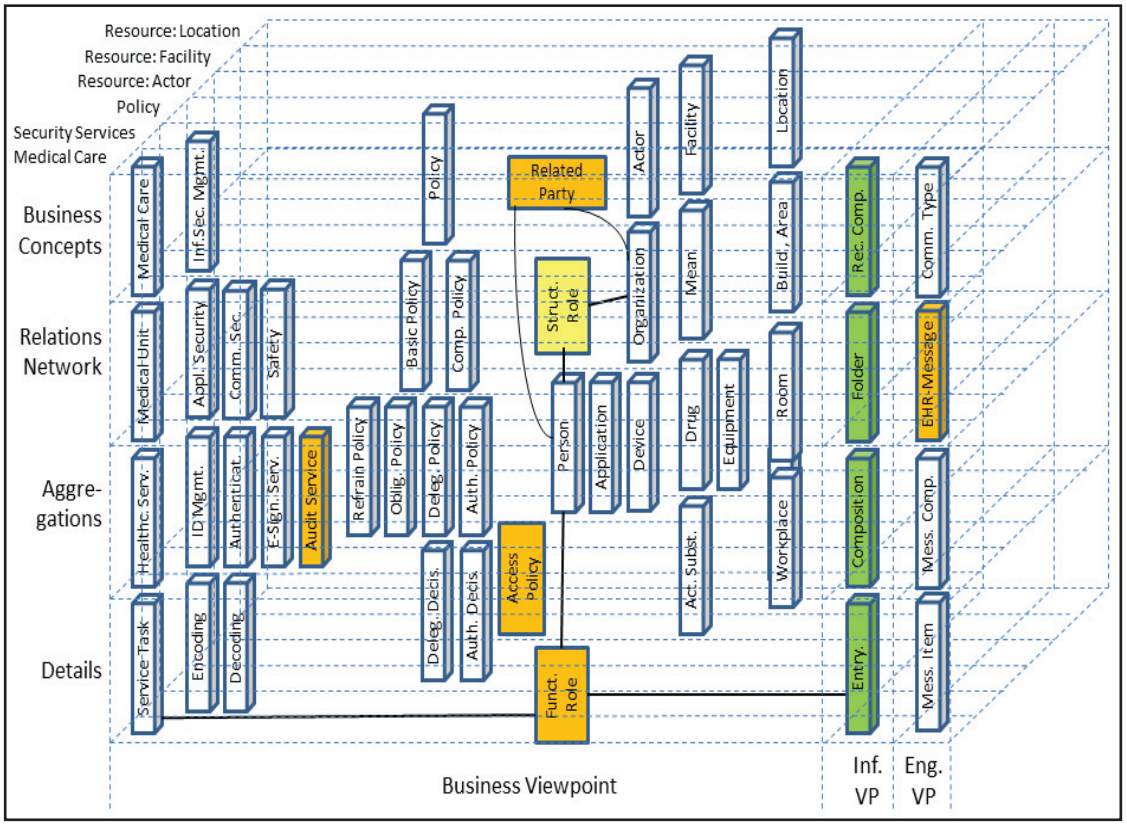


Figure 7: Re-engineering example of the ISO 13606 reference Model.

designing, implementing and maintaining the related ICT solution.

6 Discussion and Conclusion

For harmonizing domain or use case specific specifications when adding a new specification or changing/ extending the business case, especially when including another domain, currently a mutual adaptation and harmonization is performed, resulting in a revision process of the impacted standards and specifications. With increasing complexity and variability of the system and the diversity of its subsystems and components, the lifetime of domain specific specifications goes down, and a comprehensive pre-coordination is impossible.

The alternative way of a priori harmonizing the aforementioned highly complex, highly dynamic, multi-disciplinary/multi-domain advanced healthcare system by representing it by one domain's terminology/ontology or - even worse - by using ICT ontologies fails. The same holds when using one domain's representational style and models or standards as reference or master that all other domains and their experts must adhere to, e.g., by enforcing biologists, physicians, philosophers and artists to think and represent in UML and the 78 concepts of the ICT base ontology [20].

Therefore, an adaptive approach is required to sharing and harmonizing ICT, domain, and personal ontologies and conditions at runtime.

The presented approach has been successfully deployed in several cross-domain ISO specifications, such as ISO 22600, ISO 21298, HL7 Composite Security and Privacy Domain Analysis Model [21]. Its feasibility has been practically demonstrated for automatically harmonizing HL7 v2.x and HL7 v3 specifications [16, 22], for enabling use case and domain crossing interoperability in the context of ISO 13972 Health informatics - Detailed clinical models [23, 24], or for automatically designing inter-domain Web services to facilitate multi-disciplinary approaches to Type 2 Diabetes Care management [25, 26]. The approach also allows a comparative analysis and evaluation of ICT Enterprise Architectures [6].

The increasingly complex interoperability challenge cannot be met starting from an information object point of view, as most of the existing approaches including many HL7 work products do. When constructing a picture of the real world from those information artefacts, we will result in the tragic situation of creating an excellent map, however the landscape is wrong.

Acknowledgements

The author is indebted to thank his colleagues from ISO TC 215, CEN TC 251, and HL7 International for their kind cooperation and support.

References

[1] Amelung VE, Krauth C, Mühlbacher A. Managed Care. Springer: Gabler Wirtschaftslexikon. http://wirtschaftslexikon.gabler.de/

[2] Hansson E, Spencer B, Kent J, Clawson J, Meerkatt H, Lanson S. The Value-Based Hospital and Translational Medicine. Bcg.perspectives, September 09, 2014. https://www.bcgperspectives.com/content/articles/health_care_payers_providers_transformation_large_scale_change_value_based_hospital_translational_medicine/

[3] Wikipedia. Translational medicine. https://en.wikipedia.org/wiki/Translational_medicine

[4] Blobel B. Architectural Approach to eHealth for Enabling Paradigm Changes in Health. Methods Inf Med. 2010; 49: 123-134.

[5] Blobel B. Translational Medicine Meets New Technologies for Enabling Personalized Care. Stud Heaqlth Technol Inform. 2013; 189: 8-23.

[6] Blobel B, Oemig F. The Importance of Architectures for Interoperability. Stud Health Technol Inform. 2015; 211: 18-56.

[7] Völz H. Information. Berlin: Akademie-Verlag; 1982.

[8] Gruber TR. Toward principles for the design of ontologies used for knowledge sharing. Int J Human – Computer Studies. 1995; 43: 907-928.

[9] IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries, IEEE; 1990.

[10] International Organization for Standardization. ISO 13606:2017 Health informatics – EHR communication. Geneva: ISO; 2017.

[11] International Organization for Standardization. 13940:2015 Health informatics - System of concepts to support continuity of care. Geneva: ISO; 2015.

[12] Health Level 7 International, Inc. Fast Health Interoperability Resources (FHIR). Available at http://www.hl7.org.

[13] Oemig F and Blobel B. Natural Language Processing Supporting Interoperability in Healthcare. In: Biemann C, Mehler A (Edrs.) Text Mining – From Ontology Learning to Automated Text Processing Applications, 137-156. Series Theory and Applications of Natural Language Processing. Heidelberg: Springer-Verlag GmbH; 2014.

[14] Blobel B, Ruotsalainen P, Lopez DM, Oemig F (2017) Requirements and Solutions for Personalized Health Systems. Stud Health Technol Inform. 2017; 237: 3-21.

[15] Oemig F. Development of an ontology-based architecture for ensuring semantic interoperability between communication standards in healthcare (in German). PhD Thesis, University of Regensburg, Medical Faculty, Regensburg, Germany, 2011.

[16] Oemig F, Blobel B. A Formal Analysis of HL7 Version 2.x. Stud. Health Technol Inform. 2011; 169: 704-708.

[17] International Organization for Standardization. ISO 22600:2014 Health informatics – Privilege management and access control. Geneva: ISO; 2014.

[18] International Organization for Standardization. ISO 21298:2016 Health informatics – Structural and functional roles. Geneva: ISO; 2016.

[19] International Organization for Standardization. ISO/IEC 10746-2:1996 Information Technology – Open Distributed Processing – Reference Model: Foundations. Geneva: ISO; 1996.

[20] Akerman A, Tyree J. Using ontology to support development of software architectures. IBM Systems Journal. 2006; 45: 813–825.

[21] Health Level 7 International, Inc. HL7 Composite Security and Privacy Domain Analysis Model. Available at http://www.hl7.org.

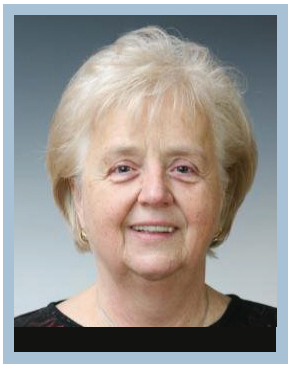
[22] Oemig F, Blobel B. A Communication Standards Ontology Using Basic Formal Ontologies. Stud Health Technol Inform. 2010; 156: 105-113.

[23] Goossen WTF. Detailed clinical models: Representing knowledge, data and semantics in healthcare information technology. Healthcare Informatics Research. 2014; 20: 163-172.

[24] Goossen W, Gossen-Baremany A, Van Der Zel M. Detailed clinical models: A review. Healthcare Informatics Research. 2010; 16: 201-2014.

[25] Uribe GA, Blobel B, López DM, Schulz S. A Generic Architecture for an Adaptive, Interoperable and Intelligent Type 2 Diabetes Mellitus Care System. Stud Health Technol Inform. 2015; 211: 121–131.

[26] Uribe GA, Blobel B, López DM, Ruiz AA. Specializing Architectures for the Type 2 Diabetes Mellitus Care Use Cases with a Focus on Process Management. Stud Health Technol Inform. 2015; 211: 132–142.



In Remembrance of Our Beloved Editor-in-Chief
Prof. Jana Zvarova, PhD (1943–2017)

Prof. RN Dr. Jana Zvárová, DrSc, the founder of EuroMISE Mentor Association and she founded and served as Chief Editor of the European Journal of Biomedical Informatics (EJBI).

The legacy of prof. RNDr. Jana Zvárová, DrSc. will always be alive, and she will always be remembered as an outstanding mentor and colleague by her innumerable students, colleagues and friends.

An Integrated Model to Capture the Provision of Health and Social Care Services Based on the ContSys and FHIR Standards

Fabrizio Pecoraro, Daniela Luzi, Fabrizio L Ricci

Institute for Research on Population and Social Policies, National Research Council, Rome, Italy

Abstract

In the last decades, there is a shift in the provision of health services from the formal facilities to home care. The delivery of care at home is considered a sustainable approach that not only results in substantial cost savings but is also a key step towards achieving optimal outcomes in the delivery of high-quality health services. Moreover, the adoption of this approach is facilitated by the prompt development of medical and non-medical technologies that can simplify care coordination and enable distant monitoring of the patient. However, home care requires the integration and coordination of health and social stakeholders in a continuity of care perspective. This requires the adoption of standards to define a common shared conceptual model that improves the interoperability among information systems and organizations. In this paper to accomplish this task we combine the ContSys standard and the HL7 FHIR (Fast Healthcare Interoperability Resources) framework to identify concepts to be exchanged between systems, organizations and providers involved in the health and social context. This model has been developed to improve the integration and coordination of social and health care services within the Italian H@H project.

Keywords

Continuity of care; Integrated care; Home care; ContSys standard; HL7 FHIR

Correspondence to:

Fabrizio Pecoraro

Institute for Research on Population and Social Policies, National Research Council, Via Palestro, 32 – 00185 – Rome, Italy

E-mail: f.pecoraro@irpps.cnr.it

EJBI 2017; 13(1):17-26

received: May 31, 2017

accepted: July 20, 2017

published: October 10, 2017

1 Introduction

The demographic, epidemiological and social changes related to a growing number of older people with an increased incidence of multiple chronic conditions and of functional and cognitive impairments are leading to an increased demand of health and social service integration at home [1]. This has led also many European countries to adopt policies to strongly shift the organization and provision of health and social services from formal institutional facilities (e.g. hospitals) to home care [2, 3, 4]. This is considered a sustainable approach that results in a substantial cost savings [5, 6, 7, 8] mainly preventing unnecessary or long-term hospitalizations [7]. Moreover, it supports patients with chronic illness as well as temporarily frail individuals in a protected and comfortable environment [9] in line with patient preferences to be cared in their own home [4, 10, 11]. It provides appropriate and high-quality care services also considering the support of informal caregivers (e.g. family members), with the use of appropriate technologies [2, 12], such as those developed within the Ambient Assisted Living (AAL) program [3]. The rapid development of medical and non-medical technologies together with new and modern organization of health and social service delivery can represent a viable solution to develop home care provided that healthcare systems further enhance integration and support coordination [3] with social services, such as house cleaning, home maintenance, shopping, and transport.

The integration of health and social services within a home care delivery model needs the connection between the health care and the social systems taking into account their specific characteristics [3]. However, a common criticism of home and community services offered by different providers is that they are fragmented, resulting in poor outcomes and wasted resources [5]. For this reason there is a need to improve coordination and cooperation among stakeholders both in a vertical (e.g. primary and secondary care) and a horizontal perspectives (e.g. health and social care) taking also into account the different needs of the target population (e.g. elderly, temporarily fragile, healthy people). To accomplish this task one of the first issues that should be

overcome is the definition of a common shared conceptual schema developed using standardized data models as well as common vocabularies and nomenclatures.

In this perspective one of the most widely used framework to improve the continuity of care and to enable communication between different information systems is the ContSys [13] standard that provides a set of concepts to describe both the content and context of the health services provided in a patient-centred care delivery. Moreover, among the different standards developed in healthcare to facilitate system interoperability and under the perspective of data models HL7 [14] surely represents one of the main candidates for the integration and exchange of information [15] generally focused on patient's care delivery. A relatively recent standard released by HL7 is the FHIR (Fast Healthcare Interoperability Resources) language that is considered to be the most promising candidate standard framework for the exchange of healthcare information [16]. As reported in the HL7 website [16], "the philosophy behind FHIR is to build a base set of resources that, either by themselves or when combined, satisfy the majority of common use cases. FHIR resources aim to define the information contents and structure for the core information set that is shared by most implementations".

This solution has been adopted within the Health At Home (H@H) project with the aim of developing an open platform that provides a technological and organizational solution based on integrated and interoperable software services as well as Internet of Things. This paper describes the H@H project focusing on its architecture and proposes a conceptual model as a result of a first step that combines the use of the ContSys standard to define a shared integrated conceptual model with the FHIR standard release 3 (STU) that can facilitate the semantic interoperability among systems and organizations.

2 The H@H Project Infrastructure

H@H project provides a technological and organizational solution to support both subjects of care and organizations that deliver health and social services at home based on integrated and interoperable software services and Internet of Things (e.g. devices). Examples of health services are remote monitoring of vital signs and body measures, telemedicine, home visits provided by specialists, primary care professionals; examples of social services are household maintenance, meal preparation, personal care activities. The overall purpose of the H@H infrastructure is to improve the "quality of life" of citizens/families which have temporary or permanent "weaknesses" or health care assistance and the remote monitoring of chronic patients in outpatient settings.

In Figure 1, a high-level architecture of the project is shown [17], where the primary focus is on the user who accesses to a set of services interacting with a cluster of sensors that monitor user's vital signs, environmental parameters, user lifestyle, etc. These devices can be placed at home (e.g. temperature and humidity, motion and occupancy, flood and leak sensors) and/or worn by the user (e.g. fitness tracker, fall monitor, blood pressure measurement). Further data are produced by other cluster of sensors to enrich the home setting monitoring activities carried out by the user outside

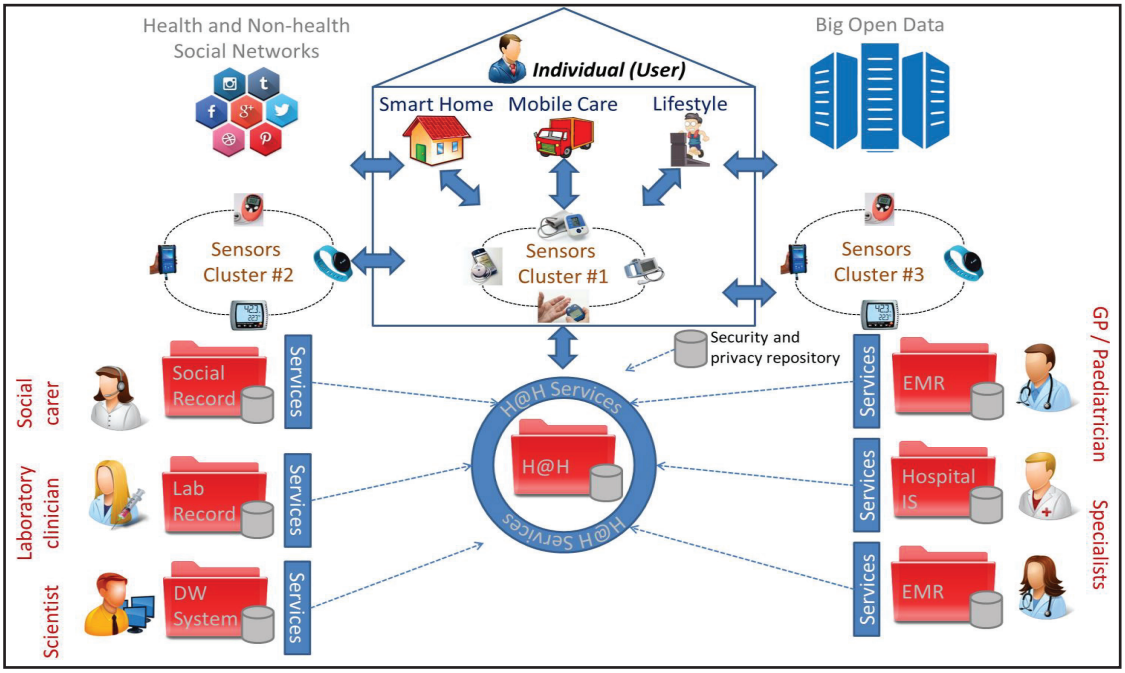


Figure 1: Architecture of the H@H project at high-level of description.

home or to capture other contextual information, such as weather conditions and forecast, transport facilities. Data provided by these sets of clusters are gathered in different databases composing the H@H system and are interoperable with information provided by different information systems managed by social and health care providers, such as the General Practitioner and specialist Electronic Health Records, Hospital and Laboratory Information Systems, Electronic Social Record. The interaction and data exchange are executed through a set of web services developed on the basis of the cloud computing paradigm. An important set of functionalities provided by this platform concerns the real-time monitoring of data coming from sensors and information systems and their communication to an operating centre that is responsible to identify and manage alerts and periodical measurements of functional parameters of the devices. Moreover, a data warehouse (generated by export functionality applied on current databases) is foreseen to analyse data for secondary purposes in a clinical governance framework to evaluate the effectiveness and appropriateness of the provision of health and social services and others uses (epidemiology, clinical researches, education, etc.) [18]. Privacy and security services are also considered to manage rules to be applied to manage databases and feed the data warehouse.

The platform is a robust and flexible Cloud Services Framework designed to be easily customizable through the use of JAVA and XML technologies. These characteristics will allow to easily extend the data domain and to provide the immediate remote access to data. Furthermore, it will interact with other distributed technologies including SQL and NoSQL (e.g. object and document oriented) databases.

3 Materials and Methods

This section describes the steps of the methodology adopted to define a conceptual model that integrates both health and social services as well as the mapping procedures implemented to represent these concepts with the resources of the HL7 FHIR standards. In particular the first step is to define the health model by selecting the main concepts from the ContSys standard that fit the purposes of the H@H project. Starting from this model, the second step is to detect the main concepts that can be shared with the social model to define an abstract core backbone that can link both the health and social schemas. This meta-model is the starting point to define the social schema identifying the concepts specifically needed in the social care context. The third step is to define the integrated model where the health and social care concepts are compared and mapped with the FHIR resources in the final step of the methodology.

3.1 Definition of the Health Conceptual Model

In this first step of the proposed methodology the main concepts that fit the purposes of the health part of the H@H project are identified. This selection is based from the concepts presented in the ISO 13940 (also known as ContSys) [13] that is an important and widely used standard in the provision of a common conceptual model for continuity of care [19] as well as to facilitate the communication between different information systems at the semantic level. This model meets the identified requirements of the H@H project where the continuity of care is one of the main tasks to be accomplished. The ContSys standard, which includes approximately 84 concepts organized into eight areas closely related to each other, is needed to represent both the content and context of the healthcare services delivered. Its aim is to define a common conceptual framework to describe generic concepts that represent both the content and context of the health services provided during the interaction between a subject of care and health professionals in a patient-centred care delivery.

3.2 Definition of the Social Conceptual Model

Given that there are not common standards in social care that play a significant role in the identification of a conceptual model to improve the interoperability among systems and organizations, to identify the main concepts of the social schema we firstly identify a meta-model composed by a set of ContSys concepts that can be shared with the social model to achieve an abstract description of a generic activity as well as the different concepts needed to describe it. Once this abstraction is performed to identify additional concepts needed to describe the social context, the different types of social services to be provided at home to support the subject of care and/or the family are analysed on the basis of a set of regional, national and European nomenclatures. This is an important step of the methodology considering that, even if social care is an important determinant of care [20], its activities are not included in the ContSys standard.

3.3 Definition of the Integrated Model

Once both models have been defined, the subsequent step of the methodology is to propose a conceptual model able to capture the interaction between the individual and the professionals in the provision of health and social service at home in a continuity of care perspective. To accomplish this task each concept is classified within the following categories taking into account its definition as well as the information gathered:

1. Shareable concepts that are in common between the health and social models.

2. Generalized concepts that are represented using similar classes in both models and that can be related with ISA relationship.
3. Concepts that are modelled only in one context (e.g. social care) but that can be included in the other one (e.g. health care).
4. Specific concepts that are explicitly used in one model and that are not needed in the other one.

3.4 Mapping of the Integrated Model with the FHIR Schema

In this final step, each concept of the integrated model has been subsequently mapped with the resources of the FHIR model, identifying in particular:

1. Classes that directly mapped with the resources of the FHIR (same semantic of the class as well as attributes).
2. Resources of the FHIR to be accommodated to map the classes of the conceptual model in terms of: semantic as well as vocabulary and nomenclatures type to be modified.
3. Resources of the FHIR to be extended to map the classes of the conceptual model in terms of: attributes to be added, associations with other resources to be added.

This mapping exercise can be considered as a test to verify to what extent the two standards can be integrated to facilitate the information exchange in a continuity of care perspective.

4 Results

In this section the results of the above described steps of the methodology are reported.

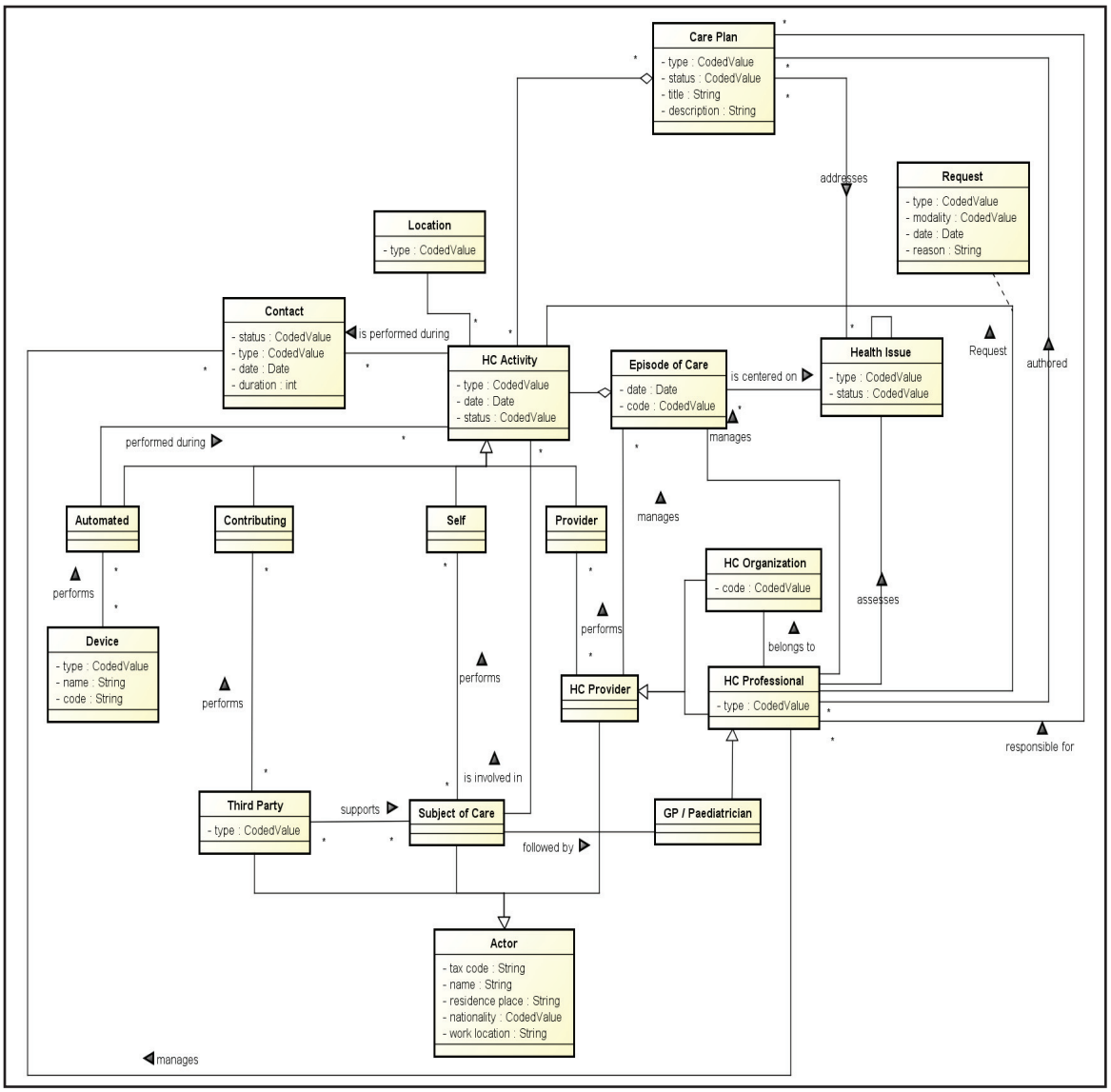


Figure 2: Healthcare conceptual model.

4.1 Healthcare Conceptual Model

The result of the first step of the methodology is shown in Figure 2 using the UML class diagram. It highlights: (i) the main concepts of the ContSys standard (i.e. activity, contact, episode of care, health issue); (ii) the main actors involved (i.e. subject of care, healthcare provider, third party such as a family member); (iii) the device and the location.

In particular the main classes are summarized in Table 1 reporting the related definition on the basis of the ContSys standard. As described in the standard document [13] healthcare is delivered through a set of Healthcare Activities that are performed in a Location,that may be required by a Healthcare Professional and included within a specific Care Plan. These activities are provided within a specific time interval (Episode Of Care) during which a Health Issue of the Subject of Care is addressed. Moreover, these set of activities are performed during a specific encounter (Contact) where the Subject of Care interacts directly or indirectly with one or more Healthcare Professionals. Note that an Healthcare Activity can be performed not only by a Healthcare Provider but also conducted directly by the Subject of Care (e.g. measuring the blood pressure), by a Third Party who is responsible for supporting the Subject of Care (e.g. a family member who executes the haemoglobin glycated test) or by a Device without the intervention of a human being (e.g. Holter).

4.2 Social Care Conceptual Model

As highlighted in the previous paragraph the core concept of the health model based on the ContSys standard is the activity that is posed at the centre of the model and connected to five main concepts: Health Issue to detect why the activity is performed; Subject Of Care and Healthcare Provider to specify who is involved in the performance of the activity as well as the related Organization; Location to identify where the activity is performed; Care Plan to specify the activities, objectives and goals related to specific health issues of the Subject Of Care. It is important to note that a generalization of the backbone of the health schema is needed to be integrated in the corresponding concepts of the social part of the conceptual model, as shown in Figure 3.

Table 1: List of health related concepts modelled within the H@H schema.

Class	Definition as reported in the ContSys standard
Healthcare Activity	Activity intended directly or indirectly to improve or maintain a health state.
Healthcare Provider Activity	Healthcare activity performed by a healthcare provider.
Self-care Activity	Healthcare activity performed by the subject of care.
Healthcare Contributing Activity	Healthcare activity performed by a healthcare third party
Automated Activity (Automated Healthcare)	Method of delivering healthcare initiated by a responsible healthcare actor and thereafter delivered automatically by an automatic medical device
Care Plan	Dynamic, personalized plan including identified needed healthcare activities, health objectives and healthcare goals, relating to one or more specified health issues in a healthcare process
Contact	Interaction between a subject of care and one or more healthcare personnel.
Location = Point of care	Location / point of care where direct healthcare activities are performed
Episode of care	Health related period during which healthcare activities are performed to address one health issue as identified by one healthcare professional
Health issue	Representation of an issue related to the health of a subject of care as identified by one or more healthcare actors. See also Health condition: observed or potential observable aspects of the health state at a given time.
Healthcare Actor	Organization or person participating in healthcare
Healthcare Provider	Healthcare actor that is able to be assigned one or more care period mandates. It can be a professional or an organization.
Healthcare Professional	Healthcare personnel having a healthcare professional entitlement recognized in a given jurisdiction.
Healthcare Organization	Healthcare provider having an organization role. See also healthcare provider = healthcare actor that is able to be assigned one or more care period mandates
Subject of Care	Healthcare actor with a person role; who seeks to receive, is receiving, or has received healthcare
Third Party	Healthcare actor other than a healthcare provider or the subject of care. Relative (family member) aiding the subject of care, any actor responsible for social support, or for the funding, payment, or reimbursement of healthcare provision are healthcare third parties.
Device = Automatic medical device	Medical device capable of performing automated healthcare activities
General Practitioner	Additional health concept not included in the ContSys standard that describes a specific healthcare professional who plays the role of care giver of the Subject of care.

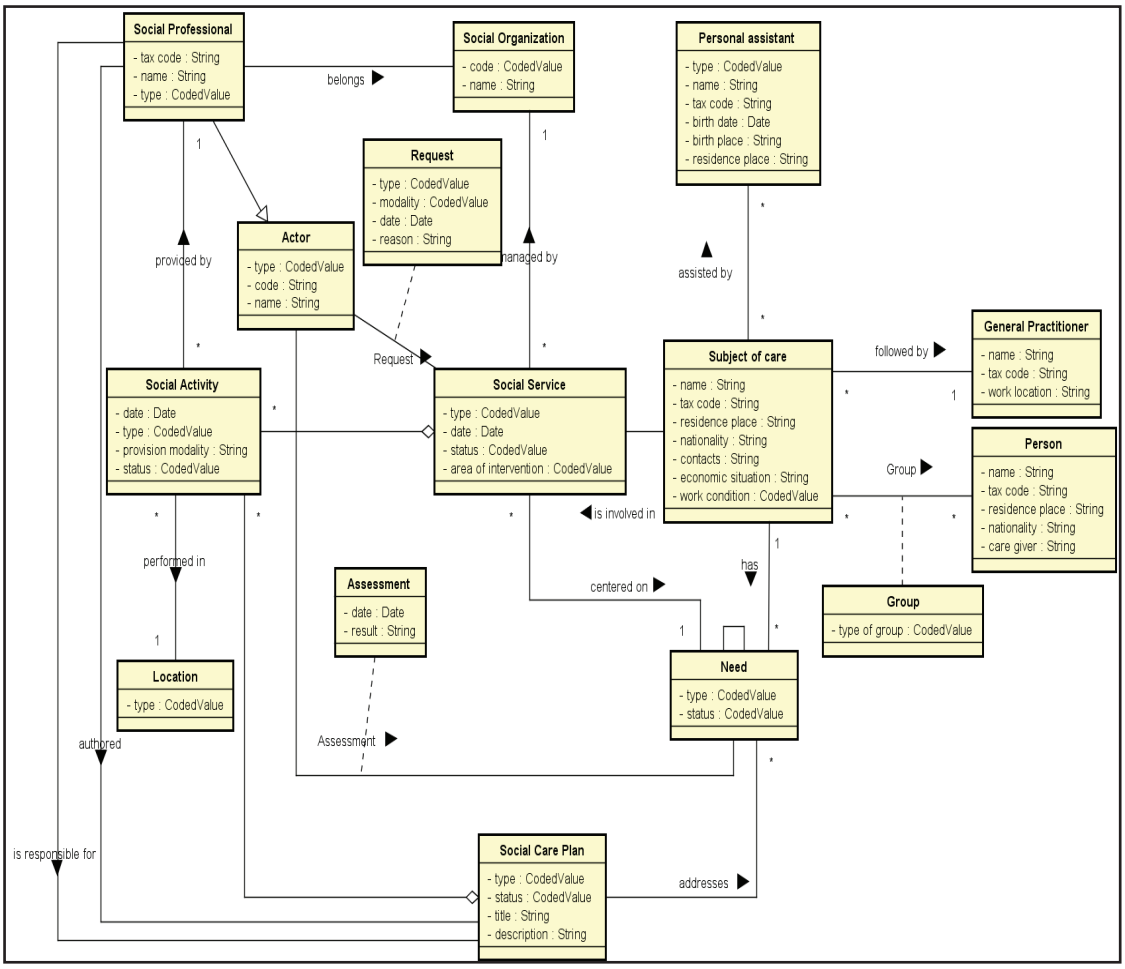


Figure 3: Social care conceptual model.

The core concept of the model is the Social Activity, performed by a care professional in a specific Location (e.g. home of the subject of care). A set of social activities are provided within an overall Social Service that is required by an Actor (that can be either a social or a health care professional) and focused on a specific Area of Interest (elderly, disable, minority, family, migration, poverty, mental health, etc.) and on a specific Need (economic, employment, health, house, educational, etc.). Each Need is assessed by an Actor who determine whether the patient is eligible to access the relevant social services. Each Subject of Care can be involved in many social services each one provided by a specific Provider Organization. As social services are provided also to a group of persons (even if associated to a specific subject of care), the conceptual model describes the membership of the Subject of Care to a family or a social network composed by a set of Persons (e.g. wife/husband, children, brothers/sisters, etc. as well as roommates, neighbours, friends, etc) also highligthing if a care giver can be detected among these persons. Finally, the subject of care is related with his/her General Practitioner.

Table 2: List of concepts that are shared between health and social conceptual model.

Concept	Comments
Subject of care	The ID can be different in health and social care.
Location	No differences between health and social models.
Actor	No differences between health and social models.

4.3 Integrated Model

As reported in the methodology, prior to define the integrated schema a mapping between the health and social concepts has been performed and the results of this analysis is reported in Tables 2-5 highlighting, respectively, sharable, generalizable and extendable concepts as well as those that are specifically devoted for one context and not included in the other one.

The concepts reported in Table 2 represent the main elements that are used to describe an activity carried out in both health and social contexts.

Each of the concepts reported in Table 3 has a particular specification that assumes a specific meaning within the context where it is applied. For instance, both the Care plan

and the Social care plan are referred to the abstract concept of plan (i.e. workflow) which is intended to describe the set of activities performed by a provider in a time sequence. This concept is specialized in each of the two contexts to capture the specificity of the health and the social part of the plan.

Each of the concepts reported in Table 4 has an important role to describe the context of reference. Although these concepts have not been considered in the model of the other context, they can be used to enrich it and to specify additional useful information. For instance, the Contact which is a medical concept, can be also used from a social perspective to collect all the relevant Social activities performed during a specific encounter between the Subject of care and a Social professional.

The concepts reported in Table 5 are specifically adopted by one of the two contexts and not included in the other one given that they seem to have a little use in it. For this reason, in this mapping procedure we do not extend it in different contexts, as evident considering the Health issue and the Social need classes.

4.4 Mapping of the Integrated Model with the FHIR Schema

Once the conceptual model is defined, this step provides the result of a mapping procedure between the concepts of the schema with the resources of the FHIR standard. In particular Tables 6-8 highlight, respectively, resources

Table 3: List of health concepts that are generalizable to be included in the social conceptual model.

Health concept	Social concept	Comments
Healthcare Provider	Social care Provider	Representing in both context a professional or an organization that delivers / manages or is generally involved in an activity.
Healthcare Professional	Social care Professional	Representing in both context a professional that delivers / manages or is generally involved in an activity.
Healthcare Organization	Social care Organization	Representing in both context an organization that delivers / manages or is generally involved in an activity.
Healthcare Activity	Social Activity + Social Service	Same concept to be generalized to capture difference in the attribute list (and coded values) as well as in the association with the provider.
Care plan	Social care plan	Same concept to be generalized to capture difference in the attribute list (and coded values) as well as in the association with the provider.

Table 4: List of concepts that are included within a context and that can be also adopted in the other model.

Concept	Context	Comments
General Practitioner	Social	This concept is adopted in the social care model to determine who is responsible for the health part of the patient's care. Even if ContSys this association is not modelled, it can be a useful information also in the healthcare model.
Third party	Health	This concept is adopted to determine who performs the third party activities as well as to identify who supports the subject of care, and can be used to specify who is the care giver of the subject of care within the social care perspective.
Episode of care	Health	This concept is adopted to group a set of activities performed for a specific health issue. It can be used also to specify which are the social activities provided for the same episode of care.
Contact	Health	This concept is adopted to capture the encounter between the subject of care and an healthcare professional grouping the set of activities performed in that specific consultation. It can be used also to specify which are the encounters between the subject of care and the social professionals. Of course attributes relevant for the social care have to be determined and included.

Table 5: List of concepts that are included within a specific context and that are not suitable in the other model.

Concept	Context	Comments
Other Person and Group	Social	The concept *Other person* is used in the social model to associate the subject of care with other persons in a specific *Group* (e.g. family, working group). These concepts seem to have a little use in the health context especially under a patient-centred perspective.
Automated; Contributing; Self; Provider activity	Health	These are concepts that generalize the healthcare activity class. They are useful in the healthcare context to specify the type of actor that provide the service. Not useful in the social setting given that services are provided only by social care professionals and organizations.
Health issue	Health	These two concepts are used in the relevant context to model the care condition of the subject of care. They are modelled using two different classes that can be related with each other to specify whether a health issue can determine a social need.
Need	Social	
Device	Health	If an automated activity is delivered this concept identifies the device involved in its performance. No automated activities are expected in the social care context.

Table 6: List of concepts that are straightforward mapped with the FHIR resources.

ContSys	FHIR	Description of FHIR resource(s)
Healthcare Professional	AdministrativeResources. Individuals. Practitioner	A person who is directly or indirectly involved in the provisioning of healthcare. Practitioner role = A specific set of Roles/Locations/specialties/services that a practitioner may perform at an organization for a period of time.
Healthcare Organization	Base. Entities. Organization	A formally or informally recognized grouping of people or organizations formed for the purpose of achieving some form of collective action. Includes companies, institutions, corporations, departments, community groups, healthcare practice groups, etc.
Care Plan	Clinical. Care provision. Care Plan	Describes the intention of how one or more practitioners intend to deliver care for a particular patient, group or community for a period of time, possibly limited to care for a specific condition or set of conditions.
Third Party	AdministrativeResources. Individuals RelatedPerson	Information about a person that is involved in the care for a patient, but who is not the target of healthcare, nor has a formal responsibility in the care process
Episode of care	Workflow. Encounters. Episode of care	An association between a patient and an organization / healthcare provider(s) during which time encounters may occur. The managing organization assumes a level of responsibility for the patient during this time.
Contact	Workflow. Encounters. Encounter	An interaction between a patient and healthcare provider(s) for the purpose of providing healthcare service(s) or assessing the health status of a patient
Other Person	AdministrativeResources. Individuals Related Person	Information about a person that is involved in the care for a patient, but who is not the target of healthcare, nor has a formal responsibility in the care process
Group	AdministrativeResources. Groups. Group	Represents a defined collection of entities that may be discussed or acted upon collectively but which are not expected to act collectively and are not formally or legally recognized; i.e. a collection of entities that isn't an Organization.
Health issue	Clinical. General. Condition	A clinical condition, problem, diagnosis, or other event, situation, issue, or clinical concept that has risen to a level of concern.
Device	AdministrativeResources. Device. Device	This resource identifies an instance or a type of a manufactured item that is used in the provision of healthcare without being substantially changed through that activity. The device may be a medical or non-medical device. Medical devices include durable (reusable) medical equipment, implantable devices, as well as disposable equipment used for diagnostic, treatment, and research for healthcare and public health. Non-medical devices may include items such as a machine, cell phone, computer, application, etc.
	& AdministrativeResources. Device. Component	Device component = The characteristics, operational status and capabilities of a medical-related component of a medical device

Table 7: FHIR resources that have to be accommodated to be mapped with the relevant H@H concept.

Concept	FHIR resource	Definition of the resource	Accommodation
General Practitioner	Base. Individuals. Practitioner + Practitioner Role	A person who is directly or indirectly involved in the provisioning of healthcare.	The role of the practitioner is set to GP.
Need	Clinical. General. Condition	A clinical condition, problem, diagnosis, or other event, situation, issue, or clinical concept that has risen to a level of concern.	Condition refers to health problems. So this resource should be accommodated to describe social needs at least in terms of vocabulary as well as nomenclatures.
Social care Professional	AdministrativeResources. Individuals. Practitioner	A person who is directly or indirectly involved in the provisioning of healthcare. Practitioner role = A specific set of Roles/ Locations/specialties/services that a practitioner may perform at an organization for a period of time.	Extends the vocabulary items to include the type of social care practitioner

that can be straightforward mapped with the concepts of the H@H schema (Table 6), FHIR resources that have to be accommodated to map with the relevant concept of the H@H schema (Table 7) and resources that have to be extended to capture the information of the relevant H@H concept (Table 8). As previously highlighted, this mapping exercise is an important first step to verify to what extent the ContSys concepts can be integrated with the FHIR resources to facilitate the information exchange in a continuity of care perspective.

Table 7 and 8 highlight on the one hand the need of capturing through the use of a standard the context and context of the ContSys concepts to improve the integration between the health and social care. On the other hand, the feasibility of the proposed approach should be verified

Table 7 (cont'd): FHIR resources that have to be accommodated to be mapped with the relevant H@H concept.

Concept	FHIR resource	Definition of the resource	Accommodation
Social care Organization	Base. Entities. Organization	A formally or informally recognized grouping of people or organizations formed for the purpose of achieving some form of collective action. Includes companies, institutions, corporations, departments, community groups, healthcare practice groups, etc.	Extends the vocabulary items to include the type of social care organization
Location	Administrative Resources. Entities. Location	Details and position information for a physical place where services are provided and resources and participants may be stored, found, contained or accommodated.	Extends the vocabulary items to include the type of social care delivery location

Table 8: List of FHIR resources that have to be extended to be mapped with the relevant H@H concept.

Concept	FHIR resource	Definition of the resource	Extension
Subject of care	Aministrative Resources. Individuals. Patient	Demographics and other administrative information about an individual or animal receiving care or other health-related services	Includes attributes specifically devoted to capture social and economic status.
Healthcare Activity	Clinical. General. Procedure	An action that is or was performed on a patient. This can be a physical intervention like an operation, or less invasive like counselling or hypnotherapy	Includes an attribute to capture the type of healthcare activity performed (e.g. visit, lab exam, etc.)
Social Activity	Clinical. General. Procedure	An action that is or was performed on a patient. This can be a physical intervention like an operation, or less invasive like counselling or hypnotherapy	Includes an attribute to capture the type of social activity performed (e.g. cleaning, good provision, etc.)
Automated; Contributing; Self; Provider activity	Clinical. General. Procedure	An action that is or was performed on a patient. This can be a physical intervention like an operation, or less invasive like counselling or hypnotherapy	The type of activity is not specified within the Procedure resources but it can be detected on the basis of the provider (Device for Automated, Related Person for contributing, Patient for self-activity and Practitioner for provider). However, an additional attribute can help identifying this information.

considering the standards under development, such as the ISO 13606-3:2017 [21].

5 Conclusions

The paper describes the H@H conceptual model integrating the social care with the health care concepts. This was designed on the basis of the ContSys standard deployed in the healthcare context to improve the continuity of care. The semantic interoperability is assured through the use of the HL7 FHIR standard framework that also improves the coordination and cooperation activities of the different stakeholders involved in the provision of services in an comprehensive continuity of care process. Our model is a starting point to analyse the various determinants of care able to guarantee a high-level of individual's quality of life. As previously highlighted, this mapping exercise is an important first step to verify to what extent the ContSys concepts can be integrated with the FHIR resources to facilitate the information exchange in a continuity of care perspective. Considering that the FHIR the infrastructure has been proved to be stable while the resources are still subject to improvements [16] our work was a necessary test to capture feasible resources that conform to the content and context of the ContSys concepts. Of course the introduction of these resources and/or the introduction of new ones should be subject to significant review through ballot and other HL7 processes. Moreover, we intend to analyse the ISO 13606-3:2017 [21] which at the moment is under development to further improve the methodology in the perspective of identifying clinical archetypes.

At the moment H@H information system is at the implementation phase. During its implementation and also the next deployment the evaluation of social technological alignment will be also is taken into account to establish a shared understanding between the social context and the software domain.

6 Acknowledgements

This work is part of the H@H (Health At Home) – Smart Communities for the wellbeing of the citizens – project granted within the National Operational Programme for Research and Competitiveness 2007-2013. Identification code: SCN_00558.

References

[1] Goodwin N, Dixon A, Anderson G, Wodchis W. Providing integrated care for older people with complex needs: lessons from seven international case studies. London: The King's Fund. 2014.

[2] Home-based and long-term care: home care issues and evidence. Geneva, World Health Organization, 1999. Available from: www.who.int/chp/knowledge/publications/Chronic_conditions/en/index.html

[3] Tarricone R, Agis TD. Home care in Europe: the solid facts. WHO Regional Office Europe. 2008.

[4] Lindberg B, Nilsson C, Zotterman D, Söderberg S, Skär L. Using information and communication technology in home care for communication between patients, family members, and healthcare professionals: A systematic review. Int J Telemed Appl. 2013.

[5] Clarfield M, Bergman H, Kane R. Fragmentation of care for frail older people—an international problem. Experience from three countries: Israel, Canada, and the United States. J Am Geriatr Soc. 2001; 49: 1714-1721.

[6] Boerma W, Kroneman M, Hutchinson A, Saltman RB. Home care across Europe. N. Genet (Ed.). London: European Observatory on Health Systems and Policies. 2013.

[7] Landi F, Onder G, Russo A, Tabaccanti S, Rollo R, Federici S, Tua E, Cesari M, Bernabei R. A new model of integrated home care for the elderly: impact on hospital use. J Clin Epidemiol. 2001; 54: 968-970.

[8] Chan M, Campo E, Estève D, Fourniols JY. Smart homes—current features and future perspectives. Maturitas. 2009; 64: 90-97.

[9] Home – The Best Place for Health Care. A positioning statement from The Joint Commission on the state of the home care industry. Available from: http://www.jhartfound.org/images/uploads/resources/Home_Care_position_paper_4_5_111.pdf

[10] Kok L, Berden C, Sadiraj K. Costs and benefits of home care for the elderly versus residential care: a comparison using propensity scores. Eur J Health Econ. 2015; 16: 119-131.

[11] Shepperd S, Iliffe S. Hospital at home versus in-patient hospital care. Cochrane Database Syst Rev. 2005; 3.

[12] Community home-based care in resource-limited settings: a framework for action. Geneva, World Health Organization, 2002. Available from: www.who.int/chp/knowledge/publications/Chronic_conditions/en/index.html

[13] International Organization for Standardization. ISO 13940: Health Informatics – System of Concepts to Support Continuity of Care (ContSys). Geneva: ISO; 2015.

[14] Benson T. SNOMED CT. Principles of Health Interoperability HL7 and SNOMED. 2010.

[15] Schadow G, Mead CN. The HL7 Reference Information Model under scrutiny. Stud Health Technol Inform. 2007; 124: 151-156.

[16] Health Level Seven, Inc. FHIR: Fast healthcare interoperability resources. Available from: http://hl7.org/implement/standards/fhir/

[17] Pecoraro F, Luzi D, Pourabbas E, Ricci FL. A conceptual model for integrating social and health care services at home: the H@H project. Proceedings of the 18th International conference on E-Health, Networking, Application and Services, 2016. DOI: 10.1109/HealthCom.2016.7749489

[18] Pecoraro F, Luzi D, Ricci FL. Secondary uses of EHR systems: A feasibility study. IEEE E-Health and Bioengineering Conference (EHB). 2013; 1-6.

[19] Hägglund M, Chen R, Koch S. Modeling shared care plans using CONTsys and openEHR to support shared homecare of the elderly. J Am Med Inform Assoc. 2011; 18: 66-69.

[20] World Health Organization. Integrated health services: what and why? Health Services Delivery Technical Brief No.1, 2008.

[21] International Organization for Standardization. ISO/DIS 13606-3: Health Informatics – Electronic health record communication – Part 3: Reference archetypes and term lists. Geneva: ISO; 2017.

HL7 Standards and Components to Support Implementation of the European General Data Protection Regulation (GDPR)

Alexander Mense[1], Bernd Blobel[2,3,4]

[1] University of Applied Sciences Technikum, Wien, Austria
[2] Medical Faculty, University of Regensburg, Germany
[3] eHealth Competence Center Bavaria, Deggendorf Institute of Technology, Germany
[4] First Medical Faculty, Charles University Prague, Czech Republic

Abstract

Objectives: Aiming to strengthen EU citizens' fundamental privacy rights in the digital age the new European General Data Protection Regulation shall apply from May 25th 2018. It will require companies processing personal data to implement a set of organizational and technical controls for ensuring proper handling of these data. Obviously this applies for companies providing eHealth services. As HL7 offers a lot of material to support security and privacy for handling personal healthcare data, this paper aims at showing which HL7 standards and components can be used to support the implementation of GDPR related controls.

Methods: The paper shows some key facts of the European GDPR as well as analyzes HL7 standards and components in the security and privacy domain to provide a basic mapping.
Results: As a result the paper provides a table mapping HL7 artifacts to GDPR requirements.
Conclusion: The paper shows, that consequently using HL7 security and privacy standards and components efficiently helps to implement GDPR requirements.

Keywords

Privacy; Security; HL7; CDA; FHIR

Correspondence to:

Prof. Alexander Mense
University of Applied Sciences Technikum,
Hoechstaedtplatz 6, Wien, Austria.
E-mail: mense@technikum-wien.at

EJBI 2017; 13(1):27-33
received: June 11, 2017
accepted: July 16, 2017
published: October 10, 2017

1 Introduction

On May, 24th 2016, the new European General Data Protection Regulation (GDPR) [1, 2] came into force and shall apply from May, 25th 2018 [3]. It replaces Directive 95/46/EC [4] from 1995 and all related national laws. As regulation (in contrast to a directive) it is in its form legally binding for the European Union Member States. The process of developing this regulation was started in 2012 as "an essential step to strengthen citizens' fundamental rights in the digital age and facilitate business by simplifying rules for companies in the Digital Single Market" [5].

The GDPR "*lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data.*" [2, p.32]. This single set of rules for all countries of the European Union implements a "one-stop shop" for all organizations and companies within the Union.

The regulation will require companies processing personal data and particularly sensitive data to implement a set of organizational and technical controls for ensuring proper handling of these data according to security and privacy requirements. So it protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data without restricting or prohibiting the free movement of personal data within the Union in that context [2]. In consequence this also impacts software companies offering systems for handling personal and sensitive data. Thus, beside the "Directive on security of network and information systems (NIS Directive)" (concerning measures for a high common level of security of network and information systems across the Union) [6] the GDPR will utmost impact the healthcare domain.

"*The regulation applies if the data controller (organization that collects data from EU residents) or processor (organization that processes data on behalf of data controller e.g. cloud service providers) or the data subject (person) is based in the EU. Furthermore the Regulation also applies to organizations based outside the European Union if they collect or process personal data of EU residents.*" [7].

In case of violation of the rule the following sanctions and fees can be imposed [2]:

- a warning in writing in cases of first and non-intentional non-compliance
- regular periodic data protection audits
- a fine up to 10,000,000 EUR or up to 2% of the annual worldwide turnover of the preceding financial year in case of an enterprise, whichever is greater (Article 83, Paragraph 4)
- a fine up to 20,000,000 EUR or up to 4% of the annual worldwide turnover of the preceding financial year in case of an enterprise, whichever is greater (Article 83, Paragraph 5 & 6)

HL7 International provides "*a comprehensive framework and related International standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services*" [8]. Therefore HL7 International offers several specifications and standards that can be used to support the implementation of the GDPR requirements in EHR and PHR systems as well the health data exchange in an interoperable manner.

This paper aims at introducing the fundamental principles and rules of the GDPR as well as at providing an overview about relevant HL7 standards and a mapping of HL7 artifacts to GDPR requirements.

2 Methods and Materials

In a first step, technical core aspects are extracted from the GDPR (2.1) and possibly relevant HL7 standards and frameworks are identified. They are differentiated into base standards (2.2), CDA R2 based specifications (2.3), FHIR based resources (2.4) and HL7 V2 (2.5).

2.1 Core Aspects of GDPR

The GDPR defines personal data as "*any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*". Thereby, it properly reflects organizational, methodological and technological paradigm changes health systems are facing [9].

The GDPR defines several obligations for data controllers accountable to demonstrate compliance and thus, setting a framework for accountability. This includes the request for maintaining certain documentation, for performing a data protection impact assessment for more risky processing, for designating a data protection officer in some cases and for implementing data protection measures by design and by default, for instance for data minimization. This places the legal obligation on the Data Controller to notify the Supervisory Authority on a data breach without undue delay.

From the text of the regulation the following detailed technical core requirements can be derived [2]:

R1: Data protection by design and by default

Taking into account the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing of their personal data, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organizational measures which are designed to implement data-protection principles to meet the requirements of the regulation and protect the rights of data subjects Also the controller shall implement appropriate measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. ([2], Article 25). Besides an appropriate, well documented development cycle this technically requires possibilities to specifically mark (label) information that falls under specific requirements or have to be treated specifically according to a person's privacy policy.

R2: Data portability

"This right allows for data subjects to receive the personal data that they have provided to a data controller, in a structured, commonly used and machine-readable format, and to transmit those data to another data controller without hindrance." [10, p3]. Meaning that any person has a right to take their data elsewhere, and the data controller must provide it machine readable form. This is the requirement for interoperable systems.

R3: Right to be forgotten–notification requirement

Data subjects already have a right to have outdated information removed or updated. Now the data controller must also notify other parties that received the data about the change.

R4: Unambiguous consent

A person's consent for processing and storing your data must be freely given, specific, informed and unambiguous. For sensitive data it must be explicit - implied consent is not accepted. A data subject's consent to processing of their personal data must be as easy to withdraw as to give. The data controller is required to be able to demonstrate that consent was given, which means that detailed consent information has to be maintained on file as well as exchanged with communication partners.

R5: Privacy notices

The GDPR requires that information of how private information is processed has to be presented in an easy to understand, clear and plain language. The same holds true also for the aforementioned consent documents. Data controller's policies have to be transparent and easily accessible.

R6: Right to Access / Records of processing activities

Every data controller must keep records pertaining to all aspects of the data processing operations under its responsibility. Records of processing activities must be maintained, that include purposes of the processing, categories involved and envisaged time limits. These records must be made available to the supervisory authority on request. The GDPR also imposes such record-keeping obligation on data processors and requires data controllers and data processors to cooperate (see also R9). This clearly defines the requirement for maintaining provenance information.

R7: Explicit and formally represented policies

By defining the terms "binding corporate rules" ([2], Article 46) and "joint controllers" ([2], Article 26) implicitly the definition and use of explicit and formally represented policies becomes necessary. An overview how to model a policy-driven system for managing personal health information can be for instance in [11]. For healthcare data exchange this also implies the requirement for setting up a common security and privacy policy domain.

The aforementioned requirements can only be met by declaring and managing multiple policies. Those policies including individual ones must be formally represented to enable dynamic and possibly automated policy harmonization [9]. Requirements R4 and R5, but also some others establish a demand for a system-oriented, architecture-centric, ontology-based approach to interoperability as defined at ISO 215 and CEN 251 with the Interoperability Reference Architecture Model for their interoperability standards and meanwhile approved for ISO 13606 [12].

2.2 HL7 General Concepts for Security and Privacy

HL7 provides some general concepts that can be used as a general basis for implementing a security and privacy framework to cover parts of the GDPR requirements.

S1: HL7 Version 3 DAM: Composite Security and Privacy Domain Analysis Model – Release 1

This DAM [13] contains a harmonized analysis of security and privacy policies required to support the security and privacy needs of healthcare organizations. It is an implementation of the ISO 22600 policy ontology [14] and identifies the information and system behaviors required to implement technological controls enforcing healthcare security and privacy policies, therefore representing the basis for policy based access control systems.

S2: HL7 Healthcare Privacy and Security Classification System (HCS), Release 1

The HCS [15] provides a common syntax and semantics standard for interoperable security labels to bind security labels to (primarily) healthcare data to enable data segmentation, fined grained access control and communication of security information related to a resource. Therefore it defines a normative set of interoperable healthcare security label fields (see Figure 1) to be assigned as a security label to healthcare information passed between systems within a security domain and specifies a conforming standard HL7 security

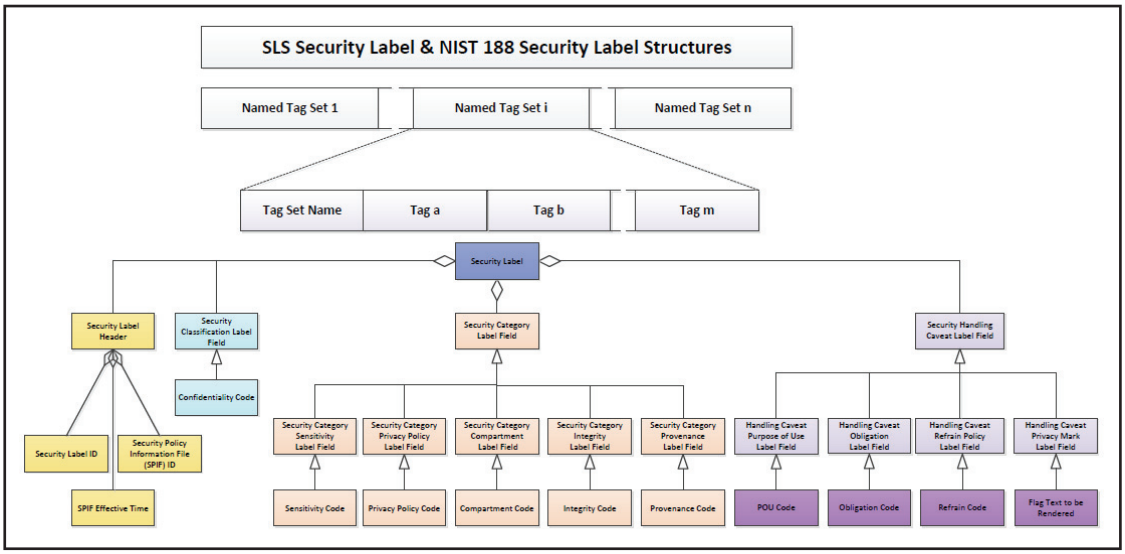


Figure 1: HL7 security and privacy labels [14].

label vocabulary. These definitions have been used as the fundamental basis for CDA R2 implementation guides (see section 2.3) as well as security metadata for FHIR artifacts (see section 2.4). The use of definitions has been show in several projects (e.g. [16]).

S3: HL7 Version 3 Standard: Privacy, Access and Security Services; Security Labeling Service, Release 1 (SLS)

This standard specifies interoperable Security Labeling functional capabilities that are exposed through well-defined, technology agnostic service interfaces. Functional capabilities will likely include the following component services and infrastructure: - Security Labeling Manager (SLM) - Security Labeling Service (SLS) - Trust Fabric Services - Security and Privacy Ontology Based Terminology Services - Privacy Protective Services [17]. The definitions have been used for instance in combination with a XAMCL based access control system in the Axle EU project [18] (presented in [19]).

S4: HL7 Version 3 Standard: Healthcare (Security and Privacy) Access Control Catalog, Release 3

Provides HL7 permission vocabulary in constructing permissions {operation, object} pairs supporting Role based Access Control (RBAC) as well as definition additional high-level concepts and vocabulary of Attribute-Based Access Control (ABAC). [20]

S5: HL7 Version 3 Standard: Privacy, Access and Security Services (PASS); Access Control, Release 1

This standard offers a conceptual framework for access control services applicable to Privacy, Access, and Security domains within the healthcare environment and therefore enabling the creation of standard based services and capabilities implementing the policy framework of any domain. [21]

S6: HL7 Version 3 Standard: Privacy and Security Architecture Framework - Trust Framework for Federated Authorization, Release 1

The document defines a trust framework model for federated authorization by presenting a policy driven approach for controlling access to and use of information across security domains. It shows a high-level harmonized view of the trust, security and privacy policy, and information required to support the interoperability needs of healthcare providers. The policies are negotiated (harmonized) in real-time by participating domains through a process called Policy Bridging, and agreed to via a trust contract also established at run time [22].

2.3 HL7 CDA R2 Implementation Guides

HL7 provides several CDA R2 implementation guides that can be taken into account for implementing GDPR compliant systems.

CDA1: HL7 CDA® R2 Implementation Guide: Privacy Consent Directives, Release 1

The implementation guide specifies templates for a CDA document to exchange signed Consent Directives, which can be represented as a narrative, signed document, and computable statements/entries using standard-based terminology. Thus, it can be eventually used to generate enforceable assertions or rules (e.g. SAML, XACML) [23].

CDA2: HL7 CDA® R2 Implementation Guide: Data Provenance, Release 1 - US Realm

This implementation guide is a result of a collaboration of HL7 and the US Health and Human Services Office of National Coordinators Standards and Interoperability Framework Data Provenance Initiative (DPROV) and enables basic provenance information about clinical (and other care related information), who created it, when was it created, where was it created, how it was created, and why it was created, to be integrated into HL7 CDA documents in a consistent and interoperable way [24].

CDA3: HL7 Implementation Guide: Data Segmentation for Privacy (DS4P), Release 1

The DS4P implementation guide defines models, concepts and templates to enable segmenting clinical records so that personally identified information (PII) can be appropriately shared as may be permitted by privacy policies or regulations. It introduces reusable privacy building blocks and CDA templates to support the association of information object (e.g. document, section, entry) with security labels, which then can be linked to privacy policies [25].

CDA4: HL7 CDA® R2 Implementation Guide: Patient-Friendly Language for Consumer User Interfaces, Release 1

This standard provides a patient friendly language/ plain language healthcare vocabulary which is targeted specifically toward healthcare consumer user interfaces which create outputs for consumer consumption such as consent directives, reports of disclosures, and notices of privacy practices [26]. It specifies a mapping of the technical/ legal security and privacy language, which patients are often uncomfortable with, to a plain language, which is defined as "communication your audience can understand the first time they read or hear it" (see [27] for complete definition) [26]. The standard provides a mapping to the English language, but can be a template for any other language.

2.4 HL7 FHIR® Artifacts

The HL7 FHIR specification [28] provides a summery page regarding security and privacy principles providing guidance and also an overview on FHIR specific controls [29]. According to the GDPR requirements the following FHIR elements can be considered relevant:

FHIR1: Security Labels

According to the "Healthcare Privacy and Security Classification System (HCS)" [15] the FHIR specification supports the implementation of the security label concept. Security and privacy labels can be attached to a resource or bundle as metadata to provide specific security and privacy attributes and information. Details can be found at [30]. Currently three core Security Labels are defined: **Context of Use** - Purpose Of Use, **Data Sensitivity** - Confidentiality codes, **Control of Flow** - Delete After Use / Do Not Re-use, but supports all categories defined by HCS. Figure 2 shows an example of applying a security label to a FHIR resource [30].

FHIR2: Compartment Resource

The compartment resource was designed to logically group resources which share common properties [31]. Thus, it can be used as the basis for applying access control mechanisms. Currently defined compartments are Patient, Encounter, RelatedPerson, Practitioner and Device. For example a Patient compartment is set of resources associated with a particular patient [32]. For detailed definition and usage of the compartment resource refer to [31].

FHIR3: Consent Resource

The purpose of the consent resource is to enable expressing specific consents regarding healthcare. This can be for instance a Privacy Consent Directive, Medical Treatment Consent Directive, Research Consent Directive or Advance Care Directives [33]. It primarily serves „*as record of a healthcare consumer's policy choices, which permits or denies identified recipient(s) or recipient role(s) to perform one or more actions within a given policy context, for specific purposes and periods of time.*" [33]. But based on different characteristics (e.g. human readability or legal binding) there are several definitions how this resource can be instantiated and used. For detailed descriptions as well as examples see [33].

FHIR3: Provenance Resource

FHIR defines two resources suitable for tracking the origins, authorship, history, status, and access of resources. The Provenance resource enables recording of activities regarding specific resources (entities). This may include consumption, processing, transformation, modification, relocation, usage, or generation of entities. It allows for tracking what has happened to a specific entity in the course of time (i.e. who created it when, who modified or transformed it, etc.). The provenance resource model definition is in alignment with the W3C provenance specification [34] (see Figure 3). The Provenance resource covers "Generation" of "Entity" with respect to FHIR defined resources for creation or updating; whereas AuditEvent (see following section) covers "Usage" of "Entity" and all other "Activity" as defined in W3C Provenance [35].

FHIR4: Audit Event Resource

The Audit event resource is the second resource for tracking activities of specific entities and provides a record of an event made for purposes of maintaining a security log [36].

2.5 HL7 Version 2

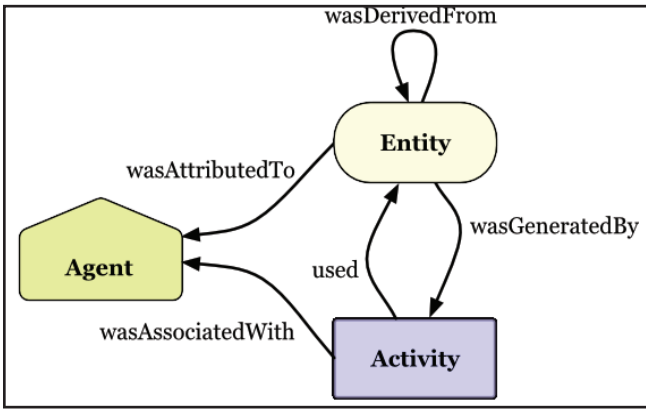
Even though HL7 Version 2 [37] states in its introduction that it does not explicitly provide elements for security and privacy, it offers some elements that can be used to convey information to support also some of the requirements of the GDPR.

V2: CON Segment

The CON segment "*identifies patient consent information relating to a particular message. It can be used as part of existing messages to convey information about patient consent to procedures, admissions, information release/exchange or other events discussed by the message. It may also be used in messages focusing on recording or requesting consent and for*

```
<Patient xmlns="http://hl7.org/fhir">
  <meta>
    <security>
      <system value="http://hl7.org/fhir/v3/Confidentiality"/>
      <code value="R"/>
      <display value="Restricted"/>
    </security>
  </meta>
...  [snip] ...
</Patient>
```

Figure 2: Example for security label as meta-data for a FHIR resource.



Target - An entity that is a FHIR resource instance that is created, updated or deleted.
Entity - An entity is a physical, digital, conceptual or other kind of thing with some fixed aspects; entities may be real or imaginary.
Agent - An agent is something that bears some form of responsibility for an activity taking place, for the existence of an entity, or for another agent's activity.
Activity - An activity is something that occurs over a time period and acts upon or with entities. It may include consuming, processing, transforming, modifying, relocating, using, or generating entities.

Figure 3: Provenance model based on W3C provenance specification [35].

Table 1: Mapping HL7 artifacts to GDPR requirements.

	R1 Priv.by Design	R2 portability	R3 right to be forgotten	R4 consent	R5 privacy notices	R6 right to access	R7 explicit policies
S1 (DAM)	x	x		x			x
S2 (HCS)	x	x					
S3 (SLS)	x	x					
S4 (HACC)	x	x					x
S5 (PASS-AC)	x	x					x
S6 (PSAF-AuthZ)		x					x
CDA1 (consent)				x	x	x	x
CDA2 (prov.)						x	
CDA3 (segment.)	x					x	
CDA4 (language)					(x[1])		
FHIR1 (labels)	x						
FHIR2 (consent)		x		x	x		x
FHIR3 (prov.)		x				x	
FHIR4 (audit)						x	
V2 (CON)				x			

[1]As it is provided only in English language it can only serve as template!

consents related to employees or service providers." [37]. While the purpose is mainly for consent to medical treatment it can be also used to express authorization to disclosure protected health information (consent type = 001). [37]

3 Results

There is no doubt, that the GDPR on the one side requires the use of international interoperability standards for systems operating in the healthcare domain and on the other side poses the obligation to implement an well documented appropriate access control system to protect private and especially sensitive data. Table 1 shows the mapping of the aforementioned HL7 standards and definitions to the GDPR core requirements.

4 Discussions

It could be shown, that consequently using HL7 security and privacy standards and components efficiently helps to implement the technical core requirement of the GDPR. Unfortunately many systems in the healthcare domain (EHR as well as PHR systems, especially when it comes to the thousands of applications for mobile devices) are far away from implementing the appropriate controls and even worse many companies have not even realized the absolute necessity to start planning to reach compliance with GDPR in 2018.

However, most HL7 specifications still focus on the IT systems interoperability based on ICT ontologies [38]. For overcoming this limitation and responding to the social, cultural, knowledge and language related requirements of the GDPR, we have to extend the interoperability scope beyond the ICT domain, also directly including non-ICT domains and specialties and their terminologies and ontologies based on the aforementioned Interoperability Reference Architecture Model. Pushed by the crucial impact of multiple non-ICT domains, the HL7 Security Working Group has moved quite early to a system-oriented, architecture-centric, ontology-

based approach to interoperability, also supported by ISO and CEN specs following the same approach. We still hope that other HL7 WGs will adapt that approach quite soon. For deeper reasoning on this, see [39].

References

[1] European Parliament and Council: Regulation (EU) 2016/679 – Summary Page. Online: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC

[2] European Parliament and Council: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016. Online: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

[3] European Commission: Protection of personal data. Online: http://ec.europa.eu/justice/data-protection/

[4] European Parliament and Council: Directive 95/46/EC. 1995, Online: http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31995L0046

[5] European Commission: Reform of EU data protection rules. Online: http://ec.europa.eu/justice/data-protection/reform/index_en.htm

[6] European Parliament and Council: Directive (EU) 2016/1148. Online: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.194.01.0001.01.ENG&toc=OJ:L:2016:194:TOC

[7] Wikipedia, General Data Protection Regulation. https://en.wikipedia.org/wiki/General_Data_Protection_Regulation

[8] Health Level 7 International, Inc., Ann Arbor, USA. www.hl7.org

[9] Bobel B, Lopez DM, Gonzalez C. Patient privacy and security concerns on big data for personalized medicine. Health and Technol. 2016; 6: 75-81.

[10] European Commission – Articel 29 Data Protection Working Party: Guidelines on the right to data portability. Online: ec.europa.eu/newsroom/document.cfm?doc_id=44099

[11] Blobel B, Ruotsalainen P, González C, López D. Policy-driven management of personal health information for enhancing interoperability. Stud Health Technol Inform. 2014; 205:463–467.

[12] International Organization for Standardization. ISO 13606-1 EHR communication – Reference model. Geneva: ISO; 2017.

[13] HL7 International Inc. HL7 V3 DAM: Composite Security and Privacy Domain Analysis Model – Release 1. Ann Arbor: HL7 International; 2014.

[14] International Organization for Standardization. ISO 22600 Health informatics – Privilege management and access control. Geneva: ISO; 2014.

[15] HL7 International Inc. HL7 Healthcare Privacy and Security Classification System (HCS), Release 1. Ann Arbor: HL7 International, Online: http://www.hl7.org/documentcenter/private/standards/v3/PRIV_SEC_CLASS_SYS_R1_2014AUG.zip

[16] Data Segmentation for Privacy VA/SAMHSA RI/Pilot at HIMSS 2013. Online: https://www.healthit.gov/techlab/ipg/node/4/submission/82

[17] HL7 International Inc. HL7 Version 3 Standard: Privacy, Access and Security Services; Security Labeling Service, Release 1 (SLS). Ann Arbor: HL7 International, Online: http://www.hl7.org/documentcenter/private/standards/v3/V3_SECURITY_LABELSRV_R1_2014JUN.ZIP

[18] Axle Project: Advanced Analytics for Extremely Large European Databases. https://axleproject.eu/

[19] Meijer H-J, Gaba A, Havinga Y. Axle Projekt. Privacy-aware analytics on healthcare data. Online: http://nl.portavita.com/sites/default/files/whitepapers/Presentation-Portavita-HL7-Workgroup-Meeting-January-2014.pdf

[20] HL7 International Inc. HL7 Version 3 Standard: Healthcare (Security and Privacy) Access Control Catalog, Release 3. Ann Arbor: HL7 International, Online: http://www.hl7.org/documentcenter/private/standards/v3/HL7_V3_HACC_R3_2016OCT.pdf

[21] HL7 International Inc. HL7 Version 3 Standard: Privacy, Access and Security Services (PASS); Access Control, Release 1. Ann Arbor: HL7 International, Online: http://www.hl7.org/documentcenter/private/standards/v3/V3_PASS_AC_R1_2017JAN.pdf

[22] HL7 International Inc. HL7 Version 3 Standard: Privacy and Security Architecture Framework - Trust Framework for Federated Authorization, Release 1. Ann Arbor: HL7 International, Online: http://www.hl7.org/v3ballotarchive_temp_F8D8D58E-1C23-BA17-0CC585AA2A888F40/v3ballot/html/infrastructure/security/V3_PSAF_R1_I2_2017MAY.zip

[23] HL7 International Inc. HL7 CDA® R2 Implementation Guide: Privacy Consent Directives, Release 1. Ann Arbor: HL7 International, Online: http://www.hl7.org/documentcenter/private/standards/cda/CDAR2_IG_CONSENTDIR_R1_2017JAN.zip

[24] HL7 International Inc. HL7 CDA® R2 Implementation Guide: Data Provenance, Release 1. Ann Arbor: HL7 International, Online: http://www.hl7.org/documentcenter/public/standards/dstu/CDAR2_IG_DATAPROV_R1_DSTU_2016SEP.zip

[25] HL7 International Inc. HL7 Implementation Guide: Data Segmentation for Privacy (DS4P), Release 1. Ann Arbor: HL7 International, Online: http://www.hl7.org/documentcenter/private/standards/v3/HL7_V3_IG_DS4P_R1_2014MAY.zip

[26] HL7 International Inc. HL7 CDA® R2 Implementation Guide: Patient-Friendly Language for Consumer User Interfaces, Release 1. Ann Arbor: HL7 International, Online.

[27] The Plain Language Action and Information Network (PLAIN): What is plain language. Online: http://www.plainlanguage.gov/whatisPL/index.cfm

[28] HL7 International Inc. Fast Healthcare Interoperability Resources Release 3(STU). Online: http://fhir.hl7.org

[29] HL7 International Inc. FHIR Release 3 (STU) Security. Online: http://hl7.org/implement/standards/fhir/security.html

[30] HL7 International Inc. FHIR Release 3 (STU) Security Labels. Online: http://hl7.org/implement/standards/fhir/security-labels.html

[31] HL7 International Inc. FHIR Release 3 (STU) Resource Compartment Definition – Content. Online: http://hl7.org/implement/standards/fhir/compartmentdefinition.html

[32] HL7 International Inc. FHIR Release 3 (STU) Compartment Patient. Online: http://hl7.org/implement/standards/fhir/compartmentdefinition-patient.html

[33] HL7 International Inc. FHIR Release 3 (STU) Resource Consent – Content. Online: http://hl7.org/implement/standards/fhir/consent.html

[34] World Wide Web Consortium (W3C). PROV-Overview. Online: https://www.w3.org/TR/2013/NOTE-prov-overview-20130430

[35] HL7 International Inc.: FHIR Release 3 (STU) Resource Provenance – Content. Online: http://hl7.org/implement/standards/fhir/provenance.html

[36] HL7 International Inc.: FHIR Release 3 (STU) Resource AuditEvent – Content. Online: http://hl7.org/implement/standards/fhir/auditevent.html

[37] HL7 International Inc: HL7 Version 2.8.2. Ann Arbor: HL7 International, 2014, Online: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=403

[38] A. Akerman, J. Tyree. Using ontology to support development of software architectures. IBM Systems Journal. 2006; 45:813-825.

[39] B. Blobel, F. Oemig.The Importance of Architectures for Interoperability. Stud Health Technol Inform. 2015; 211:18-56.

17th INTERNATIONAL HL7 INTEROPERABILITY CONFERENCE IHIC 2017

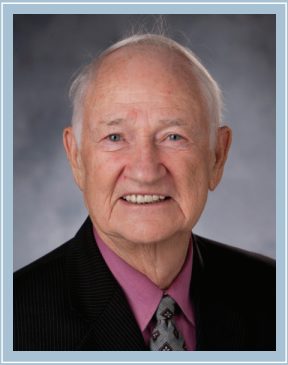
Technopolis
City of Athens
22-24 October
2017

17th International HL7 Interoperability Conference (IHIC 2017)

The 17th International HL7 Interoperability Conference (IHIC 2017) will be held from 22-24 October 2017 in Athens, Greece. The meeting will be hosted by HL7 Hellas, with support from HL7 Germany as permanent IHIC supporter - other HL7 Affiliates have been invited to join the supporters' team.

Keynote Speakers

Prof. Bernd Blobel
PhD, FACMI, FACHI, FHL7, FEFMI,
University of Regensburg, HL7 Germany



Prof. William E. Hammond
FACMI, FAIMBE, FIMIA, FHL7, Duke
University, HL7 International-US

Gora Datta
CAL2CAL Corporation
HL7 International-US

Exchange and Reconciliation of Clinical Decision Support Outputs across Systems for Coordinated Quality Improvement: Results and Future Direction from an Implementation in a U.S. Population Health Partnership

Junqiao Chen[1], Aldo Tinoco[1], Lauren Drinkard[1], Diane Hunt[2], Gregory Stevens[2], John Calhoun[2], Jason Cassidy[1], Epson Chiang[1] and Jesse James[1]

[1] Evolent Health, Virginia, 22203, USA

[2] Deaconess Health System, Indiana, 47710, USA

Abstract

Background: In a population health partnership, shareable clinical decision support (CDS) can reduce service duplication and promote patient wellness by presenting consistent information to all members of a cross-functional, distributed care team. However, existing health IT standards present challenges in the exchange of CDS outputs from multiple systems.

Objectives: To exchange and reconcile CDS outputs across systems in a partnership between an integrated health system (Deaconess Health System) and a population health services organization (Evolent Health).

Methods: We developed a bi-directional HL7-based interface for CDS outputs between Deaconess's Electronic Health Record (EHR) and Evolent's population health services platform (PHSP). The mapping of CDS logic between systems enabled this interface to automatically reconcile inconsistent CDS outputs. Fifteen quality measure (QM)-based CDS rules to identify care gaps were selected for this initiative. These care gaps remind Deaconess's clinicians to provide care or documentation necessary for gap closure, and also guide Evolent's care management services.

Results: Two months after launch, Deaconess reconciled 14,040 care gaps from Evolent using data only available in the EHR. Additionally, 1,047 care gaps were resolved due to patient refusal or clinical inappropriateness, and 246 gaps were closed by services or prescriptions provided during clinical encounters.

Conclusions: We implemented an HL7-based interface to exchange and reconcile a large volume of CDS outputs between a health system EHR and a PHSP. Future direction is to standardize the linkage between a CDS rule and its reference QM by universal identifiers and a taxonomy of variations.

Keywords

Population health; Health information exchange; Quality improvement; Clinical decision support; Health level seven

Correspondence to:

Dr. Jesse James
Evolent Health, 800 N Glebe Road, Suite 500,
Arlington, Virginia, 22203, USA.
E-mail: jjames@evolenthealth.com

EJBI 2017; 13(1):35-42
received: June 11, 2017
accepted: July 19, 2017
published: October 10, 2017

1 Background

Despite significant investment and effort during the last two decades, progress towards shareable clinical decision support (CDS) has been slow [1]. Existing standards to share CDS artifacts or provide CDS guidance service have not fully meet the need for shareable CDS in a population health partnership. Thus, we developed and implemented a novel bi-directional HL7-based interface to exchange and reconcile CDS outputs. The population health management market is anticipated to have more than fourfold increase in value by 2025, with growth not only in North America but the fastest in Asia-Pacific [2]. With this growth, we anticipate a concurrent increase in the integration of Electronic Health Record (EHR) systems and population health services platforms (PHSP), as well as the need to reconcile CDS outputs from each systems.

1.1 Deaconess Health System-Evolent Health Partnership

Deaconess Health System (Evansville, Indiana) is an integrated health system of six hospitals, fifteen multispecialty

community clinics, and a network of community-based providers. Since 2009, Deaconess Health System has used the Epic EHR system (Madison, Wisconsin) with integrated CDS capability. Deaconess Health System has partnered with Evolent Health (Arlington, Virginia), a population health services organization, to assist their transformation from traditional fee-for-service to value-based care. Evolent Health's care management programs are designed to engage patients in collaborative management of their medical, social, and mental health needs, with the goal to optimize wellness and coordinate care. Evolent Health's care managers use its PHSP with integrated CDS capability to identify gaps in care based on quality measures (QM). Care managers and clinicians use the PHSP to review and address care gaps. Because each CDS system has access to different data about patients in the health systems, these two CDS engines together yield a more holistic and comprehensive picture of patients' clinical needs than either system alone.

Deaconess Health System and Evolent Health set out to improve the use of CDS-guided coordination of care. In this partnership, CDS guidance is generated by each of the health IT systems: the EHR system and the PHSP. To ensure the consistency of CDS output presented to users at both Deaconess Health System and Evolent Health, implementation team members examined how to share and reconcile CDS output from both systems.

1.2 Challenge and Solution

Several initiatives to accelerate the sharing of CDS have been carried out. For example, the US Health eDecision initiative aims to create standards for CDS artifact sharing and CDS guidance services. CDS artifact sharing is the representation of logic in an electronic format so the receiving application can run it directly; CDS guidance service is the representation of data with a CDS service in exchange for the output of CDS (e.g., care advice) to the requesting system [3].

In the Deaconess-Evolent partnership, the implementation of PHSP alongside the existing EHR system presented challenges that prevented the complete adoption of a traditional approach to share CDS. Integrated health organizations have had systems with locally developed, locally maintained, or commercial-off-the-shelf CDS artifacts deployed in many different representations, based on a variety of data models, and highly contextualized to local settings and workflow [1]. Translating them into a standard format and implementing them across the partnership's care continuum would have taken significant time and effort.

Challenges also appeared in adopting CDS guidance services, which are based on a role-based model in which there is a CDS requester and a responder. In our partnership, both the Deaconess EHR system and the PHSP have had their own CDS engine in place and both have had access to data from different sources. Both system databases contain laboratory result data; however, the Epic EHR system's CDS has had access to provider documentation (e.g., patient refusals for clinical preventive services) and clinical observation data (e.g., vitals and point-of-care lab test results). The PHSP's CDS has had access to medical and pharmacy claims data. Thus, even if two engines were to have the same clinical logic, they might still return different CDS results.

The overarching goal of this project was to exchange and reconcile CDS outputs from both systems, especially when the CDS output varies for a given patient. We developed a solution to address this need: a bi-directional HL7-based interface of CDS outputs supported by quality measure-enabled integration and backend reconciliation logic.

The need to align CDS and QM has been widely recognized by HL7 [4], National Quality Forum [5] and US federal agencies [6, 7]. This alignment has become imperative in this current time of US health care reform where providers are being judged by their ability to manage population health defined by QM. The harmonization of QM in our work was partially inspired by the Clinical Decision Support-Quality Measure (CDS-QM) framework pioneered at the University of Utah [8]. In 2014, Kukhareva and colleagues at the University of Utah pioneered a prototype of a standards-based CDS framework aligned with electronic quality measurements [8]. This prototype converted data in a data warehouse into the HL7 virtual medical record (vMR) format and transmitted it to OpenCDS (Figure 1). OpenCDS in turn sent CDS guidance back to the data warehouse.

In this article, we describe the implementation of vital components of the Utah CDS-QM framework in the Deaconess-Evolent partnership. We then discuss the design of the bi-directional, HL7-based interface and the downstream workflow enabled by it.

2 Methods

The Deaconess-Evolent Model shares some features of the Utah CDS-QM Framework, but includes additional data sources and supports downstream connectivity with multiple systems, as described below (Figure 2).

2.1 From Data to CDS Outputs by Evolent

Firstly, structured data is loaded into Evolent's enterprise data warehouse (EDW) from multiple sources belonging to Deaconess and other entities (health insurers, independent laboratories, etc.). One noted extension to the Utah framework is the addition of pharmacy claims data, which offers important details about fill/refill dates and the actual

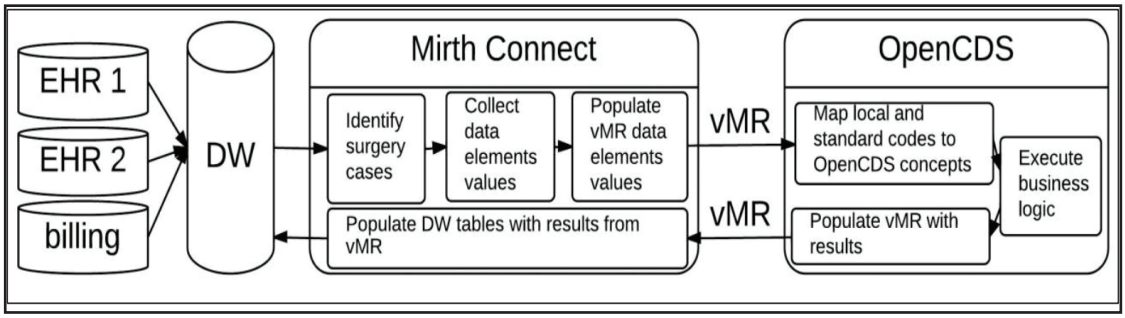


Figure 1: Major systems and processes involved in the Utah CDS-QM approach. Source: [8]. Re-used with permission from the American Medical Informatics Association.

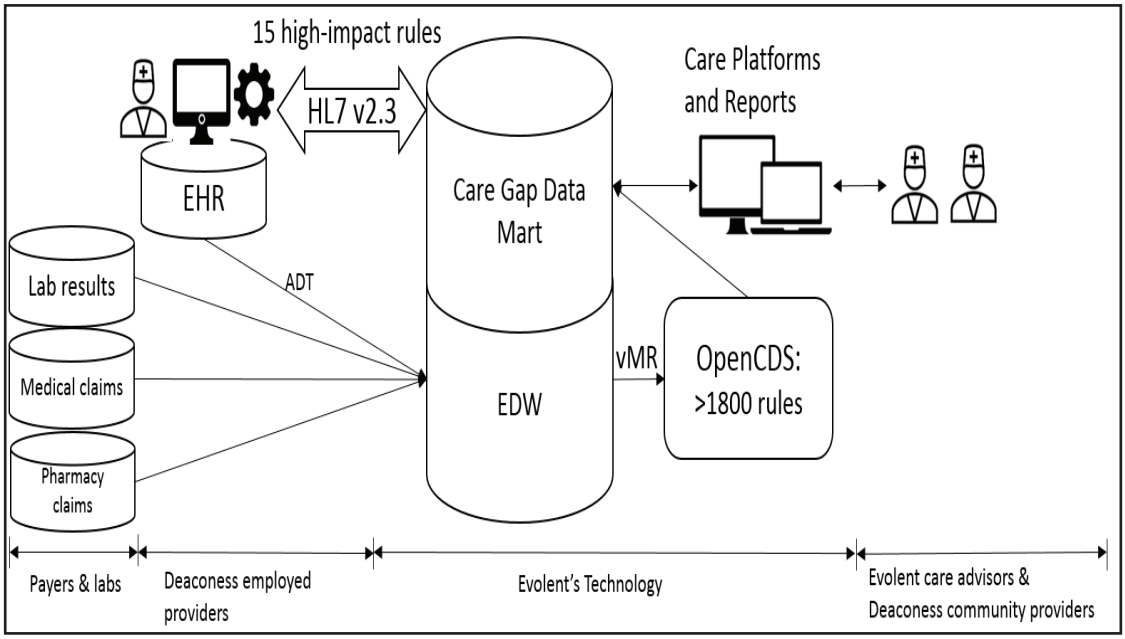


Figure 2: Major systems and processes involved in the Deaconess-Evolent enhanced CDS-QM implementation.

drug dispensed (brand, package, dosage, etc.). Loaded data is transformed into vMRs, which is then transmitted to an OpenCDS-based rules engine. Evolent Health staff developed rules to detect care gaps based on QM. The relationship between the reference QM, the derived care gap decision support rule and their internal unique identifiers are maintained in the rules engine's knowledge base. Output from the rules engine is stored in the data warehouse alongside metadata and other patient-level observations. Evolent's care managers and other users access the CDS outputs via the PHSP, which also capture interventions performed to close care gaps.

2.2 Exchange of CDS Outputs

The decision to build a bi-directional interface between Deaconess and Evolent was the result of a rigorous assessment performed by a cross-functional team. Following a physician-led solution design process, the Deaconess Physician Advisory Council and Outpatient Decision Support Committee met to design the project. A subset of CDS rules was selected for integration into the EHR. Selection criteria included clinical value, current perfomance level, physician impactability, technical readiness, and business priority. These fifteen "care gap" rules (Table 1) cover chronic disease management, clinical preventive services, medication monitoring, and pediatric access to care. For some of these CDS rules, the need of reconciliation is well-known due to factors like long look-back period for historical data [9].

Prior to the integration, eleven of these 15 rules existed in both Evolent and Deaconess platforms with different degrees of variations in logic. Four of these 15 rules heavily rely on pharmacy claims and were only available to Evolent prior the integration. Quality experts from Deaconess and Evolent Health compared the CDS rules from both systems with the corresponding QM specifications. Deconess modified its CDS rules, where necessary, to match logic across systems; thus, inconsistent rule results could be attributed to the differences in data accessible to each system rather than variation in the logic.

Reference Measure: the quality measure that the care gap rule is based upon; NQMC: National Quality Measures

Table 1: List of care gaps being exchanged and reconciled.

Care Gap Rules	National Identifiers of the Reference Measures
Appropriate Medications for Patients with Asthma	NQMC: 009940, or NQF:0036
Diabetes Care - HbA1c Test	NQMC: 010520, or NQF: 0057
Diabetes Care - Eye Exam	NQMC: 010524, or NQF: 0055
Diabetes Care - Nephropathy Screening	NQMC: 010525, or NQF: 0062
Breast Cancer Screening	NQMC: 009931, or NQF: 2372
Colorectal Cancer Screening	NQMC: 009933, or NQF: 0034
Cervical Cancer Screening	NQMC: 010930, or NQF: 0032
Chlamydia Screening	NQMC: 009934, or NQF: 0033
Annual Flu Vaccination	NQMC: 010565 & 010566, or NQF: 0039 & 0040
Pneumococcal Vaccination	NQMC: 010570, or NQF: 0043
Annual Monitoring for Patients on ACE/ARB	NQMC: 010542 Rate 1, or NQF: 2371 Rate 1
Annual Monitoring for Patients on Digoxin	NQMC: 010542 Rate 2, or NQF: 2371 Rate 2
Annual Monitoring for Patients on Diuretics	NQMC: 010542 Rate 3, or NQF: 2371 Rate 3
Well-Child Visits 3 to 6 years-old	NQMC: 010611, or NQF: 1516
Adolescent Well-Care Visits	NQMC: 010612

Clearinghouse; NQF: National Quality Forum; ACE/ARB: Angiotensin Converting Enzyme inhibitors or Angiotensin Receptor Blockers..

Deaconess maintained a crosswalk between the Evolent CDS rule ID and the clinically equivalent CDS rule in the EHR. Messages sent from the EHR system back to Evolent contain the original ID used by Evolent.

This interface encapsulates care gap messages in HL7 version 2.3 Admission, Discharge, Transfer (ADT) format. HL7 version 2 is used in 95% of US healthcare organizations [10]. Version 2.3, in particular, is the most widely accepted version by Evolent's partners. Using this standard leverages the existing interfaces in Deaconess's EHR and Evolent's platform. Also, since a significant portion of health systems in the US have installed the same EHR as Deaconess, lessons learned from this implementation will inform future development in other organizations.

There is no existing standard or implementation guide for the use of HL7 v2 segments for care gaps. Due to its extensibility and use for clinical information, we chose to use Observation (OBX) segment capabilities in ADT messages to represent care gaps. Each attribute of a care gap message is represented in one OBX segment, identified by an Observation Identifier. Attributes include: care gap status, Evolent rule ID, open and close dates of the care gap, service provider, and other supporting information. Upon receiving messages from Evolent, Deaconess utilizes the interface engine Cerner OPENLink [11] to process and route messages. Once a message is routed to the EHR, patient identity in the care gap message is matching to the appropriate EHR records by the EHR system.

2.3 System Reconciliation of Care Gaps

Automatic reconciliation Figure 3 is triggered when the care gap status is not the same between two systems. The goal is not just reconciling the status, but also prompting the care team to either provide care or data to support the closure of care gaps. The first step is to identify an inconsistent status. Because most care gaps have been tuned to represent the exact same clinical logics, inconsistent status of care gaps could be automatically identified by Deaconess's EHR. When the care gap status is "Open" in Evolent's messages but "Closed" in Deaconess's EHR, supporting data of the gap closure will be automatcially extracted from Deaconess's EHR and sent back to Evolent via the interface. This will update relevant records in Evolent's system.

2.4 Actions upon Open Care Gaps by Deaconess Providers

When the care gap status is "Closed" in Evolent but "Open" in Deaconess's EHR, providers are prompted to update the data in EHR to ensure a complete record of care gap closure. When the status is " "Open" in both systems, then a Deaconess provider attempts to offer the clinically appropriate service to close the identified gap or to document reasons why a particular type of service could not be offered/was declined. Providers perform these activites using the Health Maintanence module within the EHR. This is a routinely used module that alerts physicians to various patient needs. Utilizing this module enables a more seamless development and implementation. More importantly, it requires no workflow change to current practices.

2.5 Actions upon Open Care Gaps by Evolent Care Managers

Updated care gap statuses are automatically sent back to Evolent every 24 hours and reflected in Evolent's PHSP. Care managers can then address the updated and more

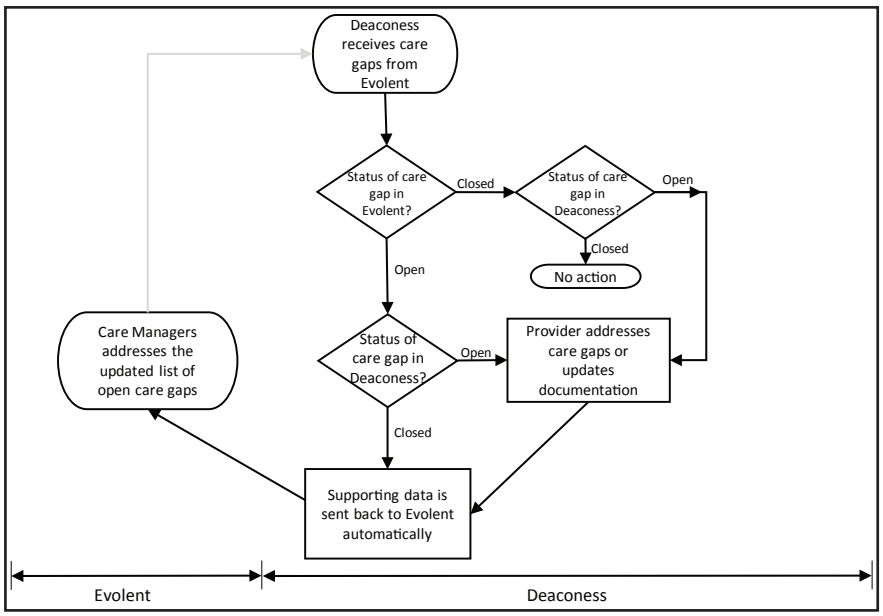


Figure 3: System reconciliation of and care team actions on care gaps.

accurate list of open care gaps by various patient engagement techniques and care management services.

3 Results

The bi-directional care gap interface was launched in March 2017. Although there was no change in workflow, provider engagement activities were carried out to inform them of the richer CDS content in the EHR. In the first two months of implementation, Evolent received 15,333 responses from Deaconess (Table 2) via this interface. 14,040 false positive care gaps in Evolent's system were automatically closed by data received from Deaconess via the interface. Additionally, 1,047 responses were either patient refusal or clinical inappropriateness, and 246 gaps were closed by procedures or prescriptions offered during clinical encounters.

Closing a true care gap requires time and resources to engage patients, schedule appointments and provide services. It is encouraging to observe that during the first two months of implementation, Deaconess providers intervened on 1300 care gaps with the help of reconciled CDS outputs. We anticipate this number to increase over time.

4 Discussion

Alignment and reconciliation of CDS "are developing areas in informatics that will need continued investigation as CDS implementations increase" [12]. Benefits include the reduction of waste (e.g., human effort and duplication of services), provider dissatisfaction, and confusion (e.g., basing outreach to patients on outdated information). Thus it enables a consistent user experience and increases the coordination of care across systems [13].

Table 2: Number of care gaps received by evolent from deaconess categorized by closure reasons; March to May, 2017.

Closed by offering care	Patient refused or not Appropriate	Closed by data only available in EHR	Total
246	1047	14040	15333

Evolent Health utilizes an open-source CDS-QM framework to turn QM into actionable CDS outputs to alert providers of gaps in patients' care. To reconcile these outputs with Deaconess Health System's practice EHR, an HL7-based bi-directional interface was built and implemented for 15 care gap rules selected by physician leaders. Due to the physician-led mapping of CDS logics between systems prior the implementation of the interface, care gaps were automatically reconciled using data that was previously inaccessible to Evolent Health.

This work presents a novel solution to electronically exchange and reconcile the output of different CDS systems via existing interface rather than build new interfaces to exchange raw observation data between the underlying data warehouses (e.g., vital sign values, lab results, medications, diagnoses, and procedures). A traditional approach to ensure that different CDS systems generate the same output would be to ensure that different systems have the same data input available to the same logic. However, adding additional clinical data exchange interfaces incurs significant expense. A survey of 125 organizations found that financial cost of building interfaces is the number one challenge to interoperability. In all, 112 organizations had to construct multiple interfaces, and 18 reported having to construct more than 25 interfaces [14]. This is burdensome work with substantial cost from developing and mapping multiple

interfaces, terminology mapping, quality assurance, and maintenance. Research has shown that only one-fifth of US hospitals engage in key domains of interoperability [15], and therefore significant care coordination gaps exist due to the lack of interoperability between health IT systems [16]. The fact that patient data is represented in various schemas and coded in different vocabularies across data sources is not only barrier to building a raw data interface, but also important challenges to clinical knowledge sharing. This is referred to as the "curly braces problem" in the space of HL7 Arden Syntax [17, 18]. While our approach uses information that is less granular than raw clinical observation data, it enables successful exchange and reconciliation of CDS outputs that add to the patient's clinical picture. It presents meaningful information for the end users.

We acknowledge three limitations to our approach to this implementation. First, the lack of a mechanism to reference a universal identifier for each care gap required time-consuming manual review of clinical logic between two organizations. This is a barrier to achieving scale when increasing the number of healthcare organizations or CDS rules in this exchange. Second, although HL7 v2 has wide adoption and fulfilled the needs of the project, it requires a significant amount of customization to fully achieve interoperability in each unique context. Additionally, even with all the necessary leg work, the message structure still tends to be complex, flat and delimited [19]. Examining the use of HL7 version 3 or Fast Healthcare Interoperability Resources (FHIR) standards is beyond the scope of this initial project, but the use of FHIR and other application programming interface-based exchange of CDS output should be discussed and explored by the industry. Finally, this interface does not always capture sufficient detail to with other related use cases that require data at the observation level, such as the regulatory requirement of QM reporting. Our specific use case is to deliver care gaps within a workflow tool to proactively and prospectively improve the quality of care; this is certainly related to but not the same as QM reporting programs. Additional configuration (e.g. mandating physician's detailed input of all required data elements and attestation) will be needed for other use cases.

5 Future Work

Future research is needed to support the sharing of CDS output between health IT systems. Care gap detection logic is derived from either QM or from the clinical practice guidelines that the QM are based on. How can systems consistently manage the relationships between CDS rules and the reference QM from which it was derived? There are several situations in which a CDS content developer chooses to modify rule logic: for example, to address local workflow needs, to adapt to unique data models, and to allow locally defined data values. How should rule developers and implementers classify variations introduced in CDS rules from the reference QM and from other similar CDS rules? Since CDS output is used in a variety of settings across the continuum of care, what attributes are required to inform or proscribe interventions to close care gaps that are relevant to the setting of their use (e.g., clinical setting, care management setting, or directed to patients themselves)? Findings from research in these areas will inform the development of exchange standards to streamline the implementation of CDS output deployed across the various settings of care, enabling proactive population health management.

5.1 Universal Quality Measure Identifier System

This implementation highlights the urgency of establishing a formal identifier system in HL7 for QM and derivative CDS logic. In our example, a challenge rests within the receiving entity (Deaconess's EHR) to understand, process and correctly display care gaps in the user interface. It required a group of experts to manually review the care gap logic from the sending entity (Evolent), and where applicable, map to or modify existing internal rules within the EHR. This is not a new challenge, as Greenes and colleagues [20] demonstrated a similar issue in a previous informatics project (Morningside Initiative, 2008).

The most direct solution is to develop a mechanism to reference a centralized, standardized and authoritative identifier system for QM, which could enable the identification of reference measures in CDS-QM related development [21]. A major barrier is that there is no single universal identifier system for QM and CDS in the United States. For example, the National Quality Forum (NQF) has a numbered inventory of 1,086 measures [22], and the National Quality Measures Clearinghouse (NQMC) has a numbered inventory of 2,342 measures [23]. The distributed nature of the U.S. healthcare industry leaves no single entity responsible for maintaining a complete set of QM publicly available. Furthermore, the existing major repositories do not support a crosswalk between their identifiers. Internationally, this may be less a barrier in countries with centralized healthcare systems. China, for example, has been maintaining a single Chinese Healthcare Quality Indicators System (CHQIS) since 2009 with more than 5,000 numbered indicators in scientific hierarchies [24].

The need to unambiguously reference a QM has been recognized by HL7 in its recent development of FHIR. In the Measure resource of FHIR [25], there is an "Identifier system" element, which refers to one or more measure development bodies like NQF by a Uniform Resource Locator (URL). Currently there is no official identifier system established in FHIR yet, but it is understood that work is in progress (Bryn Rhodes, personal communication, 2017 May 23). An excerpt from FHIR website is shown below Figure 4. A similar approach might be applied to HL7 v2 in its extendable segments as well.

Id:
Measure/measure-cms146-example
Identifier:
system: http://hl7.org/fhir/cqi/ecqm/Measure/Identifier/cms
value: 146
Identifier:
system: http://hl7.org/fhir/cqi/ecqm/Measure/Identifier/nqf
value: 0002
Title:
Appropriate Testing for Children with Pharyngitis
Status:
active

Figure 4: "Identifier System" example from FHIR. Source: [26].

5.2 Taxonomy to Classify Variations from Reference Measure

If the above reference system is in place, then the next question is how to represent variations from the "reference measure" in a way that computers can interpret. This is very difficult due to the complexity and variabilitvy in changes made to the reference measure across systems.

Yauch and colleagues drafted a taxonomy to "capture the types of variations introduced in CDS rules as they are deployed to a specific setting" [27]. The first level of its taxa includes Threshold Factor, Timing Factor, Setting Factor, Event, Contraindicating Factor, Intervention, CDS Destination, Event Inducing Action, and CDS Realization Method. Capturing information like this will add to the complexity of initial implementation. The benefit of doing this could be realized as CDS output is increasingly exchanged. As an industry, we should achieve consensus on a first, core set of variation types, and then build upon this initial set to expand the taxonomy, covering all necessary variations. The allowed degree and type of variation should be carefully examined. It is recommended that if the variation in CDS rule greatly differs from the reference QM, it may be better to work with the QM steward to create a new measure corresponding to that CDS rule [21].

5.3 Contexts-Aware Attributes of a Care Gap

While the above two topics focus on the semantic exchange and reconciliation of care gaps, we must also consider what supporting information is necessary to selectively target CDS output to the right recipients based on their respective roles, their scope of practice, and per the partnership agreements made between the health care organizations involved in CDS output exchange.

Care managers who currently use the PHSP suggested the creation of context awareness attributes - "action by" and "action type" – to indicate the type of providers who should perform the type of action necessary to close a care gap. Take the Diabetes Eye Exam care gap as an example. It should have two sets of attributes: first, [action by] = "ophthalmologist or optometrist" and [action type] = "perform a retinal exam", and second, [action by] = "non-ophthalmologist or optometrist" and [action type] = "refer this patient to ophthalmologist or optometrist". Depending on the specialty of the provider offering the service ("context-aware"), one of the "action type" values will be shown. A standard, core set of attributes can be developed after the implementation of additional care gap rules. Broad participation in industry is essential in this adventure.

6 Conclusion

We expanded an HL7-based CDS and QM framework and applies it in a population health partnership. A novel bi-directional HL7-based care gap interface is built to exchange and reconcile CDS outputs ("care gaps") between a commercial EHR and a PHSP. Preliminary data shows that the interface has automatically reconciled false care gap alerts across organizations in the short time it has been implemented. This is an effective, HL7 standards-enabled, pragmatic implementation in a complex distributed care landscape. Future work is proposed to automate semantic mapping of CDS outputs between organizations by both a universal identifier system and a taxonomy of variations, and to standardize contextual attributes to better inform responses to care gaps.

7 Acknowledgements

We thank all the colleagues who contributed to the design, development, implementation and continuous maintenance of this interface. We thank Mr. Nicholas Dus, Mr. Rami ElGawly, Ms. Rahel Ephrem, Ms. Connie Weber, Mr. Milesh Patel and Mr. Robert Goska, for their insightful comments on earlier versions of the manuscript.

References

[1] Greenes RA, Peleg M, Rector A, Osheroff J. Reusable Knowledge for Best Clinical Practices: Why We Have Difficulty Sharing And What We Can Do. InMedinfo 2013;1237.

[2] Population Health Management (PHM) Market 2014 - 2025. Grand View Research, 2016.

[3] Fridsma, D. Two Significant S&I Framework Milestones: the Health eDecision Initiative Close-Out and Clinical Quality Framework Launch. 2014. Available from: https://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/interoperability-electronic-health-and-medical-records/standards-interoperability-framework-milestones-health-edecision-initiative-closeout-clinical-quality-framework-launch

[4] Health Level Seven International, Inc. Quality Improvement Core (QI-Core) Implementation Guide [Internet]. Ann Arbor, Michigan: Health Level Seven International; 2016 May 31 [cited 2017 May 23]. Available from: http://www.hl7.org/fhir/current/qicore/

[5] Driving Quality and Performance Measurement - A Foundation for Clinical Decision Support: A Consensus Report. National Quality Forum. 2010.

[6] Clinical Decision Support: More than Just 'Alerts'. Centers for Medicare and Medicaid Services, 2014.

[7] Berner, E. S. Clinical Decision Support Systems: State of the Art. Agency for Healthcare Research and Quality; 2009; Report No.: 09-0069-EF

[8] Kukhareva PV, Kawamoto K, Shields DE, Barfuss DT, Halley AM, Tippetts TJ, et al. Clinical Decision Support-based Quality Measurement (CDS-QM) Framework: Prototype Implementation, Evaluation, and Future Directions. InAMIA Annual Symposium Proceedings.American Medical Informatics Association, 2014.

[9] Wagholikar K, Sohn S, Wu S, Kaggal V, Buehler S, Greenes RA, et al. Workflow-based Data Reconciliation for Clinical Decision Support: Case of Colorectal Cancer Screening and Surveillance. AMIA Summits on Translational Science Proceedings. 2013 March 18-22; San Francisco, USA. Washington DC; 2013.

[10] Health Level Seven International. HL7 Version 2 Product Suite [Internet]. Ann Arbor, Michigan: Health Level Seven International; 2017 March 21 [cited 2017 Jun 5]. Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=185

[11] Cerner OPENLink [Internet]. Cerner Corporation. 2009. Available from: http://www.cerner.com/page.aspx?pageid=17179878387&libID=17179878602

[12] Tso GJ, Tu SW, Oshiro C, Martins S, Ashcraft M, Yuen KW, et al. Automating Guidelines for Clinical Decision Support: Knowledge Engineering and Implementation. AMIA Annual Symposium Proceeding; 2016 Nov 12-16; Chicago, USA. Washington DC: American Medical Informatics Association; 2016.

[13] Asanjan SMH. Consolidation of CDA-Based Documents from Multiple Sources: A Modular Approach [dissertation]. 2016.

[14] Hall SD. Interoperability remains a barrier for health data exchange organizations [Internet]. In: FierceHealthcare [Internet]. 2014. Available from http://www.fiercehealthcare.com/it/interoperability-remains-a-barrier-for-health-data-exchange-organizations

[15] Holmgren AJ, Patel V, Charles D, Adler-Milstein J. US Hospital Engagement in Core Domains of Interoperability. Am J of Manag Care. 2016; 22: e395-e402.

[16] Samal L, Dykes PC, Greenberg JO, Hasan O, Venkatesh AK, Volk LA, Bates DW. Care coordination gaps due to lack of interoperability in the United States: a qualitative study and literature review. BMC Health Serv Res. 2016; 16:143.

[17] Health Level Seven International, Inc. HL7 Cross-Paradigm Specification: Clinical Quality Language, Release 1 [Internet]. Ann Arbor, Michigan: Health Level Seven International; 2016 July [cited 2017 Jun 8]. Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=400

[18] Greenes RA. Clinical Decision Support: The Road to Broad Adoption. 2nd ed. Amsterdam: Academic Press.2014; 930.

[19] Musal S. The Differences Between HL7 and FHIR (and Why You Should Care) [Internet]. 2016 In: CloudMine Blog [Internet]. 2016. Available from: http://blog.cloudmineinc.com/the-differences-between-hl7-and-fhir-and-why-you-should-care

[20] Greenes R, Bloomrosen M, Brown-Connolly NE, Curtis C, Detmer DE, Enberg R, et al. The morningside initiative: collaborative development of a knowledge repository to accelerate adoption of clinical decision support. Open Med Inform J. 2010; 4: 278-90.

[21] National Quality Forum. Variation in Measure Specifications: Sources and Mirigation Strategies (Final Report). 2016; Report No.: (ISBN) 978-1-68248-032-8. Contract No.: HHSM-500-2012-00009I.

[22] National Quality Forum. Quality Positioning System [Internet]. 2017. Available from: http://www.qualityforum.org/Qps/QpsTool.aspx

[23] National Quality Measures Clearinghouse [Internet].Agency for Healthcare Research and Quality. 2017. Available from: https://qualitymeasures.ahrq.gov/

[24] Zhao M, Liang, M, Yu R, Jian W, Zhang L, Ma W. Designing and application on CHQIS medical quality evaluation system. Chinese Hospitals. 2009; 13: 1-7.

[25] Health Level Seven International. Resource Measure – Content [Internet]. Ann Arbor, Michigan: Health Level Seven International; 2017 Jun [cited 2017 Jun 9]. Available from: http://build.fhir.org/measure.html

[26] Health Level Seven International. FHIR Example Instance [Internet]. Ann Arbor, Michigan: Health Level Seven International; 2017 Jun [cited 2017 Jun 9]. Available from: http://build.fhir.org/measure-cms146-example.xml

[27] Yauch D, Kum P, Tu S, Haug P, Hulse N, Fry E, et al. Development of a Taxonomy of Setting-Specific Factors for Adaptation of Clinical Decision Support Rules. AMIA Annual Symposium, American Medical Informatics Association, 2012.

A Terminology Service Compliant to CTS2 to Manage Semantics within the Regional HIE

Roberta Gazzarata[1], Maria Eugenia Monteverde[1], Elena Vio[2], ClaudioSaccavini[2], Lorenzo Gubian[3], Idelfo Borgo[4] and Mauro Giacomini[1,5,6]

[1] Healthropy, Corso Italia 15/6, 17100 Savona, Italy

[2] Arsenal.IT, Veneto's Research Centre for eHealth Innovation, Viale Oberdan 5, 31100 Treviso, Italy

[3] Veneto Region Health Information System, Palazzo Molin, S. Polo, 2514 – 30125 Venice, Italy

[4] Local Health Authority 16 of Padova, Via E. degli Scrovegni, 12, 35131 Padova, Italy

[5] Department of Informatics, Bioengineering, Robotics and System Engineering (DIBRIS), University of Genoa, Via Opera Pia 13, 16145 Genoa, Italy

[6] Centre of Excellence for Biomedical Research (CEBR), University of Genoa, Viale Benedetto XV 9, 16132 Genoa, Italy

Abstract

Objectives: The correct semantics management represents a fundamental requirement to create health information exchange (HIE) systems able to support clinical data sharing. A research is now underway to set up an infrastructure to aggregate data coming from health information systems, and it will be experimented to support regional HIE in Veneto region, Italy.

Methods: In the first period, the focus was on the semantics management of Clinical Document Architecture Release 2 (CDA R2) laboratory reports, in which observations must be encoded using LOINC® vocabulary. The existing components considered were the Laboratory Information Systems (LISs) of the local departments of the region (23 units) and the regional HIE. To manage the semantics of the data, the design and the implementation of a terminology service, the Health Terminology Service (HTS), compliant to the Common Terminology Service Release 2 (CTS2) standard was considered.

Results: The HTS, formed by a set of web services, as CTS2 indicates, is the core of the proposed infrastructure. It is connected with an application that allows creating, deleting, updating and managing the versioning of the maps between the laboratory observations and the LOINC® entities.

Conclusions: The adoption of CTS2 specification allows the HTS to support different use cases, which are essential to supports the clinical data sharing of the HIE content. During the development, the authors faced some problems that cause a delay in the implementation of the solution. The authors are still working to connect the HTS with the existing systems and to improve the solution.

Keywords

Semantics management; Laboratory Information System (LIS); Health Information Exchange (HIE); Common Terminology Service Release 2 (CTS2); Laboratory reports

Correspondence to:

Dr. Roberta Gazzarata

Healthropy s.r.l, Corso Italia 15/6, 17100 Savona, Italy.

E-mail: roberta.gazzarata@healthropy.com

EJBI 2017; 13(1):43-50

received: June 7, 2017

accepted: July 10, 2017

published: October 10, 2017

1 Introduction

The clinical data sharing represents a fundamental process to improve the medical research, patient care and reduce health costs [1, 2]. The Health Ministries of many developed countries are planning the creation of national health information exchange (HIE) systems by defining the functionalities to support the sharing of its content [3, 4].

In order to realize distributed system architectures able to satisfy this requirement, the management of semantics is a critical and essential aspect that must be considered. In this context, standardized terminologies, universally recognized for the specific application domain, must be adopted to semantically enrich data that are shared [4].

The use of these well-defined tools is not enough to solve the problem. In clinical practice, each medical department

uses its internal terminology that is usually very different from standardized vocabulary [5]. Therefore, it is important to represent the concepts and the relations between these concepts, defined in different local and standardized terminologies, so that can be understood and processed my machine, passing from the description of data to description of knowledge.

In addition, the medical knowledge evolution causes the change of concepts and terms, used to define them. For example the standardized vocabulary LOINC® (Logical Observation Identifiers Names and Codes) [6] is updated about twice every year. For this reason, it is necessary to maintain terminologies and their relations updated and coherent over the time, that is manage historical evolution and versioning [7].

From 2011, some of the authors were involved in two similar but separated initiatives to create prototypal solutions to manage semantics of laboratory tests, in particular the mapping between concepts at regional level. In both cases, concepts of local terminologies were encoded in the corresponding LOINC® codes through a translation process. The first initiative was performed in Veneto region within the regional project Veneto Escape [8]. In this case the translation project was performed by laboratory technicians through an ad hoc web application for a limited number of lab tests. At the end on the mapping, the resulting harmonization was stored within a database and it could be downloaded through a table that contains a view of all laboratory tests and the selected LOINC® entities. These *table* were uploaded in LISs in order to use LOINC® codes in the CDA R2 laboratory reports.

The second initiative was carried in Liguria region and the harmonization process was performed on Excel files by the authors with the cooperation of laboratory technicians through several cycles of translation proposals, review and approval on small set of codes. In this case, the resulting harmonization was stored within a database and was available through specific stored procedures in order to read the maps between codes [5]. After this experience, some of the authors decided to improve this prototype in order to design and develop a solution able to support both more terminologies and more functionalities as query and maintenance operations, adopting standardized solutions.

Taking in account these two experience, since 2015 a research is now to set up an infrastructure able to manage semantics at regional level that will be experimented to support regional HIE in Veneto Region. The aim of this infrastructure is to aggregate data coming from a regional HIE basing on semantic meaning of all available data [9]. In this paper, the authors present the first steps of this research, the current state of implementationa, the issues faced and the next steps.

2 Methods

In order to develop the overall solution that this ambitious research aimed to, different macro phases were considered. In particular, the first period of 2 years was designed to focus on the management of semantics in laboratory reports. For this reasons, the existing considered components were the Laboratory Information Systems (LISs) of the local departments of the Veneto region (23 units) and the regional HIE.

As indicated by the Italian Health Ministry, laboratory reports must be structured adopting the HL7 Clinical Document Architecture Release 2 (CDA R2) standard following the HL7 Italy Implementation Guide for Laboratory Reports [10] and the IHE (Integrating the Healthcare Enterprise) Laboratory Technical Framework [11]. According to this guide, the LOINC® vocabulary must be adopted to translate laboratory observations. For this reason, the authors decided to use LOINC® as reference code system. Another code system considered was the regional catalog of prescriptions.

To manage the semantics of the data involved in the contextual workflow, the design and the implementation of a terminology service was considered. In particular, in order to satisfy all requirements indicated by the project, the Common Terminology Service Release 2 (CTS2) standard was adopted [7, 12]. CTS2 is one of the products of the HL7 and Object Management Group (OMG) initiative called Healthcare Service Specification Project (HSSP) [13]. The aim of HSSP was to define health industry standards based on Service Oriented Architecture (SOA) approach to provide and guarantee an effective interoperability between distributed and heterogeneous applications which belong to independent socio-health system organizations [4, 13, 14, 15, 16].

The CTS2 standard provides a consistent specification to develop service interfaces to manage, search and access terminology content independent of the terminology content and underlying technological stack. For this reason, a terminological service compliant to CTS2 contributes to interoperability by supporting an easy access to the foundational elements of shared semantics. CTS2 defined different elements called *terminology resources* and different set of operations, called *functional profiles*, which could be performed on them. These functionalities can be implemented adopting several approaches [7, 12].

As all HSSP standards, CTS2 is distributed through the HL7 Service Functional Model (SFM), which provides a service interface specification at a functional level, and the OMG Service Technical Model (STM), which specifies the

technical requirements of the service [4, 7, 16]. In details, each OMG STM is formed by a set of human readable specification documents (e.g. pdf) and some computer processable files (e.g. WSDL and XSD) that can be used to automatically generate the interfaces adopted by web services and client applications to interact. The HL7 SFM for CTS2 is available at [7] while the OMG STM is available at [12].

The idea of the authors was to design and develop a complete terminology service considering the overall CTS2 specification. The authors decided to start to focus on the CTS2 *terminology resources* that were needed to satisfy the requirements of the first phase of the research:

- CodeSystem: either a classification system or a code system or a ontology or a thesaurus, etc. that may also carry information about its publisher, release cycles, purpose, etc
- CodeSystemVersion: a version of a CodeSystem that may also carry information about the release date, release format, contact information, etc.
- EntityDescription: a description about either a class or a role or a individual along with links to the CodeSystemVersion in which this description originates.
- Map: a collection of rules for transforming information represented using one CodeSystem into information represented in second one that may also carry information about the creators, intended use, CodeSystem involved, etc.
- MapVersion: a version of a Map that carries the *from* and the *to* CodeSystemVersion as well as a representation of the rules and process for the conversion.
- MapEntry: a definition of a set of rules that identifies how a single Entity that belongs to the *from* CodeSystemVersion maps to zero or more target Entities that belong to the *to* CodeSystemVersion.

For all these elements, the following *functional profiles* were considered:

- Maintenance: set of capabilities to *create*, *update*, *delete* a *terminology resource* through sets of changes, called *ChangeSets*, with specific properties to indicate the creation date, the user that changes the terminology content, the change information, the effective date, etc.
- Read: set of capability to *read* a *terminology resource* in a specific context (e.g. language, time, changeset).
- Query: set of capability to *search* the *terminology resources* with specific features in a certain context (e.g. language, time, changeset).
- History: set of capabilities to get the list of all the changes applied to a particular *terminology resource* and the list of all the *ChangeSet* applied to the whole terminology content.

The SOAP (Simple Object Access Protocol) was chosen as *implementation profile* and Microsoft Windows Azure was adopted as cloud platform to host the terminology service. The reason of this choice was that it provided high availability, scalability and manageability, fault tolerance, geo-replication, limitless storage, and security on the cloud. It allows the adoption of load balancing to implement failover, i.e., the continuation of a service after the failure of one or more of its components. Load balancing also enables other important features such as scalability [16 ,17, 18].

3 Results

In order to set up an infrastructure able to manage semantics of the content of laboratory reports, the authors designed the architecture represented in Figure 1. The solution is formed by the following independent and distributed systems:

- the regional HIE,
- the 23 LISs of the local departments of the Veneto region,
- the terminology service, called Health Terminology Service (HTS),
- the client web application to manage the content of the HTS (presented in the right side of Figure 1).

3.1 HTS

The core of the architecture is the HTS that is assembled by a *terminology database*, which stores all the information related to the terminology resources, and a set of web services (indicated in Figure 1 as *CTS2 Interface*), which provides access to the terminology database through interfaces compliant to the CTS2 standard. The *terminology database*, which was designed and implemented starting from the CTS2 OMG STM information model, is a very complex relational database, hosted on Microsoft SQL Azure. The *CTS2 Interface*, which was developed starting from the computer processable files provided within the CTS2 OMG STM, is represented by a set of Windows Communication Foundation (WCF) services, hosted on Microsoft Azure. In details, as indicated by the CTS2 standard, each terminology resource has a specific service for each supported functional profile. As mentioned in Methods section, the terminology resources considered were six and the authors implemented the maintenance, the read, the query and the history functional profiles, therefore the HTS is formed by 24 WCF services that are:

- CodeSystemCatalogReadService
- CodeSystemCatalogQueryService

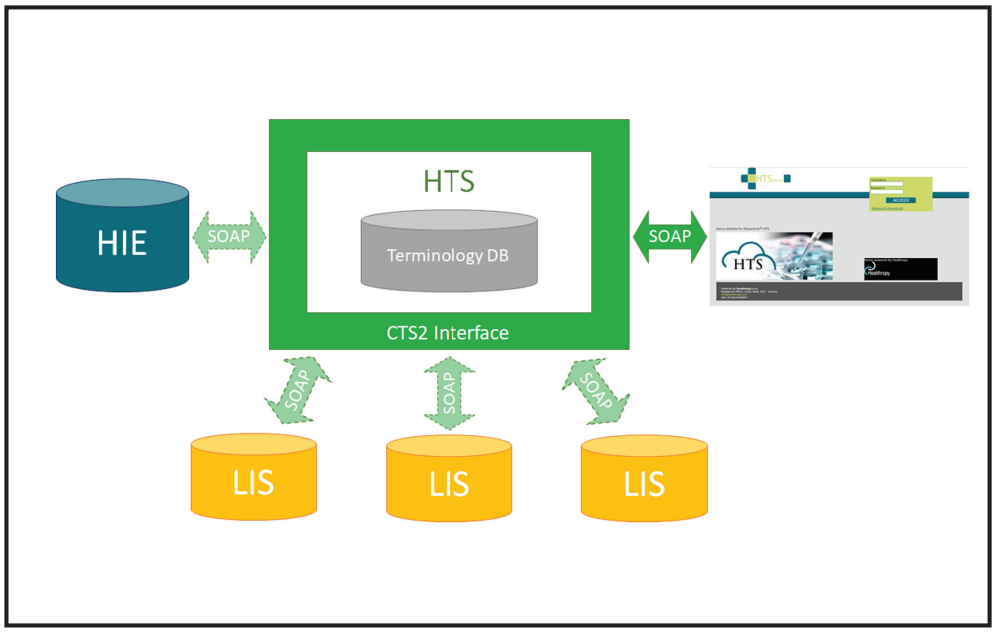


Figure 1: Proposed architecture to support semantics management at regional level.

- CodeSystemCatalogMaintenaceService
- CodeSystemCatalogHistoryService
- CodeSystemVersionCatalogReadService
- CodeSystemVersionCatalogQueryService
- CodeSystemVersionCatalogMaintenaceService
- CodeSystemVersionCatalogHistoryService
- EntityDescriptionReadService
- EntityDescriptionQueryService
- EntityDescriptionMaintenaceService
- EntityDescriptionHistoryService
- MapCatalogReadService
- MapCatalogQueryService
- MapCatalogMaintenaceService
- MapCatalogHistoryService
- MapVersionReadService
- MapVersionQueryService
- MapVersionMaintenaceService
- MapVersionHistoryService
- MapEntryReadService
- MapEntryQueryService
- MapEntryMaintenaceService
- MapEntryHistoryService

3.2 Client Web Application

In order to allow human users to manage the content of the HTS terminology database, a web application, which is client of the HTS WCF services, was involved. This application was designed and developed by the authors in deep collaboration with the medical staff that was responsible for the harmonization process between local terminologies and LOINC® vocabulary. It is continuously evolving to satisfy both the needs of laboratory technicians and the requirements that the Veneto region is designing to create the regional HIE and to manage the semantics of its content. For this first phase, the authors defined three user categories:

- *Terminology Administrators*: users that can manage all terminology resources;
- *Regional Terminology Administrators*: users that can manage only some terminology resources at regional level, so for all local departments of the Veneto region;
- *Local Terminology Administrators*: users that can manage only some terminology resources at local level, so for one or more specific local departments.

At the present, the web application is formed by three sections: user management, LIS management and mapping management.

User Management

In this section of the web application, all users can modify their credentials and *Terminology Administrators* and *Regional Terminology Administrators* can create new users. They can decide if the new user would be a *Regional Terminology Administrators* or a *Local Terminology Administrators* and in this case, they can select one or more local departments that the user would administrate.

LIS Management

The second section is used by the users to insert within the HTS the list of all laboratory observations, performed by each local department, indicating all their properties as code, description, units, specimen and the corresponding code defined within the regional catalog of prescriptions. The input can be manual, by completing a form, or can be automatic, by uploading a spreadsheet with preset structure. *Terminology Administrators* and *Regional Terminology Administrators* can insert laboratory observations for all local departments, while *Local Terminology Administrators* can do it only for the departments for which they are administrators. If the user is administrator of more than one department, the application asks to select the department that must be considered.

Mapping Management

This last section is the core of the application, which allows creating, deleting, updating and managing the versioning of the maps between the laboratory observations and the LOINC® entities. It is formed by a set of web pages that allow users to perform different sets of operations to guide the process of terminology management:

1. Map priority setting
2. Map creation
3. Elimination or update of the created maps
4. View of the created maps and of the state of map progress
5. Management of new LOINC® versions release and automatic versioning of maps
6. Manual versioning of maps where needed

The first form allows setting the map priority. As previously mentioned, with the LIS management section all laboratory observations are stored in the HTS with the corresponding code defined within the regional catalog of prescriptions. *Terminology Administrators* and *Regional Terminology Administrators* can select the codes of this catalog whose corresponding laboratory observations of local departments must be mapped with major priority.

The second set of web pages allows *Local Terminology Administrators* to create the association between local laboratory observations and LOINC® entities for the departments for which they are administrators. In order to help *Local Terminology Administrators*, a laboratory technician that has a thorough knowledge of LOINC, the *Alpha User*, was elected. His aim was to choose, for each code of the regional catalog of prescriptions, a set of possible LOINC® entities that could correspond to the prescription.

Also for this set of web forms, if the user is administrator of more than one department, she/he had to select the department that wants to consider. The system presents for the specific department all the laboratory observations that are unmapped, sorted by map priority in descending order. In addition, it highlights the observation with major priority. By this way, the user knows what observations must be managed first. After that the user has selected the observation that she/he wants to map, she/he can decide to adopt one of the following three approaches:

1. To view the set of LOINC® entities, selected by the *Alpha User*, related to the same code of the regional catalog of prescriptions of the chosen observation;
2. To view the LOINC® entities adopted in the other departments to map observations related to the same code of the regional catalog of prescriptions of the chosen observation;
3. To directly search in LOINC® website.

In all cases, after that the user has selected a LOINC® entity, the system presents all laboratory observations that was mapped with this LOINC® entity by the other departments. If the user confirms the operation, the map is created in HTS.

With the third form, *Local Terminology Administrators* can view all the maps that they created and can update or delete the associations between the laboratory observations and LOINC® entities. In case they delete a map, the corresponding laboratory observation becomes unmapped and, therefore, visible through the second sets of web pages.

Terminology Administrators and *Regional Terminology Administrators* through the fourth form can have a view of state of map progress and can see all the maps. In details, they can see in a schematic table for each department the number of all the laboratory observations and the number of the mapped ones. In addition, they can see all the maps with the corresponding department and download then via excel files. Table 1 shows examples of maps created by *Local Terminology Administrators* for different laboratories, between laboratory observations and LOINC® entity corresponding to the same entity of the regional catalog of prescriptions (code: 90.43.2_0, description: triglycerides).

When LOINC® releases a new version, *Terminology Administrators* and *Regional Terminology Administrators* can use the fifth set of web forms to manage the versioning of maps of all departments. As first step, the system presents a schematic comparison between the actual and the previous LOINC® versions. It indicates:

- the number of all LOINC® entities not deprecated in the actual and in the previous version
- the number of new LOINC® entities present in the actual version respect to the previous one
- the number of eliminated LOINC® entities in the actual version respect to the previous one

Table 1: Maps between internal laboratory observation (Internal Presctiprion Code, Internal Prescription Description, Internal Observation Code, Internal Observation Description, Units, Specimen) and Loinc code of different laboratory (Lab) corresponding to the prescription „Triglicerides" of the regional catalog of prescriptions.

Lab	Internal Prescription Code	Internal Prescription Description	Internal Observation Code	Internal Observation Description	Units	Specimen	LOINC® Code
1	a253	Triglycerides	a253	Triglycerides	mg/dL	PLASMA	2571-8
2	TGL	TRIGLYCERIDES	TGL	S-TRIGLYCERIDES	mg/dL	Serum	2571-8
3	132	Triglycerides	132	Triglycerides		P-PLASMA	2571-8
4	5	P-TRIGLYCERIDES	5	P-TRIGLYCERIDES	mg/dL	Ser/Plas	2571-8
5	TGL	P-Triglycerides	TGL	P-Triglycerides	mmol/L		14927-8
6	trig	s-TRIGLYCERIDES	trig	s-TRIGLYCERIDES	mg/dL	serum	2571-8
7	1035	S-TRIGLYCERIDES	1	S-TRIGLYCERIDES	mmol/L	Plasma	14927-8
7	1035	S-TRIGLYCERIDES	2	S-TRIGLYCERIDES second unit	mg/dL	Plasma	2571-8
8	LX148	S-TRIGLYCERIDES	148	P-TRIGLYCERIDES	mg/dL		2571-8
9	214C	TRIGLYCERIDES	214	P-TRIGLYCERIDES	mg/dL	PLASMA	2571-8
10	025	TRIGLYCERIDES	025	TRIGLYCERIDES	mg/dL		2571-8
11	210	S-TRIGLYCERIDES	210	S-TRIGLYCERIDES	mg/dL		2571-8
12	K02	TRIGLYCERIDES	K02	S-TRIGLYCERIDES	mM/L	SERUM	14927-8
12	K02	TRIGLYCERIDES	K02	S-TRIGLYCERIDES	mg/dL	SERUM	2571-8
13	TRIG	Triglycerides	TRIG	Triglycerides	mg/dL	S - Blood	2571-8
14	1035	S_TRIGLYCERIDES	1035-1	S_TRIGLYCERIDES-1	mmol/L	Serum	14927-8
14	1035	S_TRIGLYCERIDES	1035-2	S_TRIGLYCERIDES-2	mg/dL	Serum	2571-8
15	253	P-Triglycerides	253	P-Triglycerides	mg/dL	PLASMA	2571-8
16	TRIG	P-TRIGLYCERIDES	TRIG	P-TRIGLYCERIDES	mg/dL		2571-8
16	TRIG	P-TRIGLYCERIDES	TRIG	P-TRIGLYCERIDES (2° UM)	mmol/L		14927-8
17	210	TRIGLYCERIDES	210	TRIGLYCERIDES	mg/dL	PLASMA PL	2571-8
17	210	TRIGLYCERIDES	210	TRIGLYCERIDES	mmol/L	PLASMA PL	14927-8
18	CH59	TRIGLYCERIDES	CH5701	Triglycerides	mg/dL	Heparin Plasma	2571-8
19	0055	TRIGLYCERIDES	0055	TRIGLYCERIDES	mg/dL	serum	2571-8
20	0379	P-TRIGLYCERIDES	T35	Triglycerides	mmol/L		14927-8
20	0379	P-TRIGLYCERIDES	T35	Triglycerides (2° UM)	mg/dL		2571-8
20	0379	P-TRIGLYCERIDES	T46	Cholesterol LDL	mmol/L		39469-2
20	0379	P-TRIGLYCERIDES	T46	Cholesterol LDL (2° UM)	mg/dL		13457-7
21	T35	P-Triglycerides	T35	Triglycerides	mmol/L		14927-8
21	T35	P-Triglycerides	T35	Triglycerides (2° UM)	mg/dL		2571-8
22	TG	Triglycerides	TG	P-TRIGLYCERIDES	mg/dL	PLASMA	2571-8

- the number of maps between entities in the previous version and entities in the last one (provided by LOINC®).

Then the application shows a table that presents, for each LOINC® property, the number of LOINC® entities that have changed the value of the specific property. This allows *Terminology Administrators* and *Regional Terminology Administrators* to have a vision of the overall changes in LOINC vocabulary. The user can select the LOINC® property that must be considered by the system to manage the automatic versioning of maps. After this phase, the system allows *Terminology Administrators* to run the following algorithm:

a) For each local department, the system considers the maps from laboratory observations of the specific LIS to the LOINC® entities in the precedent version.

b) For each map, it gets the used laboratory observation and LOINC® entity (defined within the precedent version).

c) If the LOINC® entity exists in the actual version too, the application considers this entity (go to step e).

d) If the LOINC® entity does not exist in the actual version, it uses the map from the LOINC® previous version to LOINC® last version and gets the corresponding LOINC® entity in the actual version.

e) The system compares the values of the property previously selected by the user of the two LOINC® entities.

f) If all values are equal, the system creates a new map from the laboratory observation (get in step b) and the LOINC® entity.

g) If there is at least one value different, the system saves the map as to be manually validated by *Local Terminology Administrators.* Then it passes to the next map (go to step b).

h) When all maps of a department have been processed, the application sends an e-mail to *Local Terminology Administrators* of the specific department, to *Terminology Administrators* and to *Regional Terminology Administrators.* Then it passes to the next local department (go to step b).

The last set of web forms allows *Local Terminology Administrators* to manually manage the maps that must be validated. The system presents a list of all these maps and, after that the user has selected a specific map, it shows the two LOINC® entities, one in the previous version and one in the last version, with all their properties and highlights the values that are different. Then the user can decide if the map between the laboratory observation and LOINC® entity can be confirmed or not. In the first case, the system creates a new map, while in the second one, the laboratory observation becomes unmapped and, therefore, visible through the second sets of web pages.

3.3 Numeric Results

At the present, the HTS manages 25 CodeSystems, 27 CodeSystemVersions, about 260000 EntityDescriptions, 47 MapCatalogs, 71 MapVersions and about 30000 MapEntries. After the last release of LOINC, the automatic versioning of maps allowed automatically creating more than 11500 MapEntries that means that the system automatically validated about 80% of the existing maps in the previous LOINC® version. About 800 LOINC entities adopted in the maps changed from the previous to the last version and the *Local Terminology Administrators* had to manually validate about 3000 maps.

Discussion and Conclusions

The adoption of CTS2 specification allows the HTS to support different use cases, which are essential to manage semantics to support the clinical data sharing of the HIE content. A example of use case could be the one reported in the Figure 2: the automatic update of the HIE content with a new laboratory report. When a physician prepares a laboratory report, the client application of LIS, for each local code, used in the report, communicates with the HTS by calling several CTS2 functionalities (represented in a single interaction, Translation (local code), in Figure 2 for simplicity) to get the corresponding LOINC® code. Then, the LIS client application can use these LOINC® codes to create the CDA R2 document that can be digitally signed and sent to the regional HIE following the national specifications. Another example can be the comparison of laboratory observations contained within reports produced in different years so adopting different LOINC® versions. Thanks to history functional profile, it is possible to get all changes performed on the HTS content and retrieve the corresponding LOINC® entities in the different versions. A similar example can be the management of laboratory reports stored in the HIE adopting outdated LOINC entities. The authors are planning an algorithm that will allow interacting with the HTS and retrieving the corresponding LOINC® entity in the actual version. This could be possible using the versioning of two maps stored in HTS; the first one is between laboratory observations and LOINC® entities and the other one is between entities in different LOINC® versions, provided by LOINC®. If the specific laboratory observation was mapped with an entity in the current LOINC® version, the system will return this LOINC® entity as the corresponding entity, while if the laboratory observation was not mapped in any entity in the current LOINC® version, the system will process the maps provided by LOINC® in the different versions to return the corresponding LOINC® entity in the current version.

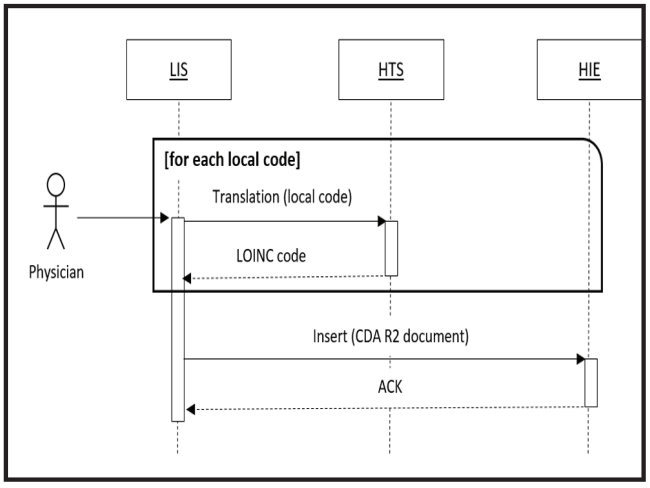


Figure 2: Example of use case supported by the architecture to manage semantics.

Another important advantage of CTS2 is that it proposes an information model where all the terminology resources are well defined and separated allowing the complete control of the terminology content. In fact, the CTS2 information model allows defining different access profiles to the terminology content. For example, it is possible to establish for each user a set of pairs of terminology resource

and functional profile for which she/he is allowed to have access. The advanced management of these access profiles is research argument of some of the authors.

During the development of HTS, the authors had to face some problems related to errors in the computer processable files provided by OMG. Three classes of errors were found: not compliance with the information model described in the human readable specification documents, syntax errors and structural errors. The authors had to correct all the errors in the processable files before generate all the needed services interfaces. The authors are preparing a white paper with all the errors that they found in order to share this knowledge with HL7 and OMG members.

The correction of errors in processable files caused a delay in the implementation of the HTS and in the first release of the client web application. For this reason, at the present this application is connected only to some services of HTS. In fact, the authors had to decide what functionalities of the web application must develop with priority basing on the deadlines that the Veneto region designed to create the regional HIE and to manage the semantics of its content. Therefore, the authors had to upload some terminology resources as all entities of the regional catalog of prescriptions and LOINC® directly in the database through stored procedures. The authors are working on the second release of the client web application that will allow managing all the terminology content only calling the CTS2 HTS interfaces.

In the next phases of the research project, the HTS will be also directly connected with the LIS of 4 local departments in order to realize the use case represented in Figure 2. At the present, the maps between the local laboratory observations and LOINC® entities are first downloaded through the fourth form via excel files and then directly uploaded in the application of the LISs that creates the CDA R2 containing the laboratory report.

Finally, in the next months the authors want to improve the comparison proposed in the step *e* of the described automatic versioning algorithm in order to reduce the number of maps that be manually managed every time that LOINC® releases a new version. In details, they are plaining to design a solution that will allow understanding if the changes of the LOINC® entities are syntactic or semantic. By this way, in case of syntactic changes, the system could automatically validate the maps for the new version and could require the manual validation only for semantic changes. In addition, they want to design and develop an algorithm to check if the LOINC® codes selected in the mappings are correct. In fact, at the present all maps are manually processed. In the future, all maintenance functionalities will be adopted to manage other regional code systems as the regional catalog of prescription.

References

[1] Lo B. Sharing clinical trial data: maximizing benefits, minimizing risk. Jama 2015; 313: 793–794.

[2] Giannini B, et al. IANUA: a regional project for the determination of costs in HIV-infected patients. Stud Health Technol Inform. 2014; 210: 241–245.

[3] Italian Health Ministry. DPCM. 2015; n. 178. 2015.

[4] Gazzarata R, Giacomini M. A Standardized SOA for Clinical Data Sharing to Support Acute Care, Telemedicine and Clinical Trials. EJBI. 2016; 12: en49-en57.

[5] Canepa S, Roggerone S, Pupella V, Gazzarata R, Giacomini M. A Semantically Enriched Architecture for an Italian Laboratory Terminology System. In: Proceedings of the XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013, 2014: 1314–1317.

[6] Regenstrief Institute. Logical Observation Identifiers Names and Codes (LOINC). [Online]. Available: http://loinc.org/.

[7] Health Level Seven International, Inc. HL7 Common Terminology Services, Release 2: Service Functional Model Specification, Normative Release. Ann Arbor: HL7; February 2015.

[8] Veneto ESCAPE. [Online]. Available: http://www.consorzioarsenal.it/web/guest/progetti/veneto-escape/dicono-di-veneto-escape.

[9] Veneto region, Bur n. 86, 08 September 2015.

[10] HL7 Italia. Implementation Guide Clinical Document Architecture (CDA), Rel. , Rapporto di Medicina di Laboratorio, 2009.

[11] Integrating the Healthcare Enterprise. IHE Laboratory Technical Framework, vol. 3 (LAB TF-3) Content, rev. 2.1 – Final Text, 2008.

[12] Object Management Group. Common Terminology Services 2, Version 1.2, 2015. [Online]. Available: http://www.omg.org/spec/CTS2/1.2/.

[13] Kawamoto K, Honey A, Rubin K. The HL7-OMG Healthcare Services Specification Project: motivation, methodology, and deliverables for enabling a semantically interoperable service-oriented architecture for healthcare. J Am Med Informatics Assoc. 2009; 16: 874–881.

[14] Gazzarata R, Vergari F, Salmon Cinotti T, Giacomini M. A standardized SOA for clinical data interchange in a cardiac telemonitoring environment. IEEE J Biomed Heal Inf. 2014; 18: 1764–1774.

[15] Gazzarata G, Gazzarata R, Giacomini M. A Standardized SOA Based Solution to Guarantee the Secure Access to EHR. Procedia Comput Sci. 2015; 64: 1124–1129.

[16] Gazzarata R, Giannini B, Giacomini M. A SOA-based platform to support clinical data sharing. J. Healthc Eng 2017; 1–24.

[17] Guleria P. ACID Support and Fault-Tolerant Database Systems on Cloud: A Review. Int J Adv Res Comput Sci Electron Eng. 2012; 1: 18–20.

[18] Rimal BP, Choi E, Lumb I. A taxonomy and survey of cloud computing systems. In: Proceedings of NCM'09. Fifth International Joint Conference on INC, IMS and IDC, 2009: 44–51.

Integrated Care Solutions for the Citizen: Personal Health Record Functional Models to Support Interoperability

Dimitrios G Katehakis, Haridimos Kondylakis, Lefteris Koumakis, Angelina Kouroubali, Kostas Marias

Foundation for Research and Technology – Hellas, Institute of Computer Science, Greece

Abstract

The ever-growing demand for acquiring, managing and exploiting patient health related information has led to the development of many e-Health systems and applications. However, despite the number of systems already developed and the apparent need for such systems, end users can only collect online and exploit, only a limited set of information for health purposes in the context of personalized, preventive and participatory medicine. To this direction, this paper initially presents the personal health record (PHR) concept, related work and best practices for the development of PHR systems in a standardized manner. It also outlines the proposal for meaningful use criteria in the United States (US) and the health level seven (HL7) personal health record system functional model (PHR-S FM). Focus is put on trying to link core functionality modules of the Integrated Care Solutions ™ PHR system, designed to support the citizen, paying emphasis on wellbeing, home care and the management of chronic diseases with PHR-S FM personal health functions, in a preliminary effort towards the exploration of functional models to support interoperability. Based on accumulated experiences from many European Union (EU) research projects, the paper concludes by providing directions towards achieving wider PHR adoption and meaningful use.

Keywords

Personal Health Record; Integrated Care; Meaningful Use; PHR-S FM

Correspondence to:

Dimitrios G. Katehakis
Foundation for Research and Technology – Hellas, Institute of Computer Science, N. Plastira 100, Vassilika Vouton, GR-700 13 Heraklion, Greece.
E-mail: katehaki@ics.forth.gr

EJBI 2017; 13(1):51-58
received: June 15, 2017
accepted: July 20, 2017
published: October 10, 2017

1 Introduction

The Personal Heath Record is a tool designed for the citizen with the goal to promote continuity of care in a reliable, accessible and secure fashion. The main expected benefit is to empower both patients and clinicians towards a more synergistic, patient-centric healthcare, promoting shared care and personalized medicine throughout a citizen's lifespan. The idea of a PHR has been developed in parallel with the development of the Electronic Health Record (EHR). The PHR was first mentioned in an early report from the US Institute of Medicine called 'The Computer-Based Patient Record: An Essential Technology for Health Care' [1]. The report described the envisaged requirements of such an endeavour. In Europe the PHR concept has been introduced through European Directive 95/ 46/ EC [2], which first allowed/ proposed the direct interaction of the person with his/ her health record including input of data from home, work and leisure places [3]. It is a fact that the advancements in healthcare practice, the limitations of the traditional healthcare processes and the need for flexible access to health information, create an ever-growing demand for electronic health systems everywhere. To this direction, PHR systems provide citizens with the ability to become more active in their own care combining data, knowledge and software tools. The PHR concept is citizen centric, in the sense that its management is the primary responsibility of the citizen. Through a PHR application, the citizen/ patient is able to provide daily life-status information, maintain his/ her own record of medical exams and define the access rights to own personal data, leveraging that access to improve own health and manage own diseases.

Over the last twenty years a large amount of PHR-like systems have been developed such as: 911 Medical ID (http://www.911medicalid.com/), CareZone PHR (https://carezone.com/), Dossia (http://www.dossia.org/), eclinicalWorks Patient Portal (http://www.eclinicalworks.com/products-

patient-portal.htm), Epic MyChart (http://www.epic.com/software-phr.php), HealtheTracks (http://www.healthetracks.com/), Indivo-X (http://indivohealth.org/), KIS PHR (http://kismedicalrecords.com/), LifeLedger (http://www.elderissues.com/), MedHelp PHR (http://www.medhelp.org/), MedicAlert (http://www.medicalert.org/), MedicKey PHR (http://medickey.com/), Microsoft HealthVault (http://www.microsoft.com/en-gb/healthvault/default.aspx), Minerva Health Manager (http://www.myminerva.com/), MyALERT (http://www.alert-online.com/myalert), myMediConnect PHR (http://www.passportmd.com/), MyOscar (http://myoscar.org/), NoMoreClipboard (http://www.nomoreclipboard.com/), OpenMRS (http://openmrs.org/), Patient Ally (https://www.patientally.com/Main), Patient Fusion (http://www.practicefusion.com/pages/phr.html), PatientsLikeMe (http://www.patientslikeme.com/), Tolven (http://www.tolven.org/), Web MD Health Manager (http://www.webmd.com/health-manager), zweena PHR (http://www.zweenahealth.com/), and others.

However, despite the wide variety of potential benefits [4] the uptake of PHRs has been proven to be really slow [5]. Recent reviews [6, 7] identify as a problem the fact that only a small subset of the PHR applications are free, web-based and open-source. Nevertheless, many different business models exist, fee-based or commercial, complicating even more the selection of an appropriate PHR. In addition, the main problems, still pending to be resolved, are the following:

Interoperability: PHR systems are rarely integrated and interoperable with other electronic service providers [8, 9]. In most of the cases, end-users need to enter the whole amount of input information by themselves.

Usability/ Adaptability: The majority of PHR systems follow the approach "one system fits all". However, different persons with different primary diseases have different needs and the PHRs so far fail to be adapted to specific needs [10].

Trust: The limitations of the methods for sharing information among patients, and their relatives, doctors and researchers. There is a sense of lack of trust as well as inefficient access control and security mechanisms [6, 11, 12].

Added Value: PHR systems are not linked with specific services. The added-value for citizens to maintain a personal health file through manual input of data has not been adequately demonstrated [6, 7, 12].

To face these challenges guidelines and standards are starting to emerge to support quality PHR systems like the US Meaningful Use Criteria [13], and the HL7 PHR-S FM [14, 15]. However, the adoption of those is still limited.

This paper focuses on the presentation of a beyond-the-state-of-the-art PHR system, which is interoperable, personalized, and adaptable for various diseases. It has been designed for easy integration with existing clinical information systems. Its development has been based on the outcomes of various EU research projects and exhibits a high technology readiness level. The PHR provides effective and efficient access control mechanisms and many added-value services. The goal of the presented system is to provide an innovative ecosystem for enhancing the self-management capacity of patients through the involvement of all stakeholders participating in the therapeutic process.

The structure of the rest of this paper is as follows: Section 2 describes the methods used. Then preliminary results are reported in Section 3. Finally, Section 4 concludes this paper and presents directions for future work.

2 Methods

In order to conduct this work, literature review was performed on established work, and additional information was retrieved from published material and web sites in order to identify current trends and good practices. In the following subsections, a short presentation is made of the US Meaningful Use Criteria, the HL7 PHR-S FM R1, and the related European research and development projects that have guided the design and development of the PHR system presented. The different models and projects have been selected to emphasize the functionalities that are required to support efficient PHR adoption.

2.1 Meaningful Use

The US Meaningful Use (MU) Criteria [13] were initially introduced as an EHR incentive. Later on, they were adopted also as a guideline for PHR systems, since they include specific requirements for patient engagement, as seen in Stage 2 of the MU program [16]. Since 2014, in order to meet the requirements of the Centers for Medicare & Medicaid Services (CMS) EHR Incentive Program, healthcare providers must provide a patient portal. The Healthcare Information and Management Systems Society (HIMSS) identified the "required" and the "helpful" core and menu objectives that the patient portals should support or consider to support in order to allow patients to interact with their healthcare provider [17], from the complete list of the MU Stage 2 program (17 core measures and 6 menu measures). The minimum core objectives for a PHR, connected to an EHR, include a mechanism to provide patients the ability to view online, download and transmit their health information, and a secure electronic messaging system to communicate.

The MU program is using certified EHR technology to reduce health disparities and improve quality, safety, and efficiency of healthcare [18]. Having guaranteed data capture and sharing, MU stage 2 aimed towards improved outcomes through the advancement of clinical processes. The patient

is expected to be able to access their health information on demand through a patient portal or PHR. As reported in [19], the PHR products available as of 2014 are likely to meet and exceed meaningful use stage 3 targets before 2020 without any incentive.

2.2 HL7 Personal Health Record System Functional Model, Release 1 PHR-S FM

The HL7 PHR-S FM, which is a Draft International Standard: ANSI [20]/HL7 PHRSFM, R1-2014; ISO/FDIS 16527:2013(E) [21], defines a standardized model of the functions that may be present in a PHR System (PHR-S). According to HL7, the PHR is the underlying record (e.g., data, information, pictures, sounds, graphs, videos, etc.) that the software functionality a PHR-S maintains. Depending on the functionalities applicable for a particular business case, certain PHR functions and criteria in the PHR-S FM will apply to any single PHR-S implementation. This, more concrete, expression of usable subsets of functions from the PHR-S FM is called *functional profile*. This profile is the standard description and common understanding of the requested or available functions in a given environment. The intention is for all functions describing the behaviour of a system, in a consumer-oriented language, to be recognizable to all key stakeholders of a PHR-S. PHR-S FM consists of three main sections: Personal Health, Supportive, and Information Infrastructure, as outlined in Table 1.

Table 1: The PHR-S functional outline.

PHR-S Function List Sections	ID#	Superset of functions
Personal Health	PH.1.0	Account Holder Profile
	PH.2.0	Manage Historical Clinical Data And Current State Data
	PH.3.0	Wellness, Preventive Medicine, and Self Care
	PH.4.0	Manage Health Education
	PH.5.0	Account Holder Decision Support
	PH.6.0	Manage Encounters with Providers
Supportive	S.1.0	Provider Management
	S.2.0	Financial Management
	S.3.0	Administrative Management
	S.4.0	Other Resource Management
Information Infrastructure	IN.1.0	Health Record Information Management
	IN.2.0	Standards Based Interoperability
	IN.3.0	Security
	IN.4.0	Auditable Records

The Personal Health (PH.1.0-PH.6.0) section functions are the subset of PHR-S functions that manage information and features related to self-care and provider-based care over time. The Supportive (S.1.0-S.4.0) section functions are the subset of PHR-S functions that assist the PHR account holder with administrative and financial requirements. The Information Infrastructure (IN.1.0-IN.4.0) section consists of PHR-S functions that ensure that the PHR-S provides information privacy and security, interoperates with other information systems (including PHR and EHR systems), and helps make PHR-S features accessible and easy to use.

According to HL7 PHR-S FM, in order to ensure the necessary functions in selecting or developing PHR systems, it is important to create accurate, clear and impartial functional profiles by selecting functions from the HL7 EHR-S FM. A functional profile is a selected set of functions that are applicable for a particular purpose, group of users, degree of interoperability, etc. The profile consists of the choice of certain functions that can be mandatory (SHALL), prohibited (SHALL NOT), optional recommended (SHOULD), optional, or permissible (MAY). The goal of creating a functional profile is to support a business case for PHR-S use by selecting an applicable subset of functions from the PHR-S FM. A formal process exists for registering and balloting functional profiles. Testing and certification procedures are then required to ensure that the subsystems and the general PHR-S conform to the selected functional profile and meet the characteristics for the proper operation of the system.

A PHR-S does not conform directly to the PHR-S FM; rather, a PHR-S conforms to a functional profile (i.e., a subset – more specifically, a tailored subset) of the PHR-S FM. Conformance to the PHR-S FM is defined for functional profiles.

2.3 European Projects

The implementation of projects co-funded by the EU, during the past few years, has provided significant experience gains on PHR development, for specific cases. Projects in which the authors of this paper have been actively involved are briefly described below.

REACTION: The REACTION project (http://www.reaction-project.eu/news.php, 2010-2014), aimed to research and develop an intelligent service platform for professional, remote monitoring and therapy management of diabetes patients in different health systems across Europe. As such, the platform is not a general-purpose PHR system but optimized especially for the empowerment of diabetic patients. The constructed platform can execute various clinical applications for monitoring of vital signs, context

awareness, feedback to the point of care, integrative risk assessment, event and alarm handling as well as integration with clinical and organisational workflows and external Health Information System [22].

P-Medicine: The p-medicine EU project (http://www.p-medicine.eu, 2011-2015) created an infrastructure that facilitates the translation from current medical practice to personalized medicine. Essential to the realization of personalized medicine is the development of information systems capable of providing accurate and timely information about potentially complex relationships between individual patients, drugs, and tailored therapeutic options. In the context of the project, a range of services were designed and developed on top of a PHR system. The p-medicine PHR is based on a general purpose PHR (Indivo-X) with extensions towards the directions of cancer patient profiling and clinical decision support for personalized oncology [23]. Furthermore, the project implemented a secure mechanism for informed secondary use of patient's biomaterial and personal data via the PHR [24]. The p-medicine tools and technologies have been validated within the concrete setting of advanced clinical research with pilot cancer trials based on clear research objectives, in the domains of Wilms tumour, breast cancer and leukaemia.

EURECA: The goal of the EURECA project (http://eurecaproject.eu/, 2012-2015) was to enable seamless, secure, scalable and consistent linkage of healthcare information residing in electronic health record (both EHR and PHR) systems with information in clinical research information systems, such as clinical trials. Achieving semantic interoperability among PHR and clinical trial systems was at the core of the EURECA project, as it was the basis for enabling many of the software services and tools developed in the project. Data management services were implemented for a variety of EHR and clinical trial systems (e.g. Obtima, OpenClinica) in order to achieve semantic interoperability with the Indivo-X PHR using terminology standards and HL7 mechanisms for exporting and importing data [6, 11].

MyHealthAvatar: The MyHealthAvatar (http://www.myhealthavatar.eu/, 2013-2016) EU project [25] (FP7) was an attempt for the digital representation of patient health status. The goal was to create a "digital avatar", i.e. a graphical representation/ manifestation of the user, acting as a mediator between the end-users and health related data collections, focusing on the interoperability and the data integration aspect. It was designed as a lifetime companion for individual citizens to facilitate the collection, the access and the sustainability of health status information over the long-term. Among others, key questions that are answered in this context is how to develop optimal frameworks for large-scale data-sharing, how to exploit and curate data from various Electronic and Patient Health Records, assembling them into ontological descriptions relevant to the practice of systems medicine and how to manage the problems of large scale medical data.

iManageCancer: The iManageCancer (H2020) EU project (http://imanagecancer.eu/, 2015-2018), has the objective to provide a cancer specific self-management platform designed according to the needs of patient groups. At the same time, it focuses on the wellbeing of the cancer patient with special emphasis on avoiding, early detecting and managing adverse events of cancer therapy but also, importantly, on the psycho-emotional evaluation and self-motivated goals [26]. In this context, developed cancer specific apps allow patients, through an easy-to-use interface for mobile devices, to keep track of their health and disease status and to keep a health diary on personal clinical observations such as side effects of therapies, which the patient can share with his healthcare providers. Health and disease status includes therapies and results of clinical interventions or tests.

3 Results

Integrated Care Solutions ™ (ICS) is a software suite developed by FORTH and includes several tools and applications for electronic health management. The majority of the already implemented software components of the platform are in operational use in several units of the national health system in Greece [27]. The ICS-C applications, as part of the ICS suite, aims towards directly supporting the Citizen. The Personal Health Record for the Citizen (PHR-C) end-user application belongs to this group of applications. Figure 1 depicts the architectural approach and key modules involved.

PHR data types have evolved since the first PHRs [28]. The components that are common to the EHR and are stored also in the PHR include medications, scheduled appointments, vital signs, medical history (problems list), laboratory information, immunizations, scanned documents, and progress notes about changes in the patient's health. Ideally, the PHR should include as much relevant data as possible over the individual's lifetime, from multiple sources, including health care facilities as well as the individual [29]. The specific data source of each item should be labelled and visible to the user. The more comprehensive the data contained in a PHR are, the more useful the data will be to patients and care providers, although there are no conventions for what data should be contained in a PHR.

PHR-C functionalities and modules can be directly linked to the PHR-S FM. In an effort to depict the relation of core functionality modules with EHR-S FM, towards establishing an initial approach for defining an initial profile to test compatibility against, we have come up with the links depicted in Table 2.

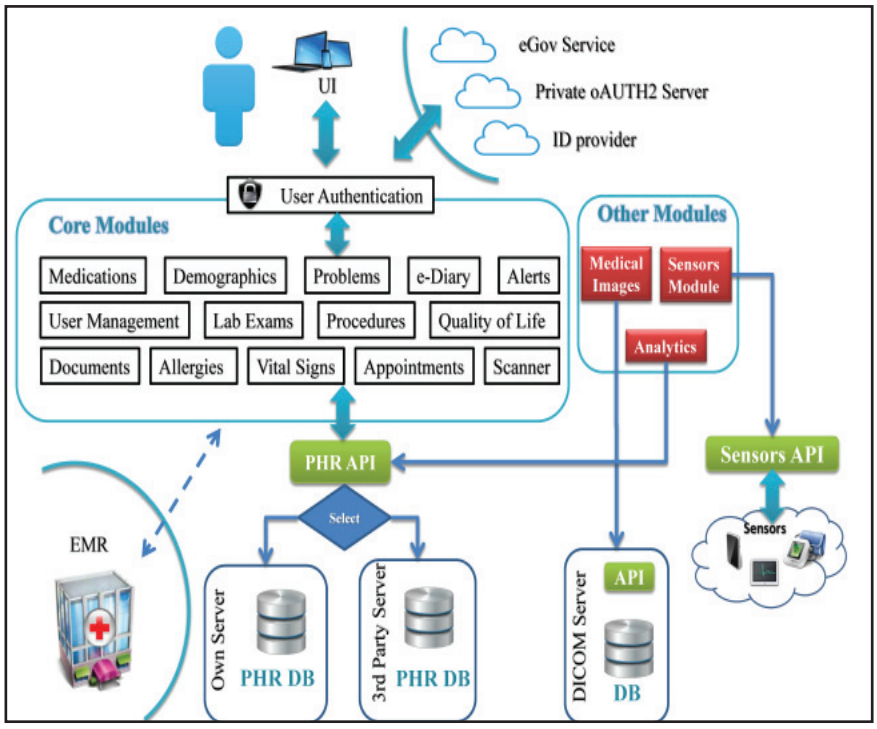


Figure 1: Integrated Care Solutions PHR high-level architecture approach.

Table 2: ICS PHR-C core functionality modules link to PHR-S FM R1 personal health functions.

Name	Potential Sources	Description	PHR-S FM ID#	PHR-S FM Statement
Alerts	PHR, external sources	This app allows the implementation of care management alerting rules to appear in end-user(s) account.	PH.3.5.5	Notify the PHR Account Holder of an event or situation that may need immediate action.
Allergies	PHR, EHR	For recording allergies and related information (allergy name, severity, allergen, adverse reactions, etc.).	PH.2.5.4	Manage the PHR Account Holder's list of known allergens and adverse reactions with all pertinent information.
Appointments	PHR, external providers	To allow an end-user to schedule his appointments. Appropriate reminders are then issued to remind him a specific appointment.	PH.3.5.1	Provide a health calendar to record and display health care events.
Demographics	PHR, eGov Service	For recording demographic information (address, gender, date of birth, etc.)	PH.1.2	Enable the PHR Account Holder to manage information about demographics.
Documents	PHR	For storing personal documents as attachments (discharge letters, prescriptions, PDFs, ECGs, DICOM images etc)	PH.3.1.1	Provide the ability for the PHR Account Holder to enter personally sourced data and to make it available electronically to authorized health care provider(s) or other authorized users or applications.
e-Diary	PHR, external providers	This app allows the diary visualization of all information available in a user profile.	PH.3.5.4	Present the PHR Account Holder with reminders either sent by external sources (such as from provider(s)), or internally generated from information in the PHR-S (such as guideline-based reminders, prescription refills, appointment reminders, or other calendar entries).
Lab Exams	Patient, EHR, commercial laboratories	For recording laboratory results and related information (lab test name, date, value, abnormal values, etc.)	PH.2.5.3	Manage results of diagnostic tests including inpatient, ambulatory and home monitoring tests.

Medications	PHR, EHR, claims history	For recording medication prescribed medicines taken and related information (drug name, quantity, unit, date etc.).	PH.2.5.2	Manage the PHR Account Holder's medication list.
Problems	PHR, EHR	For recording problems and related information (start date, category, details, etc.)	PH.2.5.1t	Manage the PHR Account Holder's health problem list and provide the ability to manage the problem list over time in accordance with organizational policy and/or jurisdictional law.
Procedures	PHR, EHR, or claims	For recording medical procedures and related information (name of the procedure, date performed, institution, cost etc.)	PH.2.5.7	The list of past procedures is a useful summary of what has been done in the past and anatomic changes have occurred that might influence current assessments and treatments. It is important to capture any surgical implants and associated lot/ serial numbers for tracking/ reporting purposes.
Quality of Life	PHR	For recording responses to validated questionnaires to assess the quality of life of individuals.	PH.3.1	Provide the ability for the PHR Account Holder to enter personally sourced data and to make it available electronically to authorized health care provider(s) or other authorized users or applications.
User Authentication	PHR, external services	To allow proper user authentication	PH.1.1	Unambiguously identify the PHR Account Holder; correctly link the information with the PHR Account Holder and vice-versa.
User Management	PHR, external services	To enables effective management for the administrator (role definitions etc).	PH.3.5.3	Each individual that accesses the PHR should be registered in a directory with his or her contact information and granted specific access rights.
Vaccinations/ Immunizations	PHR, EHR	Dates and types of (childhood) vaccinations	PH.2.5.5	Manage the Account Holder's immunization data and associated capabilities including reminders, alerts, compliance, and administration.
Vital Signs	PHR	For recording vital signs such as pulse, temperature etc. and related information (date, unit etc.).	PH.3.1.1	Provide the ability for the PHR Account Holder to enter personally sourced data and to make it available electronically to authorized health care provider(s) or other authorized users or applications.

In order to address the need for an interoperable and adaptable PHR, the ICS PHR system is expanded to include modules such as home care services, connection to wellness applications to automate input of data (e.g. steps/ day, water/ day, sleep patterns and others), links to the electronic health record of the citizen, the national e-prescription service, and other connections to third party apps for accessing data from clinical/ hospital information systems. In addition, PHR-C is incorporating various personalized medicine modules to address emerging new data, including genetic information, medical advice and recommendations, and prevention information. Specialized modules based on specific chronic conditions are under development to support patient empowerment and self-management. Home monitoring, enhanced communication with health providers, and information on guidelines for prevention and life style have been shown to reduce comorbidities and improve quality of life.

In addition, the experience accumulated and the tools developed during the EU projects (Section 2.3) have rendered PHR-C as an important tool not only for health management but also for clinical research. Advance work in progress includes modules that allow authorized end-users to make cohort analysis on all patient data, to visualize graphically the psycho-emotional profile of patients using various graphical paradigms, to perform advanced searches in specialized databases using natural language (advanced search engine tools), and to allow researchers to generate requests for specific cohorts. Finally, the PHR-C incorporates advanced decision support functionalities exploiting computer-based clinical guidelines.

The use of the PHR-S FM allows the direct linking of the core functionality modules of PHR-C with PHR-S FM Personal Health Functions. Setting conformance criteria will require the definition of certain functional profiles. In addition to the importance of designing a PHR in accordance to the functional models methodology discussed in Section 2.2, it is also vital to be able to provide disease-specific modules in order to deal with the plethora of information needed by the experts/ decision makers who play a vital role in the management of diverse diseases.

4 Conclusions

The real goals underlying the development and implementation of electronic health systems are to allow citizens to stay healthy, effectively manage chronic conditions, reduce comorbidities and improve quality of life. The PHR has the potential to become the life-long companion of citizens that can truly transform health care and establish continuity of care. PHRs offer the potential to improve patient-clinician interactions, empower citizens to become co-producers of their health, and reduce the cost of healthcare by maximizing on-line interactions and avoiding unnecessary hospitals visits. The vision is to achieve a better health service for all citizens and a better outcome for patients. In order to do that any PHR solutions must be prepared for future integration with any health information system. Therefore any information and communication technology support needs to consider current technology developments and demonstrate flexibility and capacity, by adopting related international standards and architectures in a coordinated manner.

As discussed earlier, PHR adoption has been slow due to various reasons such as lack of interoperability, low usability and adaptability and limited added-value [30]. Lack of wider acceptance and fragmentation in efforts remain obstacles for wider adoption and consequently all the expected benefits are yet to be experienced. However, there are success stories to present as a proof of concept. In Australia, almost 20% of the country's population has registered for My Health Record (https://myhealthrecord.gov.au/). This success story indicates that the vision of PHR is indeed feasible under strict regulatory strategies; however, they require strong governmental persistence. In Europe, the key issue to success is standardization, in terms of functionality offering and interoperability, requiring a coordinated governance framework and process.

Following a structured development of PHR using common established set of criteria and functionalities can help the wider adoption of PHR systems. The PHR-S FM offers a realistic and applicable promotion of functionality and interoperability components of PHR systems based on functional profiles. This in turn gives the opportunity to support PHR system certification programs underway or emerging in many countries. Focusing on solving actual end-user needs can lead to a wider adoption and meaningful use of PHR systems. Using the notion of functional profiles, a general purpose PHR can be instantiated for specific types of diseases and the individual needs of the citizens. Adaptability and personalization are key to the successful deployment of any large-scale PHR infrastructure, bearing in mind that interoperability is not possible without standards and specifications.

5 Acknowledgements

The development of ICS PHR-C is funded, in part by, by the Center for eHealth Applications and Services, and is implemented in co-operation with the Computational Biomedicine Laboratory of the Institute of Computer Science, of the Foundation for Research and Technology – Hellas.

References

[1] Dick R S, Steens E B (Eds.). The Computer-Based Patient Record: An Essential Technology for Health Care. National Academy Press. 1991.

[2] European Parliament and Council Directive 95/46/EC of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [Official Journal L 281 of 23.11.1995.

[3] Iakovidis I. Towards personal health record: current situation, obstacles and trends in implementation of electronic healthcare record in Europe, International Journal of Medical Informatics. 1998; 52: 105–115.

[4] Tang P C, Ash J S, Bates D W, Overhage J M, Sands D Z. Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption. Journal of the American Medical Informatics Association. 2006; 13: 121-126.

[5] Detmer D, Bloomrosen M, Raymond B, Tang T. Integrated personal health records: transformative tools for consumer-centric care. BMC Medical Informatics and Decision Making. 2008; 8:45.

[6] Genitsaridi I, Kondylakis H, Koumakis L, Marias K, Tsiknakis M. Evaluation of Personal Health Record Systems through the Lenses of EC Research Projects. Computers in Biology and Medicine 2015; 59:175-185.

[7] Genitsaridi I, Kondylakis H, Koumakis L, Marias K, Tsiknakis M. Towards Intelligent Personal Health Record Systems: Review, Criteria and Extensions, Procedia Computer Science. 2013; 21:327-334.

[8] Kondylakis H, Plexousakis D, Hrgovcic V, Woitsch R, Premm M, Schuele M. Personal eHealth knowledge spaces though models, agents and semantics. International Conference on Conceptual Modeling (ER). 2014; 293-297.

[9] Krukowski A, Barca C C, Rodríguez J M, Vogiatzaki E. Personal Health Record, Cyberphysical Systems for Epilepsy and Related Brain Disorders. 2015; pp 205-238

[10] Jung S Y, Lee K, Hwang H, Yoo S, Baek H Y, Kim J. Support for Sustainable Use of Personal Health Records: Understanding the Needs of Users as a First Step towards Patient-Driven Mobile Health. JMIR Mhealth Uhealth 2017; 5:e19

[11] Kondylakis H, Flouris G, Fundulaki I, Papakonstantinou V, Tsiknakis M. Flexible access to patient data through e-Consent. InProceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare 2015 Dec 22 (pp. 263-266).

[12] Gunter T D, Terry N P. The Emergence of National Electronic Health Record Architectures in the United States and Australia: Models, Costs, and Questions. J Med Internet Res 2005; 7:e3.

[13] Meaningful Use Definition Objectives [Internet]. Available from: https://www.healthit.gov/providers-professionals/meaningful-use-definition-objectives.

[14] Health Level Seven International [Internet]. Available from: http://www.hl7.org/

[15] Health Level Seven, Inc, PHR-S FM Personal Health Record System Functional Model. [Internet]. Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=88

[16] Stage 2 Eligible Professional (EP) Meaningful Use Core and Menu Measures Table of Contents, October 2012 [Internet]. Available from: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/Stage2_MeaningfulUseSpecSheet_TableContents_EPs.pdf

[17] http://www.himss.org/sites/himssorg/files/FileDownloads/2014-12-01-UsingPatientPortalstoAchieveMU2-EP.pdf

[18] HIMSS. Using Patient Portals to Achieve Meaningful Use Stage 2 [Internet]. Available from: http://www.himss.org/sites/himssorg/files/FileDownloads/2014-12-01-UsingPatientPortalstoAchieveMU2-EP.pdf

[19] Ford E W, Hesse B W, Huerta T R. Personal Health Record Use in the United States: Forecasting Future Adoption Levels - J Med Internet Res. 2016 Mar; 18: e73.

[20] American National Standards Institute [Internet]. Available from: https://www.ansi.org/

[21] ISO/HL7 16527:2016(en) [Internet]. Available from: https://www.iso.org/obp/ui/#iso:std:iso-hl7:16527:ed-1:v1:en

[22] Hoell B, Spat S, Plank J, Schaupp L, Neubauer K, Beck P, Chiarugi F, Kontogiannis V, Pieber R T, Holzinger A. Design of a mobile, safety-critical in-patient glucose management system, in proceedings of International Conference of the European Federation for Medical Informatics - MIE2011.

[23] Bucur A, van Leeuwen J, Christodoulou N, Sigdel K, Argyri K, Koumakis L, Graf N, Stamatakos G. Workflow-driven clinical decision support for personalized oncology. BMC medical informatics and decision making 2016; 16 :87.

[24] Kondylakis H, Koumakis L, Hänold S, Nwankwo I, Forgó N, Marias K, Tsiknakis M, Graf N. Donor's support tool: Enabling informed secondary use of patient's biomaterial and personal data. International Journal of Medical Informatics 2017; 97: 282-292.

[25] Kondylakis H, Spanakis M, Sfakianakis S, Sakkalis V, Tsiknakis M, Marias M, Xia Z, Yu H Q, Dong F. Digital Patient: Personalized and Translational Data Management through the MyHealthAvatar EU Project, International Conference of the IEEE Engineering in Medicine and Biology Society of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, Milan, Italy.

[26] Kondylakis H, Bucur A, Dong F, Renzi C, Manfrinati A, Graf N, Hoffman S, Koumakis L, Pravettoni G, Marias K, Tsiknakis M, Kiefer S. iManageCancer: Developing a platform for Empowering patients and strengthening self-management in cancer diseases, 30th IEEE International Symposium on Computer-Based Medical Systems - IEEE CBMS.

[27] Katehakis DG, Halkiotis S, Kouroubali A. Materialization of Regional Health Information Networks in Greece: Electronic Health Record Barriers & Enablers. Journal of Healthcare Engineering. 2011; 2 :389-403.

[28] Roehrs A, da Costa C A, da Rosa Righi R, de Oliveira K S F. Personal Health Records: A Systematic Literature Review. Journal of Medical Internet Research. 2017; 19: e13.

[29] Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption. Journal of the American Medical Informatics Association, 2006; 13: 121-126.

[30] Baird A, North F, Raghu T S. Personal Health Records (PHR) and the future of the physician-patient relationship. In Proceedings of the 2011 iConference (pp. 281-288). ACM.

A Flexible Solution for Privilege Management and Access Control in EHR Systems

G Gazzarata[1,2], B Blobel[3,4,5], M Giacomini[1,6,7]

[1] Department of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa, Italy

[2] Institute of Social Medicine and Health Economy, University of Magdeburg, Germany

[3] Medical Faculty, University of Regensburg, Germany

[4] eHealth Competence Center Bavaria, Deggendorf Institute of Technology, Germany

[5] First Medical Faculty, Charles University Prague, Czech Republic

[6] Healthropy s.r.l., Savona,Italy

[7] HL7 Italy

Abstract

Background: Inter-organizational healthcare businesses are ruled by a huge set of policies: legal policies, organizational policies, medical policies, ethical policies, etc., which are quite static, patients policy and process, social and environmental conditions, which are highly dynamic. In the context of a business case, those different policies must be harmonized to enable privilege management and access control decisions.

Objectives: The authors offer a methodology to achieve interoperability through policies harmonization in a privilege management and access control solution for EHR systems, to be later on implemented in a cancer care network using HL7 specifications.

Methods: To meet the objective, the authors make use of a system-theoretical, architecture-centric, ontology-based approach to formally representing the aforementioned polices for harmonization.

Results: Because of its flexibility and generality, a policy-driven RBAC model is used to formally represent all the other access control models such as MAC, DAC, RBAC, ABAC, HL7 Data Segmentation and Labeling Services. All the policies deployed in the context of an inter-organizational collaboration for cancer care can be formalized and then harmonized.

Conclusions: The authors provide an implementation-independent methodology to enable policies harmonization in EHR systems. The methodology described in the paper is independent on the maturity of organizations' privilege management and access control system. Furthermore, it does not hamper organizations progressing to more advanced solutions over the time. Even dynamic policies can be harmonized at run time, allowing advancement towards a patient-centered care.

Keywords

Electronic Health Record (EHR); Privilege management; Access control; Policy management; Healthcare Privacy and Security Classification System (HCS)

Correspondence to:

Giorgia Gazzarata

Department of Informatics, Bioengineering, Robotics and System Engineering, University of Genoa Via All'Opera Pia 13, 16145, Genoa, Italy.

E-mail: giorgia.gazzarata@gmail.com

EJBI 2017; 13(1):59-66

received: June 18, 2017

accepted: July 01, 2017

published: October 10, 2017

1 Introduction

The University of Genoa – supported by the Institute of Social Medicine and Health Economy at the University of Magdeburg, Germany – currently engages in the establishment of a cancer care network combining regional healthcare establishments at primary, secondary and tertiary care level. Breast cancer is the most frequent type of cancer, establishing 30% of the cancers females in Italy are suffering from [1]. In females, it is the first cause of death [2]. Breast cancer care is a multi-disciplinary challenge involving different specialties and units in a hospital, but can also cross-organizationally include different hospitals, clinics, practices and laboratories.

For improving quality and efficiency of care delivery, health systems around the globe are evolving towards

inter-organizational, inter-regional and even international communication and cooperation, increasingly based on Electronic Health Record (EHR) systems.

Two of the most important prerequisites to inter-organizational collaboration are security and privacy for establishing trust between the actors involved in the business case including the patient. Security aims at guaranteeing information availability, confidentiality, integrity, authenticity and accountability. Privacy is a human right for self-determination, respecting legal requirements, ethical principles, personal preferences and expectations regarding collection, processing, communication and use of personal data, thereby preventing harm from disclosure of that personal information [3]. One of the key elements to provide security and privacy is privilege management and access control to information and functionalities. Figure 1 illustrates the schema of a general access control system [4].

The business case deals with access to clinical objects stored in the EHR, which is not a banal information sharing. Permissions have to be managed in a way that only the medical staff involved in the patient care can access a patient's clinic information according to the 'need to know' principle. It is important to underline that both communication security and application security are relevant. Using an EHR system, the authors will focus on application security, considering communication security as a prerequisite and not healthcare specific.

In healthcare context, policies that have to be respected are a complex mix of legal, organizational, functional, medical, social, ethical and technical aspects [3]. In addition, also personal wishes, local, timely, contextual and environmental constraints have to be considered [5]. We can distinguish at least legal and domain-specific policies, organizational policies, process-related policies as well as personal policies. Domain-specific policies are, e.g., the Hippocratic Oath, the medical code of conduct and ethical principles. The patient consent is frequently mistakenly called a personal policy. However, it is just the agreement or disagreement with an organizational policy [6]. For distributed business cases, also security and privacy management must be realized

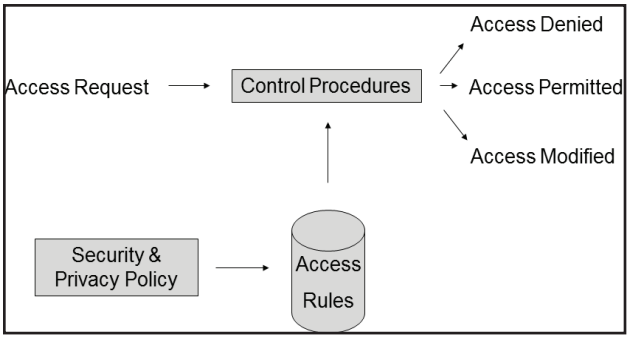


Figure 1: Access control system [4].

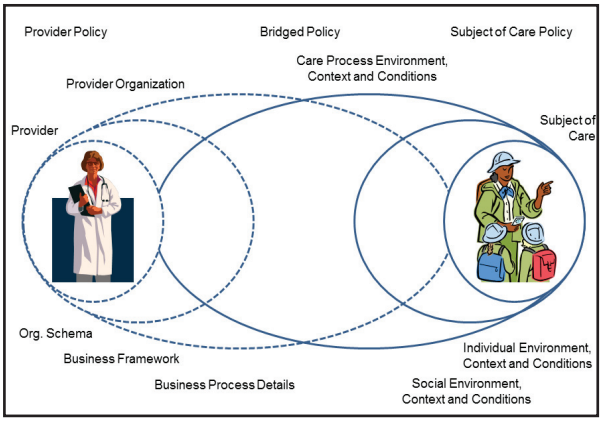


Figure 2: pHealth interoperability schema [7].

in a distributed way. Figure 2 presents the policies and the conditions defining the relation between a physician and a patient [7]. This is the most complex relation, because it involves a large number of policies: legal policies, organizational policies, ethical policies, etc., which are quite static, the subject of care policy and process, social and environment conditions, which are highly dynamic. In the context of a business case, those different policies must be harmonized to enable privilege management and access control decisions. In general, the more the subjects, their knowledge, experiences and skills are close and implicitly sharable, the easier is policy harmonization.

In this paper, the authors want to offer a system-theoretical, architecture-centric, ontology-based, policy-driven approach to achieve interoperability through policies harmonization in a privilege management and access control solution for EHR systems, to be implemented using HL7 specifications.

2 Principles and Methodologies

In order to realize appropriate privilege management and access control, it is fundamental to find a thorough model, which is an abstract representation of that part of the reality the business case deals with. The authors will deploy system theory for that purpose.

A system is group/composition of elements separated from the environment according to properties or needs in the context of a business process. A system could be a part of a super-system, or it can be split in subsystems. A system could be analyzed in two different ways, resulting in the black box approach and the white box approach. The black box approach assesses the system's input-output functional relationship. With this approach, we can describe the overall function of the system without understanding the internal processes and the reasons behind. To carry out greater control, it is necessary to move from the black box to the

white box approach. With the white box analysis, a system is conceived as a collection of interrelated elements. That way, a system is more than just the sum of its components. A system can be represented through its architecture, describing the elements that compose the system, their functions and relationships. The complete understanding of structure, function and relationships of the system elements allows controlling the system itself. The set of rules that controls the behavior of the system is named policy. For managing a quite complex business case, the authors have to use the white box approach.

An ontology is a formal "explicit specification of a conceptualization" [8] of the domain of interest. An ontology defines a controlled vocabulary and represents domain knowledge in a formal and structured form. It consists of concepts, concept definitions and relations between those concepts. A formal description decreases language ambiguity. It allows to make domain assumption explicit, to share knowledge between agents, to re-use and to analyze knowledge [9, 10].

3 Results

In a business process, an entity requests access to some information objects. The purpose of privilege management is to provide the permissions, if any, the entity has, deciding on when, where, why, for which purpose, how, under which conditions. Then, according to the assigned permissions, the request of access to the resource has to be permitted, denied or modified (as instance veiling some information).

In order to guarantee an appropriate privilege management and access control, a prior observation of the business is fundamental. Through observation, interpretation and understanding of the system, its relationships and its rules can be derived and represented as knowledge. Accepted ontologies enable creation, representation and management of the knowledge about the domain of interest in a consistent, reasonably expressive and formalized way that properly reflects the reality.

For realizing interoperability, i.e. advanced communication and cooperation, the different policies must be harmonized. If policies are static and can therefore be predefined, policy harmonization can be performed in advance by coordinating or aligning policies relevant for the predefined business case in the design and definition phase already. This is, e.g., the case when defining, negotiating and contracting disease management programs (DMPs). We must have in mind however the impossibility of predefining any thinkable policy harmonization, e.g. in open care settings and personalized health, as some of the policies to be applied are not known at that time. Furthermore and even more relevant, the consideration of any thinkable policy would make the system too complex, and therefore undefined and not computable. In consequence, we should only consider policies relevant for the corresponding business case. If the harmonization of business case related policies cannot be performed in advance, it has to be performed at run time, using decision intelligence systems for security and privacy services, especially for a specific privilege management. Dynamic policy harmonization can be performed by a system-theoretical, architecture-centric, ontology-based, policy-driven approach.

Figure 3 shows the policy-driven RBAC schema, which is provided by ISO 22600 [6].

In this schema, the principal is the user who wants to access a resource (an information object as target or a service). The structural role is the role assigned to the user by the organization, such as head physician, medical doctor, nurse, etc., but also specific qualifications or competences. The structural roles policy represents the relationships within the organization and is quite static. The process policy is the set of rules that control the business process. In the healthcare context, a process policy can be established through clinic guidelines or best practice guidelines. The functional role is the role that the user has related to the process, which is connected to the actions he/she can perform on resources in a certain process act, e.g. ordering an observation, justifying a statement by signature, or prescribing a medication, so becoming a requester, signer, prescriber. In the business case described, the target policy is strongly influenced by the fact that information objects are clinic information stored in the EHR. Thanks to its flexibility, this model is able to formally represent all the other access control models. This means that in the context of an inter-organizational collaboration all the policies can be formalized and then harmonized without touching them. So, cross-organizational interoperability can be provided irrespective of the access control level and the underlying access control model of the single organization. The model in Figure 3 can be extended by top level policies any business is bound to such as legislation, ethical rules, etc. In order to harmonize policies, it is necessary to obtain a consistent formalization of the policies themselves.

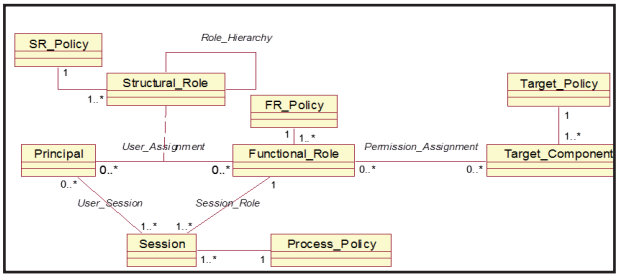


Figure 3: Policy-driven RBAC schema [6].

Following the policies, rules can be applied to assign privileges to actors or to roles played by actors. While the first case allows binding static policies to individual actors, the second enables an easier manageable rather coarse-grained binding of policies to roles, which in the worst case are structural roles and therefore static. When considering functional (predefined) roles, the business process defines the roles and privileges assignment to an entity. As an outcome, we implement statically Access Control Lists (ACLs), Mandatory Access Control (MAC), Discretionary Access Control (DAC), or Role Based Access Control (RBAC).

HL7 provides a special policy-driven solution for managing security and privacy in a business case and an individual context, and for deciding on that basic on privileges at runtime. Within the US national project of data segmentation for advancing communication and cooperation between healthcare establishments, HL7 has specified the HL7 Healthcare Privacy and Security Classification System (HCS) – Release 1 [11]. This specification defines security labels as markers bound to a resource, which connect an information object, but also process steps or actions, to a set of security and privacy attributes. This solution has already been demonstrated at different HIMSS events [5, 12].

The HL7 Healthcare Privacy and Security Classification System consists of two parts:

- A context-sensitive segmentation of health information;
- Security and privacy labeling of data segments, enabling machine processing of privilege management and access control.

In HCS specification, security labels are defined as meta-data bound to resources that transmit constraints on the use of the resources. Security labels are applied based on risk assessment of harm resulting from unauthorized disclosure. "This assessment may reflect personal perceptions or legal requirements, which may involve inherently emotional characterization of clinical information as prejudicial to a party's "interests" when exposed in unauthorized ways or to those who lack authority and responsibility for its care and use" [11]. With the implicit knowledge stored in security labels as "mini policy", privilege management and access control decisions can be performed without accessing the target information. The label refers to the explicit policy stored in a policy repository to be accessed when needed for interpretation as explained in some more detail as follows.

NIST FIPS PUB 188 specification defines a security label as a set of specified fields. Each field consists of globally unique Tag Set Name and a set of semantically interoperable security tag or field values. These labels define the classification of each item. HL7 HCS specifies a Security Classification Tag Set (Confidentiality), a Security Categorization Tag Set (Sensitivity, Integrity, Compartment, Privacy Law), and Handling Caveat Tag Set (Purpose of Use, Obligations, Refrain Policies). In the following, the label fields will be introduced in some details [13]:

- Confidentiality: classifies an IT resource (clinical fact, data, information object, service, or system capability) according to its level of sensitivity, which is based on an analysis of applicable privacy policies and the risk of financial, reputational, or other harm to an individual that could result from unauthorized disclosure;
- Sensitivity: categorizes the value, importance and vulnerability of an IT resource perceived as undesirable to share;
- Integrity: conveys the completeness, veracity, reliability, trustworthiness and provenance of an IT resource;
- Compartment: "segments" an IT resource by indicating that access and use is restricted to members of a defined community or project. An example for compartment labels is "for pharmacy only";
- Privacy Law: refers to the corresponding legislation;
- Handling Caveat: conveys dissemination controls and information handling caveats, such as constraining the purpose of use, defining concrete refrain policies and obligations to which an IT resource custodian or receiver must comply.

Confidentiality, Sensitivity, Integrity and Compartment fields characterize security and privacy rules ("mini policies") for specific health information. Instead, handling caveat fields are characteristics of activities, such as processes of using that information. The valid security labels and how they have to be compared with the users' clearances have to be expressed as explicit policies specified in the Security Policy Information Files (SPIF). The SPIF are usually XML based [5].

The newest project established at HL7 for privilege management and access control is the draft specification "Privacy and Security Architecture Framework – Trust Framework for Federated Authorization, Release 1. Being more consistent with the prosed methodology than older specifications, also this model can be represented and harmonized with others following the presented approach [14].

For managing and harmonizing the different privilege management and access control models, Figure 3 is used as Reference Architecture Model of the privilege management and access control system. For that purpose, all those models have to be architecturally represented in that schema.

All the different policies provided by the different solutions presented have to be considered for the specific privilege assignment and access control decisions. For concluding on all those policies, the concepts have to be provided at a level of expressivity and formalization allowing that all concepts established by domain experts or laymen from different domains in different context with different education, experiences and skills can be appropriately and consistently taken into consideration. For that purpose, the concepts of the policy domain have been ordered and interrelated in the policy domain ontology as described in ISO 22600 [6]. Additionally to the definition of the base classes of that ontology and their structural relations, the latter must be quantified using a proper logic representation. The knowledge (concept) processing in the decision making process is based on an ontology harmonization process. An extended study of relevant tools to perform this task is underway as well.

More details related to access control models and their policies formalization will be presented in [15].

For implementing the policy decision process, existing standards and related services can be used as shortly discussed in the following.

Within the scope of the HCS and its use in an access control system, there are two principal services: the Security Labeling Service (SLS) and the Privacy and Protective Services (PPS) [16]. The SLS evaluates the submitted clinical information objects, including clinical tagging and provenance, to determine the appropriate security labels to assign to information objects for access control based on rules. Access Decision Services can then use the labeled clinical objects as classified resource Access Control Decision Information (ADI) to check clearances. Access decision policy can be dynamic, particularly in the case of patient preferences. For this reason, labels should be applied at runtime, rather than being permanently stored with information objects. In this way, classified resource ADI could be current with the most current policy. The SLS is supported by a Security Label Management System. The latter establishes, provisions, and manages the security tagging vocabularies and security labeling rules needed to support jurisdictional, organizational privacy and security policies, including patient consent directives. Once an access control decision is made, also obligations should be met before releasing the resources. Then, the Policy Decision Point (PDP) decision and obligations are provided to a Policy

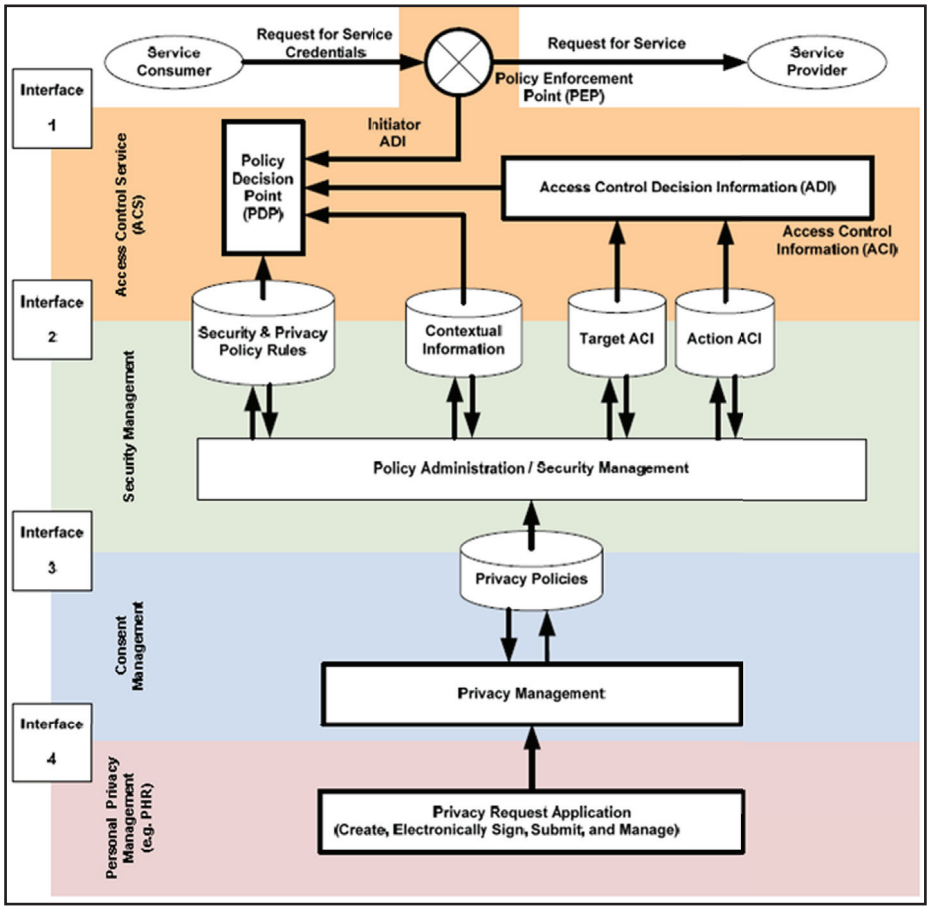


Figure 4: Authorization Reference Model [18].

Enforcement Point (PEP), which tasks appropriate obligation services, such as PPS, to impose the obligations. Basing upon rules, the PPS can apply various transforms to the security labeled resources: masking, redaction, shedding, shifting, annotations, anonymization, pseudo-anonymization, etc. The PPS is supported by its own Protective Services Management Sub-System, which establishes the type of transformations to be applied based upon rules. The latter can be determined in advance or dynamically at runtime. The transformed resource is finally sent to the recipient [5].

Once users' clearances, resources security labels and SPIF are defined, privilege and access control management in health information systems can be automated [5].

For implementing the aforementioned advanced service, the HL7 Implementation Guide: Data Segmentation for Privacy (DS4P) as well as HL7 Version 3 Standard: Privacy, Access and Security Services; Security Labeling Service [16, 17] have been specified. This Implementation Guide, including the value definitions and references, must be localized for the Italian environment. Figure 4 presents the Access Control logical architecture model [18].

4 Deployment of the Developed Methodology

The aforementioned privilege management and access control models refer to predefined policies in their informational representation. As the informational representation is usually defined by informaticians, the consistency with real world policies cannot be guaranteed. The solution offered by the paper is overcoming those limitations by:

- The definition of all policies relevant for a specific business case using the domain specific terminologies;
- The formal representation of those policies using the domain specific ontologies;
- The harmonization of real world policies at runtime.

Then, it is intuitive that static and rigid policies are not suitable. Instead, policies coming from different domains have to be mapped dynamically and in an adaptive and automated way. In order to allow this, it is essential to provide a formal description of the policies that can be used by the authorization services to obtain the security and privacy rules to apply to the resources [19]. Furthermore, the environment conditions have to be evaluated in the same moment in which the user makes the access request. To implement an access control

- To service functionalities: it is necessary to make a functional description of the service, specifying security and privacy minimum requirements for each functionality through security labels, or at the next level by dynamically representing the related explicit policy;
- To the resources: the resources have to be classified through security labels, or at the next level by dynamically representing the related explicit policy.

Through the description of the security and privacy labels (or the explicit policies) of processes and resources, the authorization service is able to obtain the policy to be applied and to check the resulting constraints at runtime. As mentioned above, the policies that have to be mapped come from different domains. So, policies must be formally expressed for allowing their integration in advanced e-health environments. Since different expression means will be used to formally modelling policies, measures and tools for expressing and mapping them have to be developed. Interoperability between different domains requires ontology management by harmonizing common ontology domains' sub-ontologies (e.g. harmonizing concepts of medical sub-ontologies such as SNOMED and LOINC) or by linking different domains' ontologies (e.g. linking policy and medical concepts or linking legal policy and medical policy). Harmonizing ontologies can be performed a-priori by merging, aligning, integrating, or at runtime by matching or mapping. Matching addresses the management of equivalent concepts, while mapping addresses the management of similar concepts. As ontologies are used to represent architectural components of the Generic Component Model (GCM) at appropriate level of generalization/specialization, the GCM process principles also apply to the ontology management (e.g., only interrelating concepts at the same level of granularity).

5 Discussion

In this paper, the authors offer a solution for policy harmonization for privilege management and access control in healthcare context to be implemented using HL7 specifications. Personal health information and related process information will be managed using the Italian Fascicolo Sanitario Elettronico (FSE). The FSE is a regional EHR approach to collect clinical data and documents produced by present or past clinical events and constitutes the patient clinical history. The FSE can be accessed through the Internet with appropriate security and privacy measures in place. The patient can have access to his/her FSE through personal credentials or a smartcard [20].

In order to guarantee adequate privilege management and access control, it is necessary to identify the subjects who request to access to the resources. For this reason, the solution must include an authentication service. The authors focused on authorization services, supposing that an adequate system for identification and authentication already exists. This

assumption is justified by the current Italian governmental effort to set up a national identity system called Sistema Pubblico di Identità Digitale (SPID), Public System of Digital Identity [21]. As its name suggests, the SPID is a national service that provides a digital identity for Italian citizens. The latter are identified through the fiscal code (Codice Fiscale), which is an alphanumeric code of sixteen characters that is associated to Italian individuals at birth and to foreigners when having contacts with Italian institutions. The fiscal code depends on the subject's name, surname, sex, place and date of birth and is unique and tailored for the person. Up to now, the SPID supplies two levels of authentication [22]. At the first level, one factor authentication is provided by a password. At the second level, two factor authentication is provided by a password and a One Time Password. During 2017, a more secure two factor authentication, provided through a password and a physical medium (such as a smart card) should be available for first services. Efforts similar to the SPID are also performed in other European countries with the opportunity of cross-border use, see the Electronic identification and electronic Trust Services (eIDAS) regulation [23, 24]. However, a strong and secure three factor authentication is fundamental for activities as awkward as healthcare.

In Italy, healthcare organizations have different access control models: in some organizations access control is static, rigid and strongly hierarchy dependent (MAC); in others authorized subjects can delegate permissions to other users (DAC); in rare cases, it is possible to find a role based privilege management (RBAC). All of these solutions have been the result of past investments of money and in many cases organizations have not the possibility to make progress towards more advanced access control. The methodology provided by the authors enables at run time the harmonization of all the policies irrespective of the access control model used in the organizations. This offers the opportunity to consider also patient policies, which are strongly dynamic, that way enabling the move from an organization-centered care to a patient-centered care. In addition, since all the access control models can be formally represented with the schema in Figure 3, the methodology does not hamper organizations progressing to more advanced privilege management and access control solutions. These are the advantages of the authors' approach over other solutions that enable harmonization only by static pre-coordination or even require the use of identical or at least equivalent solutions. As a consequence of such pre-coordination, only organizations with the stated access control model can join the intra-organizational collaboration. In addition, since policies are not harmonized at run time, a predefined solution can fit only organization centered healthcare.

6 Conclusion

The authors provide a methodology to enable policies harmonization in EHR systems by deploying a system-theoretical, architecture-centric, ontology-based, policy-driven approach, which:

- Is irrespective of the maturity of organizations privilege management and access control system;
- Does not hamper organizations progressing to more advanced privilege management and access control system over the time;
- Accepts also dynamic policies;
- Allows to advance towards a patient-centered care;
- Is implementation-independent.

References

[1] http://www.airc.it/cancro/cos-e/statistiche-tumori-italia

[2] http://www.airc.it/tumori/tumore-al-seno.asp

[3] Blobel B. Security and privacy services in pathology for enabling trustworthy personal health. Stud Health Technol Inform. 2012; 179: 203-217.

[4] Blobel B, Koeppe D. Handbook of Data Protection and Data Security in Health and Social Care, 4th ed. Frechen: Datakontext – Fachverlag; 2016 (in German).

[5] Blobel B, Davis M, Ruotsalainen P. Policy management standards enabling trustworthy pHealth. Stud Health Technol Inform. 2014; 200: 8-21.

[6] International Organization for Standardization. ISO 22600 Health informatics – Privilege management and access control. Geneva: ISO; 2014.

[7] Blobel B, Ruotsalainen P, Lopez DM, Oemig F. Requirements and Solutions for Personalized Health Systems. Stud Health Technol Inform. 2017; 237: 3-21.

[8] Gruber TR. Toward principles for the design of ontologies used for knowledge sharing. Int J Human – Computer Studies. 1995; 43: 907-928.

[9] Blobel B. Knowledge Representation and Management Enabling Intelligent Interoperability – Principles and Standard. Stud Health Technol Inform. 2013; 186: 3-21.

[10] Blanco C, Lasheras J, Valencia-García R, Fernández-Medina E, Toval A, Piattini M. A Systematic Review and Comparison of Security Ontologies. Proceedings of The Third International Conference on Availability, Reliability and Security. Barcelona, Spain, 4-7 March 2008.

[11] HL7 International Inc. HL7 Informative Guidance: Guide to the HL7 Healthcare Privacy and Security Classification System (HCS). Ann Arbor: HL7 International; May 2013.

[12] Data Segmentation for Privacy VA/SAMSA RI/ Pilot at HIMSS 2013: http://wiki.siframework.org/ Data+Segmentation+for+Privacy+Homepage.

[13] HL7 International Inc. HL7 Healthcare Privacy and Security Classification System (HCS) – Release 1. Ann Arbor: HL7 International; May 2013.

[14] HL7 International Inc. HL7 Version 3 Standard: Privacy and Security Architecture Framework – Trust Framework

for Federated Authorization, Release 1. Ann Arbor: HL7 International; January 2017.

[15] Gazzarata G, Blobel B, Giacomini M. A system-oriented, policy-driven approach to harmonize existing privilege management and access control models. In preparation.

[16] HL7 International Inc. HL7 Version 3 Standard : Privacy, Access and Security Services; Security Labeling Service, Release 1. Ann Arbor: HL7 International; January 2014.

[17] HL7 International Inc. HL7 Implementation Guide: Data Segmentation for Privacy (DS4P), Release 1. Ann Arbor: HL7 International; September 2013.

[18] HL7 International Inc. HL7 Version 3 Standard: Privacy, Access and Security Services (PASS) – Access Control Services Conceptual Model, Release 1. Ann Arbor: HL7 International; September 2015.

[19] Blobel B. Ontology driven health information systems architectures enable pHealth for empowered patients. Int J Med Inform. 2011; 80: e17-e25.

[20] Fascicolo Sanitario Elettronico (FSE): http://support.fascicolo-sanitario.it/faq

[21] SPID (Sistema Pubblico di Identità Digitale): https://www.spid.gov.it/.

[22] SPID (Sistema Pubblico di Identità Digitale): http://www.agid.gov.it/agenda-digitale/infrastrutture-architetture/spid/percorso-attuazione.

[23] Trust Services and eIdentification: https://ec.europa.eu/digital-single-market/en/policies/trust-services-and-eidentification.

[24] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.257.01.0073.01.ENG

Standardizing the Transmission of Genomic-Related Medical Data for the Optimization of Diagnostics and Research in the Field of Oncological Diseases

Elisabeth Pantazoglou#, Martin Thoma#, Lasse van de Sand, Heike Dewenter, Sylvia Thun

Niederrhein University of Applied Sciences, Faculty of Health Care, Competence Center e-Health, Krefeld, Germany

Abstract

Introduction: In Germany, there is currently no consistent analytic structure within genomic diagnostics in oncological diseases. Within the framework of the project GENeALYSE, a standardized and interoperable specification for associated uses cases shall be developed.

Intended Methods: Through process analysis and interface modeling, problems of the actual processes will be depicted between the involved actors. In the next step, the workflows and relevant findings will be displayed and adapted. In particular, the heterogeneous workflows in genome diagnostics will be represented by semantic annotation in an international terminology. The results of the semantic annotation build the basement for the creation of an implementation guide for standardized genome analytics, referring to HL7 Clinical Document Architecture (HL7 CDA).

Discussion: The problems of heterogeneous genomic diagnostics as well as unstructured findings in oncology leave the actors face comparable challenges on a regional and supranational level. Interfaces, ambiguous semantics and manual activities inhibit interoperability, promote errors and lead to risks for patients and their sufficient medical treatment. A major challenge will be consistency between the heterogeneous terms to be found in genome analysis. The problem shall be addressed via using international terminologies as well as appropriate mapping techniques.

Conclusions: The aim of the project is to create an implementation guide for standardized digital documentation and communication solutions between diagnostics, medical therapy and research in the field of genome analysis. GENeALYSE is intended to optimize the coordination between the diagnostic genome laboratory and the clinical therapy decision in order to increase the safety and success of medical treatment, as well as to improve the health-related quality of life of the affected patients.

Keywords

Interoperability; Semantics; Clinical Document Architecture (CDA); Genetics

Correspondence to:

Elisabeth Pantazoglou, B. Sc. Health Care Management
Niederrhein, University of Applied Sciences, Faculty of Health Care, Reinarzstrasse 49, Krefeld, Germany.

E-mail: elisabeth.pantazolgou@hs-niederrhein.de

 # Contributed equally

EJBI 2017; 13(1):67-71
received: June 09, 2017
accepted: July 16, 2017
published: October 10, 2017

1 Introduction

Currently, cancer is one of the most common diseases in the German society. Based on statistic data from the National Statistical Office in 2013 breast cancer, lung cancer and colorectal cancer were identified as the most frequent causes of death for women placed on rang six, seven and ten. For men the most frequent causes of death were lung cancer, colorectal cancer and prostate cancer on rang two, six and seven [1]. In the same year the incidence of cancer reached 482.000 new cases and a total of 223.000 patients died because of this disease. The ten-year prevalence was 2.6 million. According to a report about cancer in Germany published in 2016, about four million people suffered from cancer in their life time [2].

The underlying causes of cancer are complex, but the main reasons for the rising number of newly reported cases of cancer are a rapidly increasing life expectancy and the reduction of other life-threatening diseases. In general, there is also a discussion about consumption of alcohol and tobacco as well as certain aspects of nutrition relating to cancer incidences [3, 4].

Nowadays the most common therapy against cancer is chemotherapy. With this treatment approach chemical

substances, the so-called cytostatic drugs, are administered to the patient with the aim to stop the division of the malignant cells. This form of therapy can be regarded as extremely radical, because cytostatic drugs do not specifically affect cancer cells, but also every form of cells inside the human body. In this context chemotherapy can be described as a severely impairing therapy option [5].

Even if chemotherapy, especially in association with radiotherapy, is very efficient on cell destruction there are very strong side effects for the patients and a high risk of an absent treatment success or the return of recurrences. The patients suffer extremely with this treatment. Because of the massive effects and adverse side effects of chemotherapy the patient's quality of life drops rapidly. In some cases these effects cause other disorders or severe health consequences for the patients. However, the lack of a therapeutic effect is the most serious problem in cancer treatment. The reasons for reduced treatment success are often resistant cancer cells or a late diagnosis of the cancer disease [6].

Advances in molecular biology have brought new information about the development of cancer in recent years. The genetic disposition of a person gets an important role along with external risk factors in the incidence of cancer diseases. In addition, the molecular-genetic diagnostics of tumor cells provides accurate information about possible entry points for efficient therapy options. This has led to new methods in cancer treatment, the so-called "personalized medicine". The new methods of genome sequencing are summarized by the concepts "Next Generation Sequencing"(NGS) and "Precision Medicine". Combined with medical databases they possess the potential of predicting the probability of cancerous diseases and to apply more specific therapy forms and therefore they may improve the quality of life and increase the survival rate of the affected patients [3, 7].

The new methods of the upcoming medicine become more predictive and preventive, as they contribute to a new medicine concept. They form the basis of the paradigm shift in modern medicine far away from the reactive medicine towards a more proactive and personalized medicine, the so-called P4-Medicine [8]. Instead of a general treatment the personalized medicine focuses on two different approaches. The first is used to detect the patient`s potential risk of developing cancer and to initiate preventive steps. For this genome sequencing blood-material or other physical material is needed. The second is used for genome sequence analyses to verify mutations and to identify possible attack points. For that genome sequencing tumor material is needed. In addition to a specific requirement for the detection of the existence of a certain mutation in association with an individual clinical implication, it is possible to use whole genomic (sub-) sequences for diagnostics. This contains the risk, that the requirements to the molecular genetic laboratory are not clearly expressed. Furthermore the interpretation of the determined genome sequence depends especially on the physician`s or natural scientist`s knowledge. However, funded diagnostics require acknowledged reference databases, which may be license- or charge bound. Though the databases support the medical findings, the problem remains, that they for free available platforms a standard procedure doesn't exist [9]. Another difficulty lies in the different output formats and the use of different annotations of genome sequences in the existing databases. A further problem lies in the report of the diagnostic test results. Neither the generation nor the transmission of the report is standardized at the current time. Solely narrative text modules can be used to support generating the diagnosis report. Along with the lack of syntactic standards, semantic reference systems are not applied regularly. Furthermore, the transmission of molecular genetic diagnosis reports is still paper bound [10]. Due to a non-consistent term definition these circumstances may affect the therapeutic decisions for the patients.

The aim of the future research project GENeALYSE is to create an implementation guide for standardized digital documentation and communication solutions between diagnostics, medical therapy and research in the field of genome analysis. GENeALYSE is intended to optimize the coordination between the diagnostic genome laboratory and the clinical therapy decision in order to increase the safety and success of the treatment, as well as to improve the health-related quality of life of the affected patients. The project will be coordinated via the Niederrhein University of Applied Sciences Krefeld, in collaboration with the Institute for Pathology at the University of Cologne and the Clinic for Gynecology and Obstetrics at the University Hospital of Düsseldorf. The located Center for Familial Breast and Ovarian Cancer (ZFBEK-D) is part of cooperation between 17 university centers and is specialized in genomic diagnostic for genetic induced family-related cancer. The Institute for Pathology at the University of Cologne diagnoses genetic modifications on solid tumors. In addition it is a central molecular diagnostic center for patients with lung cancer in the network "Genomic Medicine" in Germany. This collaboration ensures that the findings of genome sequences consider all medicinal domains. The project is also supported by the involvement of associated partners, such as the Cancer Society North Rhine Westphalia (NRW), the Scientific Institute of Office-based Hematologists and Oncologists (WINHO) and the Federal association of German pathologists. The project is scheduled to start in September 2017 with duration of two years and it is funded by the Ministry of Innovation, Science and Research from the federal state of North Rhine – Westphalia (NRW).

2 Intended Methods

2.1 Process Analysis

The methodological basement will be the development of technical basics and the collection of information about

the actual processes. In the beginning, legal and regulatory framework conditions will be analyzed, which significantly influence the medical findings in genome analysis. Within the collection and analysis of the actual processes, scientific methods, e.g. analysis of the vulnerability, stakeholder analysis and questionnaire methodology, the specific workflows and the participatory environment will be evaluated. The goal is to examine the relevant workflows and to represent the main actors and stakeholders.

Associated information will be collected and illustrated using the Business Process Modeling Notation (BPMN) as a diagram Table 1.

Together with the project partners as well as selected experts from the specific areas, the listed questions will be discussed within several project meetings. The areas of clinical requirements, data transmission, sample entry, assessment of findings will be discussed as the main content inside the project meetings. Particular attention is also paid to the requirements concerning the tumor tissue. In contrast to the germ line determination of biomarkers, tumor tissue cannot be easily retrieved. Against this backdrop it is necessary to precisely define requirements for removal, transport and medical questionings concerning tumor tissue. The search for alternative possibilities for action or rather deviating processes as well as the collection and evaluation of exemplary results and reports of results from the experts involved and other relevant materials complete the survey of all necessary information and form the basis of the weakness analysis. Furthermore, the technical conditions will be analyzed. This includes the recording of the involved information systems, the necessary interfaces and transport standards for data transmission. This methodological approach forms the basis for the further elaborations and involves the perspective and the requirements of the actors involved as well as the legal framework. Based on the actual process recording, optimization potentials will be derived and documented in the process of genome analysis. This will determine the requirements for the following target process determination and conception.

Table 1: BPMN-displayed questions.

Clinical questions and requirements Which information is needed by the requesting physician? Which information is available about the patient? How is the data collected and recorded?
Transmission of requirement and sample material How is the transmission performed? Do certain times or conditions of delivery have to be observed?
Delivery of the samples in the laboratory Which tests are carried out? Which questions or problems arise often?
Clinical results Which methods are implemented? Which support systems are used?
Assessment and reporting How is the result obtained? How is the report submitted? Which semantic and syntactic standards are used? What are the contextual problems or questions?

2.2 Process Modeling

Using the collected and analyzed information about the actual processes, potentials will be derived and presented via BPMN. The identification of the target processes have major influence on the development of the semantic reference systems and the implementation guide. In addition, requirements and framework conditions for necessary software and hardware must be formulated. The quality management approach “Six Sigma” with its core concept "Define-Measure-Analysis-Improve-Control" (DMAIC) and the associated tools will be used to create the target processes by following and combining the scientific methodology from the previous steps. In several individual project meetings expert opinions will be incorporated into the development of the target processes [11, 12].

In comparable mode of operation, the research topic will be limited, then separately considered and formulated, and finally validated in the review process. The resulting requirements for the target processes form the basis for the semantic annotation, as well as the creation of an implementation guide. For the modeling, a continuous exchange with the involved actors is indispensable. The work will be reviewed and adapted by a continuous process of improvement in accordance with KAIZEN to ensure the best possible applicability for all stakeholders. Documentation-based quality assurance is an essential part of research and development projects, in order to understand, analyze and correct mistakes in individual steps and changes [12].

2.3 Semantic Annotation and Implementation Guide

Against the backdrop of the technological developments in the field of NGS and related comprehensive genetic diagnostics, the interpretation of the identified sequence variants relating to their possible pathogenicity is very challenging. A current, structured and standardized data query of the existing knowledge for a functioning classification of DNA sequence variants provides the foundation for evidence-based clinical treatment options and therapy decisions. Especially the molecular diagnostic for the detection of a genetic predisposition requires an evidence-based interpretation of the analysis results. Furthermore only the standardization of the diagnosis report with semantic clear terms as well as a standardized communication of the

reports may avoid misinterpretation like wrong therapeutic or preventive decisions. The technical complexity of this medicinal field and the substantial amount of the collected information is a strong challenge for standardization. However, due to the rising relevance of these diagnostic methods and new possibilities in this field, standardization is indispensable.

Semantically clear terms generate the basis for a uniform, content-based understanding in the area of genome analysis. This is the basic prerequisite for the successful cooperation of all stakeholders. Semantic interoperability ensures a smooth exchange of information with a lower risk of misinterpretation. Therefore an analysis of available semantic standards for the representation of gene analyzes is planned. The survey is carried out via scientific literature analysis and in exchange with the experts. This forms the prerequisite for the development of the matching characteristics between the terminologies and their mapping.

First researches have shown that the HGSV nomenclature (Human Genome Variation Society) is a worldwide-adopted standard for naming genetic sequences. It is the result of the so-called "Sequence Variant Description Working Group" (SVD-GV), which is composed of the specialist groups Human Genome Variation Society (HGSV), Human Variome Project (HVP) and the Human Genome Organization (HUGO). The HGSV nomenclature is used for the uniform annotation of specialist terms in genome sequencing. The standard describes among other things the designation of the determined sequence in a particular format as well as in connection with a reference sequence [13].

The use of the nomenclature is also becoming increasingly important for research in databases and is increasingly being demanded in the field of communication [14]. Logical Observations Names and Codes (LOINC) is a nomenclature for laboratory analyzes and clinical observations, which has been developed by the Regenstrief Institute since 1994 and is available free of charge via a database. LOINC terms follow also the recommendations of other nomenclatures such as HGSV. The use of the LOINC Codes is provided in HL7 Clinical Document Architecture (CDA) documents. The ISO-Standard CDA is used to exchange electronic clinical documents in a standardized form. These two standards provide the framework for structured reporting of laboratory test results. In conjunction with gene sequencing they can be used for reports to structure semantically interoperable findings. For this purpose LOINC and HL7 collaborated in two HL7 Clinical Genomic Working Groups to define implementation guides for structured reporting of genetic tests [15]. These implementation guides shall be adapted for German conditions.

Terminological experts carry out the medical coding and the creation of the mapping tables independently. In addition, the ART-DECOR software serves as an online tool for the experts' quality assurance. ART-DECOR is a web-based open source tool which supports the creation and maintenance of CDA templates, value sets, codes and data sets. Its underlying data format is XML. With an extension it can be used for the quality assurance. In addition, the semantic coverage (Figure 1) of the specialized terms as well as the quality assurance will be carried out by classification according to ISO TR 12300 [16]. The mapping tables will be used during the diagnostic evaluation and make a significant contribution to facilitating the currently highly complex searches in scientific genomic databases. Furthermore, the semantic annotation is part of the CDA (Figure 2), which will be represented in ART-DECOR.

The harmonization with international standards such as HL7, as well as e.g. LOINC, HGSV or the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) forms the basis for the creation of the implementation guide. The work will be continually presented at the German interoperability forum and coordinated by the involved experts. Finally, the implementation guide shall provide all necessary information regarding structure and semantics for the standardized and interoperable exchange of results in the genome-sequencing field.

3 Discussion

Although genetic findings in oncology have for some time been a basis for the therapy decision, these results are still unstructured and differ from laboratory to laboratory. Against this backdrop, the lack of structure generates a huge problem in the interpretation of results as well as in research for therapeutic approaches. The unstructured

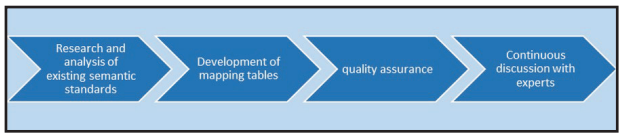


Figure 1: Process of elaboration of the semantic coverage.

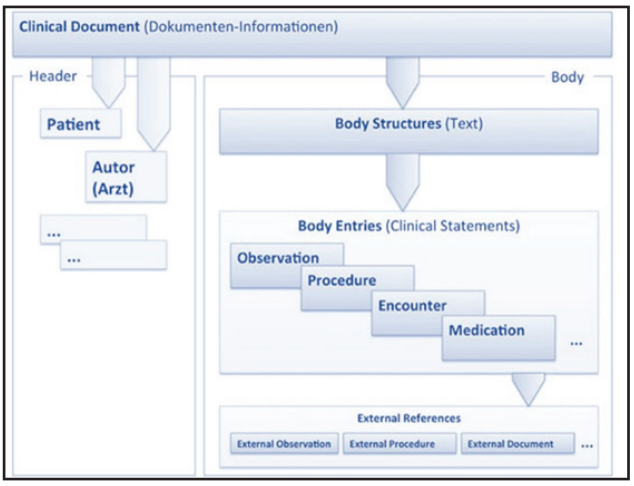


Figure 2: Simplified overview of a part of the CDA model.

report of genomic analyzes in oncology is not a regional problem. Nationwide, the actors face comparable challenges, leading to countless semiprofessional approaches and therapy solutions. The progress and the gain in knowledge in the fields of oncological laboratory analysis as well as the standardization of medical terminologies are progressing rapidly. The current findings are not standardized and developed in different forms depending on the institution. In addition, the results are provided in different quality, depending on the performing laboratory. If unstructured or incorrect information is passed on in the assessment, this may have negative effects on the treatment and therefore on the quality of life and ultimately on the survival of the patients.

With standardized digital techniques, genetic information about cancer cells as well as on biomarkers can be transferred and evaluated in high data quality and with little range of interpretation. Clearly and unambiguously structured results of the genomic sequencing are the basis for the search for targeted, individual therapeutic approaches for each patient. Thus the finding is the most important starting point and the decisive basis for the ultimately to be used therapy. In this project the findings from the fields of analytical laboratory diagnostics with the focus on oncogenes, the scientific analyzes and the development of innovative technical solutions in the eHealth sector as well as the standardization of communication in the health and social sector will be brought together. These three areas can be the basis for improved data quality and contribute to the improvement of communication by establishing a uniform findings structure.

In this way, the treatment quality will be positively influenced. The development of a structured electronic finding on the basis of semantic uniqueness has gained in importance as a result of the increasing interest in the approach of personalized medicine, new success rates in combating cancer and the numerous findings in oncological research. The scope of findings in this field will grow continuously. Therefore, it is important to standardize the basic documentation on data transmission and evaluation. Only a clear structure of the findings and the semantic uniqueness can contribute to better communication and to the improvement of the evaluation of this diverse information. But this aim requires that all above described conditions and procedures in genome analyzing are recorded and taken into account. In summary, the project`s goals will only be realized with the cooperation of experts to provide evaluation and a quality assured mapping of the existing standards.

4 Conclusion

On the one hand, the standardization in the described field reduces the effort of documentation during the analysis in the laboratory. On the other hand, the semantically clear information can be searched more quickly and more precisely for appropriate therapeutic approaches and therapeutic parameters. With this analysis, a finding can be made on the basis of which a personalized therapy can be created. This individual kind of treatment increases the chances of healing, relieves the side effects and generates medical knowledge. Through the optimization process, time resources in the laboratory as well as in the treatment can be performed more efficient.

By reducing the time needed, patients can benefit from a suitable therapy more quickly. Particularly in the case of cancer, the factor time is a decisive criterion for the success of a therapy. However, the optimization of the results cannot only save time. Publicly accessible databases can be browsed faster and more specifically with the given search masks, at least to contribute to better research requirements.

References

[1] Federal Statistical Office Germany. The ten most frequent causes of death; 2013.

[2] Barnes B. Report on cancer in Germany. Berlin: Robert Koch Institute; 2016.

[3] Collisson EA, Cho RJ, Gray JW. What are we learning from the cancer genome? Nat Rev Clin Oncol.2012; 9:621–630.

[4] Ullrich A, Sepulveda C, Yach D, et al. Public Health und Krebsprävention. Der Onkologe 2004; 10:166–174.

[5] Krege S. Allgemeine onkologische Chemotherapie. InFacharztwissen Urologie 2014. Springer Berlin Heidelberg.

[6] Seiffert S. Makromolekulare Chemie. Nachr Chem. 2016; 64:332–340.

[7] Kuß A. Next Generation Sequencing. medgen. 2014; 26:229–230.

[8] Bieber T, Broich K. Personalisierte Medizin. Zielsetzungen und Herausforderungen. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2013; 56:1468–1472.

[9] Kast K, Arnold N. Gendiagnostik bei familiären Tumoren. Gynäkologe 2012; 45:28–34.

[10] Li M M, Datto M, Duncavage EJ, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer. J Mol Diagn. 2017; 19:4–23.

[11] Helbig R. Prozessorientierte Unternehmensführung. Heidelberg: Physica-Verlag; 2003.

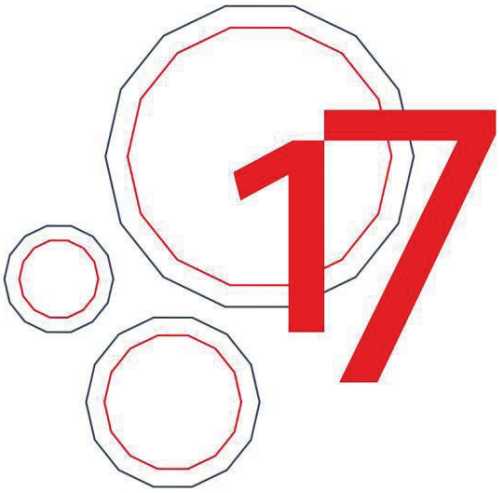
[12] Töpfer A. Lean Six Sigma. Berlin Heidelberg: Springer; 2009.

[13] http://varnomen.hgvs.org/bg-material/simple/

[14] Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et. al. HGVS recommendations for the description of sequence variants: 2016 Update. Human mutation. 2016 ;37: 564-9.

[15] Deckard J, McDonald CJ, Vreeman DJ. Supporting interoperability of genetic data with LOINC. J Am Med Inform Assoc. 2015; 22:621-7.

[16] The International Organization for Standardization ISO/TR 12300:2014 Health informatics - Principles of mapping between terminological systems. Geneva: ISO; 2014.



17th INTERNATIONAL HL7 INTEROPERABILITY CONFERENCE IHIC 2017

Technopolis
City of Athens
22-24 October
2017

17th International HL7 Interoperability Conference (IHIC 2017)

The motto of the 2017 conference is:

"Re-shaping healthcare systems"

The conference aims to play the role of an interface between science, research and practice with regards to interoperability in the health and social care domain to share experiences and best practices.

It will present scientific papers on the one hand and demonstrations, practice reports, success and failure stories on the other hand.

Program Committee

- Bernd Blobel - Co-Chair (DE)
- Alexander Berler - Co-Chair (GR)
- Christof Gessner (DE)
- Kai Heitmann (DE)
- Mauro Giacomini (IT)
- Dimitrios Katehakis (GR)
- Nikos Kyriakoulakos (GR)
- Lina Nikolopoulou (GR)
- Sylvia Thun (DE)

On the Evaluation of HL7 CDA R2 Documents Richness and Validation Reliability

Abderrazek Boufahja, Eric Poiseau

IHE-Europe

Abstract

Many test providers and certification programs define test plans in order to test the conformity of CDA documents against implementation guides. Even if the applications and systems tested provide valid CDA documents, it is not easy for test providers to decide if the validated documents are rich enough to have a good reliability on tested tools; providing the coverage of the areas tested is mandatory based on many test framework specifications. Many projects tried to define a way to describe the richness and providing scoring for validated CDA documents. In this paper, we describe a new methodology to identify the richness of CDA documents based on implementation guides specification. We define a way to provide a scoring for the richness of the CDA documents, with some applications on IHE and C-CDA documents.

Keywords

IHL7; CDA; Scoring; Richness; Validation; Reliability; Templates

Correspondence to:

Abderrazek Boufahja
Kereval 4 Rue Hélène Boucher, 35235 Thorigné-Fouillard, France.
E-mail: abderrazek.boufahja@ihe-europe.net

EJBI 2017; 13(1):73-83
received: June 11, 2017
accepted: July 14, 2017
published: October 10, 2017

1 Introduction

Since the publication of HL7 CDA R2 standard [1], evaluating the richness of a CDA document remains a problematic subject, especially in the context of testing and certifying applications creating CDA documents. In order to comply with the requirements of CDA implementation guides, an editor can be tempted to provide documents containing a minimal set of data, in order to comply with the tests requirements. So unless the test description covers the entire tree of templates in the context of a document, the confidence in the tested application remains uncertain [2]. An indicator of the richness of a tested CDA documents can be useful for tester, in order to complete the outcome of conformance checking tools. There is no complete methodology to automatically express the richness of CDA templates. The only way to calculate it is to manually parse the CDA document and to compare it with the original specification. The aim of this paper is to define a new methodology to calculate and interpret the richness of CDA documents. We will first present the state of the art regarding the scoring of the CDA documents. Then we present our analysis regarding the CDA templates richness and scoring. And finally, we describe the implementation and some applications of the richness scoring in some projects.

2 State of the Art

2.1 Richness of CDA Documents

A couple of papers and articles in the literature mentioned the richness of the CDA documents, as a notion describing "how rich are the clinical information in the CDA document" [3, 4, 5]. Technically this is interpreted by how many templates and CDA elements are present, comparing to the list of possible templates and elements (this include both required and optional templates/elements). The richness of CDA documents refers to the 'A' in CDA: Architecture. It is a way to describe the architecture of the provided documents comparing to the specifications' architecture of templates.

2.2 Scoring of C-CDA Documents

Scoring has been introduced by meaningful use with the creation of a scorecard for CDA document [6]. The scorecard provides a score to a CDA document based on the presence of specific components with the analyzed document. Smart C-CDA Scorecard is a tool that promotes best practices for C-CDA [7]. This tool is one of the first clinical checking tools to provide a scoring of information provided in a CDA

document. The tool is dedicated to C-CDA usage, and based on users experience with C-CDA specification. The tool provides as output a percentage as a scoring and a grade for the C-CDA document tested, and a description of clinical issues for each CDA section.

3 HL7 CDA R2 Templates Richness Analysis

3.1 Presentation and Notions

Test Data Characteristics

During testing process, tested applications can provide two kinds of test data: minimal test data, and relevant test data. A minimal test data contains the minimal architecture of templates allowing passing the validation process by automated validation tools. This kind of documents contains only the required templates, the required elements, and fits well the rules checked by validation tools; however, it is not enough to confirm that the tested application is able to create valid CDA documents in a real use case. A relevant test data is an expression of a real use case with fake data. Testing using relevant test data increases the reliability on the tested tool. These notions are also treated in the Test Framework specification of eHDSI project [8].

Templates Inheritance

The inheritance between CDA templates is a basic notion for the CDA implementation guides. It allows reusing templates rules, without the need to rewrite them. HL7 Templates Standard defines many types of inheritances between HL7 Templates [9]; the most important one for our study is the specialization (SPEC) between CDA templates. The other kinds of templates inheritance are more to be extensions than specializations of the specification rules [9, 10].

Inheritance between templates can be for multiple levels. Example: in eHDSI project, an epSOS medication [templateId: 1.3.6.1.4.1.12559.11.10.1.3.1.3.4] inherit from IHE medication [templateId: 1.3.6.1.4.1.19376.1.5.3.1.4.7], which inherit from CCD medication activity [templateId: 2.16.840.1.113883.10.20.1.24] [11]. In this case, the more specific template needs to fit all the requirements coming from all the parent templates.

Templates Containment

CDA templates can contain many other sub-templates. In this paper, we are only considering the containment as described in the HL7 Templates Standard [9]. A template can be a header template, a L2 template, or a L3 template. The header templates are related to CDA header elements (like the patient, the participants, etc). The L2 templates are related to the sections containments, and the L3 templates are related to the CDA entries. The richness of CDA templates can be related to the L2 or L3 [12].

3.2 Templates Richness Matrix

There are as explained two kinds of relationship between CDA templates: inheritance and containment. We expressed each of these kinds of relationship by matrices: the matrix of inheritance and the matrix of containment. The work performed here is to merge both matrices and to get a complete richness containment relationship.

Example Analysis

We suppose we have this scenario:

- We have eight CDA templates: A, B, C, D, E, F, G, H
- A inherit from B
- B inherit from C
- A contains two templates D and E
- B contains F
- C contains two templates G and H
- D inherit from G
- F inherit from C

The first step to be executed is to know the real inheritance architecture. From the example, there are an extra inheritance path that we can calculate, which is between A and C Figure 1.

Once we have the final inheritance relationships; we copy all the missing templates from the parents into the more specific templates. Example, A inherits from B, so one possible template for A will be F. Here is the diagram of containment computation Figure 2.

Right now, A can have many sub-templates: D, E, F, G, and H. However, we know that D is a specialization of G. So, we can remove the containment between A and G, as it is redundant: when A includes D, it already includes G Figures 3, 4, 5.

So here we have the final result of containments between templates in our example:

Let's now express these operations using matrix description.

We consider the vector $V = [A\ B\ C\ D\ E\ F\ G\ H]$

The first matrix of containment can be expressed by:

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

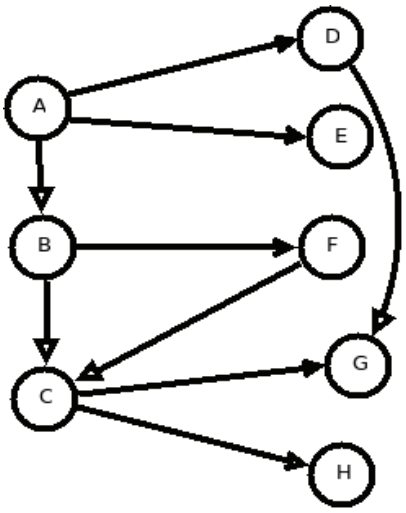


Figure 1: Original containments.

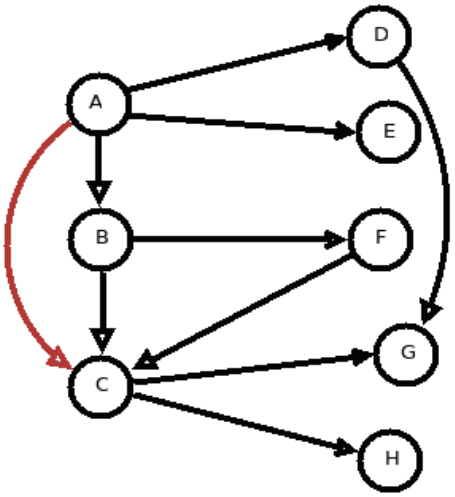


Figure 2: Inheritance computation.

The matrix of inheritance between the templates can be expressed as:

$$B = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The matrixes A and B are the expression of inheritance and containment relationship. These matrices are square matrixes, having as dimension the number of templates used in the specification. The rows and the columns describe the same vector, the vector V of the templates used. Then, to interpret these matrices, if you have 0, it means there are no

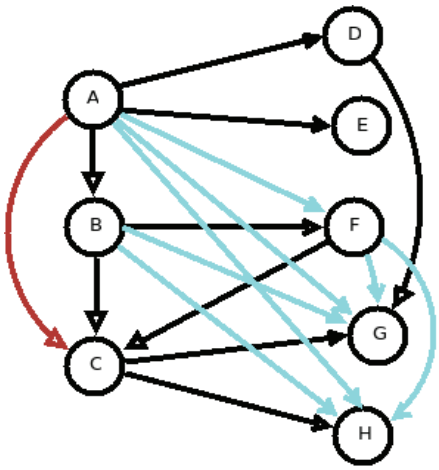


Figure 3: Containment computation.

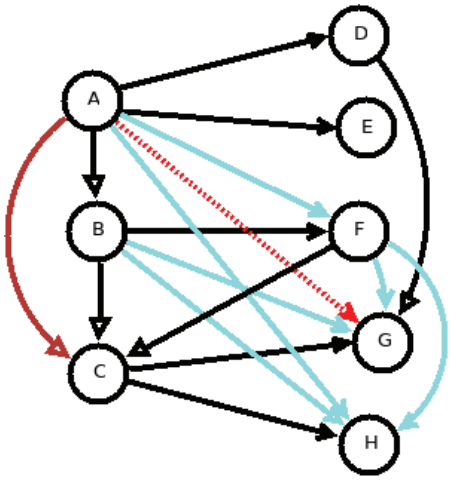


Figure 4: Diagram refinement.

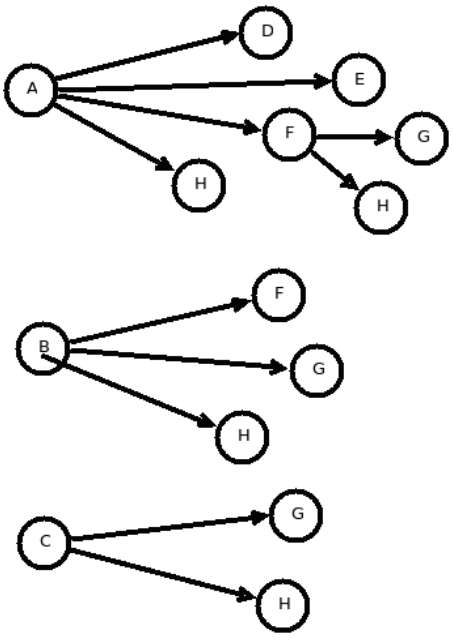


Figure 5: Tree of containments.

containment (or no inheritance) between the two templates (for the row and the column selected), and if we have 1, it means there are a containment (or inheritance). Here for example we have B[0][1] = 1, it means there are inheritance between the template A and B.

The first step was to find the real inheritance architecture, by looking for the parent of the parent of a template. The parent of the parent of the template A is C. To get this mathematically, we need to multiply the matrix B by itself:

$$BxB = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The complete inheritance for our example can be expressed by the matrix:

$$C = B + B^2 = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

We call the C matrix, the complete inheritance description matrix.

The second step was to identify all the possible missing containment in the CDA templates. For template A, it was F, G and H; for template B it was G and H. To get this information mathematically, we need to multiply the matrix C by A (the containment matrix):

$$CxA = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The third step is to clean up the redundant containment. To do so, we need first to find what was included redundantly. In our case, for the template A, we included the template G redundantly as we already include the template D. To find this information we need to multiply the matrix A by C:

$$AxC = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Then we need to remove the redundancies found. The final result of our operations is:

$$(CxA + A) \& \overline{(AxC)} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

And this result fits exactly the output from our analysis of this example.

Richness Computation

From the matrix *A* (for the CDA templates containment) and *B* (for CDA templates inheritance), we are able to extract the possible sub-templates that may appear for a specific templates. The formula allowing computing these containments is described in this equation:

$$\Phi = (CA + A) \& \overline{AC} = \left(\sum_{n=0}^{\infty} B^n A \right) \& \overline{\sum_{n=1}^{\infty} AB^n} \quad (1)$$

Where

$C = \sum_{n=1}^{\infty} B^n$ is the complete inheritance matrix between the different templates.

Explanations

B^n computes the inheritance relationship for the level n.

The matrix C is the complete inheritance architecture for the defined list of templates. As the inheritance between CDA templates is not circular, the matrix *B* can be expressed as a strict triangular matrix, and then it is a nilpotent matrix: $\exists\, n \in \mathbb{N}: B^n = 0$ [13]. And this proves that C is computable.

The matrix *CA* describes the inherited templates containment. *(CA + A)* describes the complete sub-templates architecture. *AC* describes the redundancy in the templates architecture. As *A* describes the templates containment, and *C* describes the templates inheritance, *AC* describes

the parent templates that will be contained automatically if we already include a specialization of them. $\overline{AC}$ describes the templates that can be included without redundancy. To create $\overline{AC}$ from *AC* matrix, we need to replace every "1" by "0" and every "0" by "1". The matrix ϕ describes then the containment relationship between the different templates, by removing all the redundancies.

Exploitation

Based on the final calculated matrix of containments for all the CDA templates, we are able to refine the matrix of containment, in order to extract only needed templates to express the richness of tested CDA document. For many specifications, there are multiple document header templates (example, for C-CDA 2.1 there are CCD documents, Discharge Summary documents, etc). When analyzing CDA document richness, we are only interested on the specific used document header template, and all its containments. The other document header templates are not useful for the analysis of the document. To extract the matrix of sub-templates containment for a specific template, we need to execute this equation:

$$\Psi = diag\left(\sum_{i=0}^{n} e_k \Phi^i\right) x \Phi \quad (2)$$

Where

ϕ: the final containment matrix, without refinement

Ψ: describes the final containment matrix, refined for the use case of a specific template.

n: the level of CDA containment we are looking for; if we want to have the complete CDA templates containment architecture, we fix this attribute to the dimension of the matrix ϕ. However if we want for example to analyze only CDA L2, we set n = 1.

k: The index of the selected templateId in the list of rows described by the matrix ϕ

e_k: a linear Boolean vector containing only 0 and 1 values, 0 when the element in the row of the matrix ϕ is different than the selected template, 1 when the row template is equal to the selected template.

In our example, if we want to select all the templates and containment related to template A, we have:

- n: 8
- k: 1
- e_k = [1 0 0 0 0 0 0 0]

$e_k \Phi^i$ is a linear Boolean vector describing the templates included for the i[th] level of containment between templates, based on the selected CDA template.

From our example: $e_k \Phi^2$ = [1 0 0 0 0 0 0 0] describes the second level of containment (G and H).

$e_k \Phi^i$ is the linear Boolean vector describing all the templates included in all the level of containment, and related to the selected template. In our example,

$$\sum_{i=0}^{n} e_k \Phi^i = [1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]$$

This is interpreted as: in the tree related to template A, we can have A, D, E, F, G, and H.

$diag\left(\sum_{i=0}^{n} e_k \Phi^i\right)$ is the diagonal matrix used to extract useful templates information from the global containment matrix.

For our example

$$diag\left(\sum_{i=0}^{n} e_k \Phi^i\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The final refined matrix provides:

$$diag\left(\sum_{i=0}^{n} e_k \Phi^i\right) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

This matrix allows having only the needed templates containments relationship, and this allows identifying in the validated CDA document the list of present and missing templates for the specific document header templateId. The refined containment matrix allows constructing the tree of sub-templates based on the document header templateId.

From the analyzed CDA document, we extract a matrix containing the same dimensions and row description of the refined containment matrix, called matrix document description (MDD: Λ). When we have a zero in the matrix Ψ we set a zero on Λ. When we have one in the matrix Ψ, we extract the templateId described by this row index, and its related sub-template described by the column index. We look then in the CDA document if the described containment is implemented. Using Λ, we are able to evaluate graphically the list of present and missing templates in the analyzed CDA document. Λ and Ψ are used together to compute the CDA document richness scoring.

3.3 CDA Document Richness Scoring

The aim of scoring a CDA document is to provide a metric to compare CDA documents between themselves. The metric provided is a computation of the richness of templates containment. There are two ways to calculate CDA document richness scoring: basic richness scoring and weighted richness scoring.

Basic Richness Scoring

The calculated scoring is based on existing templates against possible templates.

$$R_s = \frac{D_i}{T_i} = \frac{Number\ present\ templates}{Number\ possible\ templates} \quad (3)$$

For each template, we calculate the tree of containment based on the matrix Ψ. Once we have the complete tree of templates, T_i is equal to the total number of present sub-templates for all levels.

$$T_i = \sum_{l=0}^{\infty} sum_t(l) \quad (4)$$

Where l is the level of containment and $sum_t(l)$ is the number of possible templates in the level l. T_i is called Template richness indicator.

$$D_i = \sum_{l=0}^{\infty} sum_d(l) \quad (5)$$

Where l is the level of containment and $sum_d(l)$ is the number of existing templates in the level l included in the CDA document tested. D_i is called Document richness indicator.

From our example, the Template richness indicator for the template A is 6 (we have 6 sub-templates in all the levels included).

D_i is calculated the same way as T_i; however the computation is based on the matrix of existing templates in the CDA document (Λ).

D_i describes the number of existing sub-templates in the CDA document checked, for all the sublevels of the selected template.

For example, if we are missing two sub templates related to the template A from our example (example E and G), we will have $D_i = 4$ and $R_s = \frac{4}{6} = 0.66$

The problem with this method of scoring computation is the fact that we do not take in consideration the level of containment. This method does not make difference if we are missing a template from a higher level or from a lower level. In CDA, the most important containments are those of the L2 templates (the section). The L3 templates are less important, and the other levels are lesser important. The scoring of the CDA document needs to take in consideration these levels of containments. Missing templates from the CDA L2 shall be more dangerous than missing templates from CDA L3. That's why we need a weighted computation of the richness scoring.

Weighted Richness Scoring

To weight the scoring of the richness of CDA document, we need to weight found templates in each level.

$$R_{ws} = \frac{D_{wi}}{T_{wi}}$$

Where

$$T_{wi} = \sum_{l=0}^{\infty} \frac{sum_t(l)}{f(l)} \quad (7)$$

And

$$D_{wi} = \sum_{l=0}^{\infty} \frac{sum_d(l)}{f(l)} \quad (8)$$

Where $f(l)$ is a function based on the level l.

The more $f(l)$ is exponential, the more R_{ws} is reflecting the containment for the first levels of templates and ignoring the other levels.

Examples of $f(l)$:

$$f(l) = l$$

$$f(l) = l!$$

$$f(l) = l!$$

Let's take the example explained in this paper and we score the richness of the templates included in the template A, when we are missing the template G.

$R_s = \frac{5}{6} = 0.83$ (Basic richness scoring computation)

When we score the richness with a weighted level based on $f(l) = l$

$T_{wi} = 5$ and $D_{wi} = 4.5 \Rightarrow R_{ws} = 0.9$

When we score the richness with a weighted level based on $f(l) = l^2$

$T_{wi} = 4.5$ and $D_{wi} = 4.25 \Rightarrow R_{ws} = 0.94$

As we can remark, the weighted scoring better reflect the fact that we are missing a template from the L3 and not from the L2. Making $f(l)$ exponential modify the scoring result by making the accent only on the first levels of templates containment.

It is up to the tests provider to define the strategy regarding the definition of $f(l)$.

For a CDA document template, T_{wi} is always a fixed value for a specific $f(l)$: this number allows describing the complexity of a CDA standard.

When $f(l) = 1$, $R_{ws} = R_s$.

3.4 Tools Richness Scoring

The richness scoring is about CDA documents; however certification programs are about applications and systems. The evaluation of tools shall be calculated against a bench of CDA documents, and the result of the scoring need to take into account the different trees of the provided test documents. The matrix of richness shall be then a summary of all provided templates, and their containments.

3.5 HL7 CDA R2 Validation Reliability

The automatic validation of CDA documents against validation tools is a way to estimate the correctness of the clinical information provided. However, even if a CDA document is valid, the reliability on the validation result depends on two other parameters:

- The reliability on the tool used for the validation; this reliability is based on the requirements coverage regarding the CDA specification used for the validation [14]
- The richness of the CDA documents provided

A low richness scoring of the CDA documents decreases the reliability on the content creator tool. The validation of the CDA documents cannot confirm if in the future the tool is able to create valid CDA documents, if the CDA documents validated do not cover all the possible kind of clinical information.

Reliability = f(Validation result, Requirements Coverage, Documents richness)

4 Implementation

The CDA richness computation was implemented as a part of Gazelle ObjectsChecker [15]. The input for the CDA richness module is the customer templates design coming from ART-DECOR [16]. This templates design is the formal XML description of the CDA specification, based on HL7 Templates Standard. And any tool able to provide this architecture of requirements may be an input for the CDA richness module. This module takes advantage of Gazelle ObjectsChecker for the parsing and for the information extraction from the custom templates design. The other input for this architecture is the CDA document to be scored. The output from the richness module is the matrix of templates containment. From the custom templates design, we generate the CDA validation tool using Gazelle ObjectsChecker, and then we validate the provided CDA document and generate the report of validation. This report contains the report for all checked rules, and also contains the list of found templates on the validated CDA document. The validation report and the Matrix of templates containment are used as input for the richness scoring (Figure 6).

5 Applications

5.1 HL7 CDA Specifications Richness Comparison

As application, we selected a bunch of standards in order to compare their richness in templates and in depth. The basic richness template indicator provides the number of possible sub-templates included in the parent template of the targeted specification. Weighted richness template indicator is the sum of all possible sub-templates weighted by their

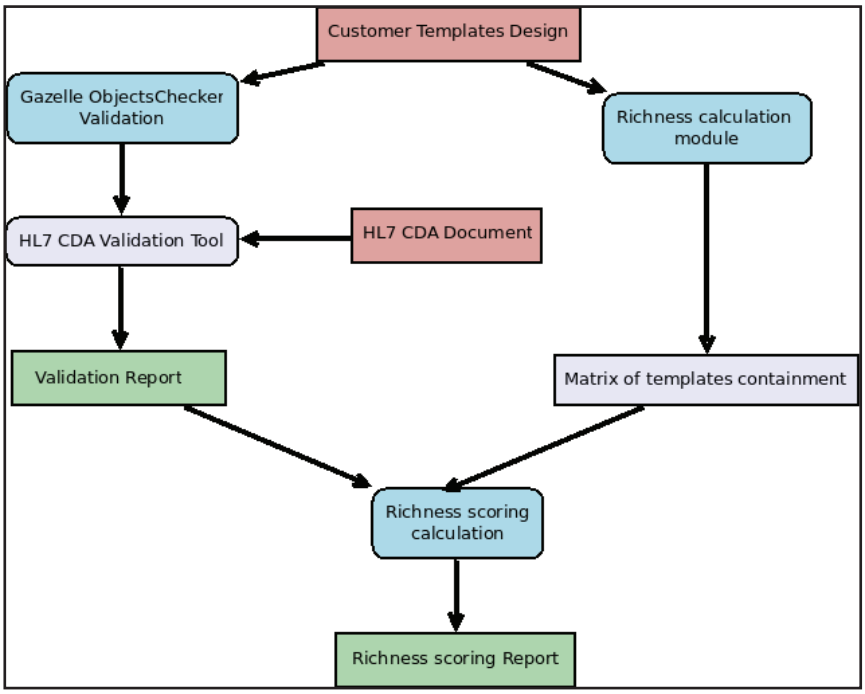


Figure 6: Implementation of richness scoring.

Table 1: CDA implementation guides comparison.

Standards	Basic Richness Template Indicator	Weighted Richness Template Indicator	Number Kind Templates Referenced	Depth
IHE Discharge Summary Specification	1271	52.91	63	10
IHE Immunization Content Specification	796	40.15	68	10
IHE PHR Extract Specification	926	60.86	84	10
IHE Referral Summary Document	862	45.6	66	10
epSOS-Patient Summary	1161	59.67	65	10
C-CDA 2.1 CCD	705	73.77	104	9
C-CDA 2.1 Discharge Summary	645	75.2	101	9
C-CDA 2.1 Diagnostic Imaging Report	42	17.03	17	6

levels ($f(l) = l!$). The depth describes the number of level of templates found in the specification (Table 1).

These metrics provide a big picture of each standard. The weighted template richness indicator provides an indicator for a comparison between the standards. Those with a high T_{wi} are more oriented for the tools processing, and their content is more oriented for CDA L3 content description. Standards with low T_{wi} are more flattened, and then more oriented to human readability. The depth of the standard provides also the same kind of indication. The basic template richness indicator provides a description of the complexity of the architecture of the standard; it describes the number of possible containments between templates. A global remark, nearly all the standards have more than 600 possible containments. This is a huge number and it gives an idea of the complexity of implementing a CDA content creator tool for such specifications.

As example, let's compare the IHE Discharge summary, and the C-CDA Discharge summary characteristics. The number of possible templates for C-CDA DS is bigger than the number of possible templates for IHE-DS (101 against 63). Although the number of possible containments in IHE DS is bigger than C-CDA DS, the weighted comparison describes the C-CDA DS as more flattened than IHE-DS; this is confirmed by the Depth of both standards (10 against 9 levels). We can interpret this by the fact that C-CDA DS is more human readable, with a better granularity of the collected information, and the IHE-DS is more tools interpretation dedicated. Also there are less redundancy of the sub-templates in the tree of containment of C-CDA DS (1271/63 > 645/101); this is interpreted by the fact that C-CDA DS has included more specialized templates than IHE DS, and we have less interpretation to do in C-CDA DS regarding the context of use of a specific sub-template.

5.2 CDA Documents Richness Comparison

In this paragraph, we took a list of CDA documents already validated by Gazelle ObjectsChecker under the database of EVSClient tool [17], and we calculated the average of richness scoring for some available CDA validators (Table 2).

We can remark that all the validated CDA documents are far from covering 100% of the templates containments defined in the CDA implementation guides. Some of the tested documents were used in IHE Connectathon [18], which is not harmful for IHE testing process, but this is dangerous for a certification program; the certification authorities SHALL check the richness of the provided documents [2].

Table 2: Richness scoring analysis of CDA documents.

Standards	Average of Weighted Richness Scoring	Number of tested documents
ePSOS ePrescription	0,229	2447
ePSOS eDispensation	0,201	1322
ePSOS ePatient Summary	0,178	11530
IHE Immunization Content	0,226	276
IHE Referral Summary	0,248	1049
IHE Discharge Summary	0,310	526
IHE PHR Extract	0,139	827
C-CDA Referal Note	0,280	129
C-CDA CCD	0,201	1908

5.3 Interpretation of Richness Scoring

A certification program may define a strategy regarding the accepted CDA documents during a testing process. Based on the C-CDA companion guide, SITE [19] defined 5 score grades regarding the scoring of C-CDA documents [20]: A+, A-, B+, B-, C and D; each of those values describes the quality of the CDA provided. Based on the study of a set of CDA documents related to the implementation guide tested, we can establish a score grade distribution. For example, we take the C-CDA CCD documents tested [20]. Here is the normal distribution of the richness scoring for the 2000 documents tested, coming from EVSClient database (Figure 7).

The distribution of grade is based on the average and the variance of the richness scoring. The average of richness scoring for the C-CDA tested is 0.201, with a variance of 0.07. Here is a possible distribution of grades:

D: $R_{ws} < \mu - 2\sigma$: very low richness

C: $\mu - 2\sigma < R_{ws} < \quad -\sigma$: low richness

B-: $\mu - \sigma < R_{ws} < \mu$: lower than the average

B+: $\mu < R_{ws} < \mu + \sigma$: higher than the average

A-: $\mu + \sigma < R_{ws} < \mu + 2\sigma$: high richness

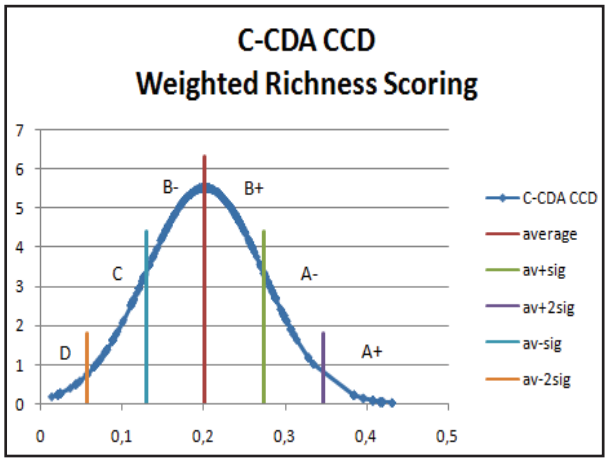


Figure 7: CDA Richness grading.

A+: $R_{ws} > \mu + 2\sigma$: excellent richness

It is up to the testing provider to define the targeted richness grade for the tested CDA tools.

5.4 Combined richness and validation process

IHE-Europe has developed a combined tool for validation and richness analysis of CDA documents. The validation is performed using Gazelle ObjectsChecker. Here is an example of the validation of a C-CDA CCD document, and the GUI provided as validation and richness analysis (Figure 8).

This tree describes all the possible templates that may appear in a CCD document. The green templates in the schema express the fact that such template was found and was valid. The red templates express that such template was found with errors. White templates express that they are missing from the validated CDA document. This output allows having a visible validation and richness report regarding the templates provided.

6 Matrices Glossary

Table 3: Matrices glossary.

Matrix	Name	Description
A	Matrix of inheritance	Describes the inheritance between different templates
B	Matrix of containment	Describes the containment between the different templates as described in the customer templates design
$C = \sum_{n=1}^{\infty} B^n$	Complete matrix of inheritance	Describes all the inheritance between the CDA templates
$\Phi = (CA + A) \& \overline{AC}$	Flattened matrix of containment	Describe the final matrix of templates containment
$\Psi = diag\left(\sum_{i=0}^{n} e_k \Phi^i\right) x\Phi$	Refined matrix of containment	This matrix is calculated based on Φ, to describes only the templates related to a specific CDA template
Λ	Matrix document description	A matrix to describe the templates containment related to a provided CDA document
$f(l)$	Level heightening	A formula to weight the levels of the CDA templates containment
$T_{wi} = \sum_{l=0}^{\infty} \frac{sum_t(l)}{f(l)}$	Template richness weighted indicator	The richness indicator of possible templates related to a specific root template
$D_{wi} = \sum_{l=0}^{\infty} \frac{sum_d(l)}{f(l)}$	Document richness weighted indicator	The richness indicator of present sub-templates for a CDA document
$R_{ws} = \frac{D_{wi}}{T_{wi}}$	Weighted richness scoring	The weighted scoring computation of a specific template under a CDA document

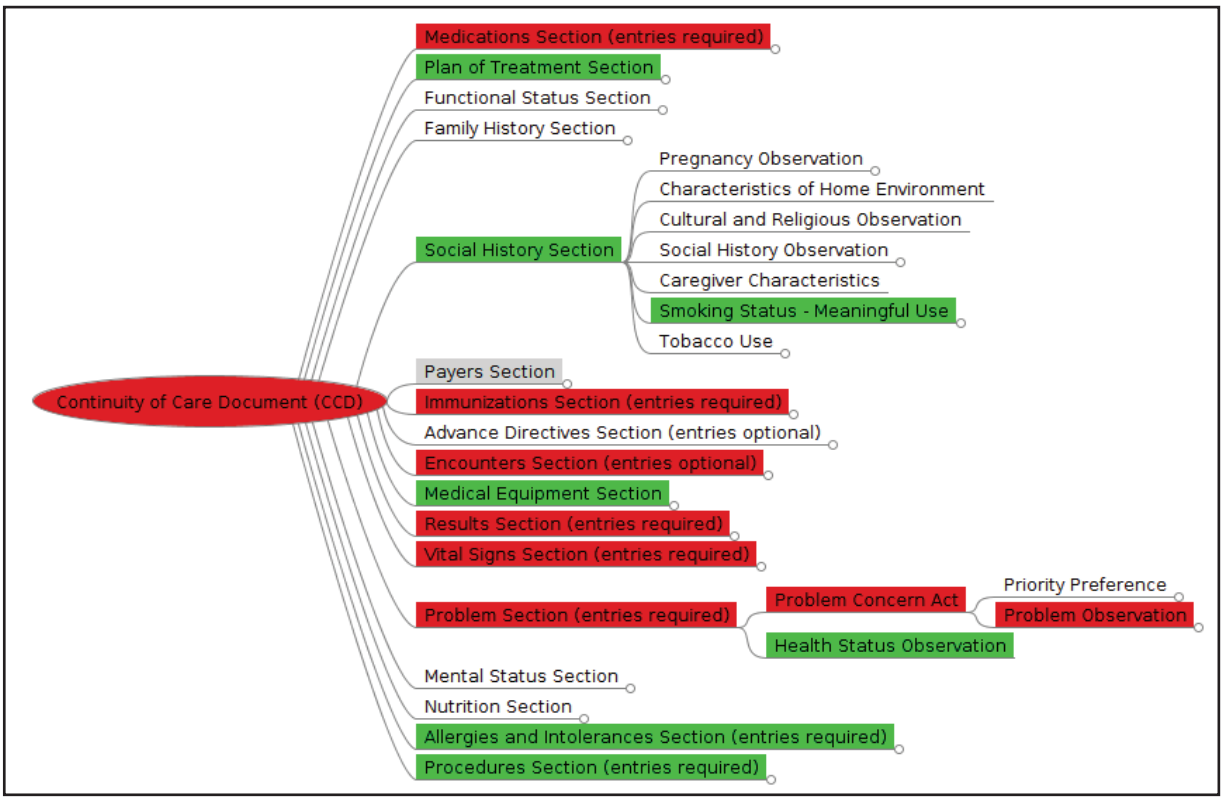


Figure 8: Combined validation and richness analysis.

7 Conclusion

The richness of CDA documents can be computed based on many criteria. In this paper, we expressed it using the templates containments. Tests providers need to have an indicator of richness to confirm their reliability on the tested applications. Richness study provides reports on areas covered by the testing process.

In this paper, we resolved the main difficulty for scoring the CDA templates, which is the extraction of the complete list of templates containment. Also we defined a methodology to compute the richness of documents tested; this method takes in consideration the specificity of CDA standard, and its architectural levels. Grading of CDA documents is also possible, based on existing test data.

The experimentation of this methodology on some HL7 and IHE specifications confirmed the complexity of these standards, and provided us a way to compare their architectures. The computation of the richness of thousands of CDA documents coming from test tools proved they are in average poor on templates and clinical information. The tests providers need to improve their testing process by taking in consideration the richness of the test data provided.

Many applications of this paper may follow this study, like in conformity assessment accredited testing of IHE CDA documents, and where reporting the coverage of testing area is mandatory during the test procedure.

References

[1] Health Level Seven Internatinoal, Inc. HL7 Clinical Document Architecture, Release 2.0

[2] Internatinoal Organization for Standardization. ISO 17025 General requirements for the competence of testing and calibration laboratories. Geneva: ISO; 2005

[3] Brull R. CDA Levels of Interoperability. Healthcare IT News. June 03, 2011.

[4] D'Amore JD, Mandel JC, Kreda DA, Swain A, Koromia GA, Sundareswaran S. et al. Are Meaningful Use Stage 2 certified EHRs ready for interoperability? Findings from the SMART C-CDA Collaborative. J Am Med Inform Assoc. 2014; 21: 1060-1068.

[5] Boufahja A. Projectathon results and findings on semantic related aspects, eHDSI Project, June 2017

[6] Health Level Seven Internatoinal, Inc. HL7 Implementation Guide: S&I Framework Transitions of Care Companion Guide to Consolidated-CDA for Meaningful Use Stage 2, Release 1 – US Realm. Ann Arbor: HL7; 2014.

[7] SMART. C-CDA Smart Scorecard, http://ccda-scorecard.smartplatforms.org/

[8] eHealth DSI – ePrescription and Patient Summary, Test Framework, version 1.1.0, 01/06/2017

[9] Health Level Seven International, Inc. HL7 Templates Standard: Specification and Use of Reusable Information Constraint Templates, Release 1. Ann Arbor: HL7.

[10] Boone K. A Template Meta-model, http://motorcycleguy.blogspot.fr/2011/04/template-meta-model.html

[11] eHealth DSI CDA Implementation Guides, June 2017.

[12] Benson T. Principles of Health Interoperability HL7 and SNOMED, Section 9.3. Berlin: Springer; 2010.

[13] Horn RA, Johnson CR. Matrix Analysis, 2nd Ed. Cambridge: University Printing House; 2013.

[14] Boufahja A, Poiseau E, Thomazon G, Bergé A. Model-based Analysis of HL7 CDA R2 Conformance and Requirements Coverage. EJBI 2015; 2: en41-en50.

[15] Boufahja A, Heitmann K, Poiseau E. Model-based Validation of HL7 CDA R2 Documents and Implementation Guides Using Gazelle ObjectsChecker and ART-DECOR. EJBI 2016; 1: en62-en69.

[16] Heitmann K. ARTDECOR Basics and User Perspective. Tutorial, IHIC 2015, Prague; 2015.

[17] Gazelle project. EVSClient tool, https://gazelle.ihe.net/EVSClient/

[18] IHE-Europe. The IHE Connectathon: what it is and how it is done, November 2015

[19] The Standards Testing & Implementation Environment (SITE), C-CDA scorecard, https://sitenv.org/scorecard/

[20] Health Level Seven Internatinal, Inc. Consolidated CDA Templates for Clinical Notes (US Realm), DSTU 2.1.

17th International HL7 Interoperability Conference (IHIC 2017)

Topics

- The advancement of interoperability
- Harmonization of interoperability standards and specifications among different SDOs
- Terminology and ontology challenge of interoperability
- Concepts and frameworks for Smart Interoperability Infrastructure Services
- Local, regional or national Electronic Health Records solutions
- Business Intelligence and Clinical Decision Support
- Security, safety and privacy
- Specification and implementation tools
- FHIR and CDA – controversy, coexistence, or synergy?
- "Show me your CDA" – CDA implementations at all levels

The Use of HL7 Clinical Document Architecture Schema to Define a Data Warehouse Dimensional Model for Secondary Purposes

Fabrizio Pecoraro, Daniela Luzi, Fabrizio L Ricci

Institute for Research on Population and Social Policies, National Research Council, Rome, Italy

Abstract

This paper proposes a semi-automatic approach to extract information stored in a HL7 Clinical Document Architecture (CDA) and transform them to be loaded in a Data Warehouse for secondary purposes. It represents a suitable solution to facilitate the design and implementation of Extract, Transform and Load (ETL) tools that are considered the most time-consuming step of the data warehouse development process. The implementation of this framework is also proposed adopting the XSLT style sheet language that converts an original CDA XML-based document to an output XML document that can be easily loaded in the Data Warehouse. A case study is also provided to demonstrate the feasibility of the approach proposed.

Keywords

Data warehousing; Dimensional model; HL7 CDA; Extensible Stylesheet Language Transformation (XSLT); XML

Correspondence to:

Fabrizio Pecoraro

Institute for Research on Population and Social Policies, National Research Council Italy, Via Palestro, 32 – 00185 – Rome, Italy

E-mail: f.pecoraro@irpps.cnr.it

EJBI 2017; 13(1):85-95

received: May 31, 2017

accepted: July 19, 2017

published: October 10, 2017

1 Introduction

In the healthcare setting there is a growing attention on secondary uses of clinical data defined as "non-direct care use of personal health information" [1]. The use of clinical data for secondary purposes provides important sources to support decision-making in different domains, such as patient safety, healthcare quality assessment, clinical and translational research including clinical trials, comparative analysis of therapy pathways and best practices application [2]. To reach this aim a comprehensive analysis is required that has to integrate clinical and administrative information provided by heterogeneous information systems often developed using different technologies, for different specialties and purposes and by different organizations [3, 4]. This makes it necessary to implement specific Extract, Transform and Load (ETL) procedures devoted to convert data from source operational systems in a common data model optimized for data analysis purposes.

In healthcare different standards have been developed to facilitate system interoperability and under the perspective of data models, HL7 [5] surely represents one of the main candidates for the integration and exchange of information [6] generally focused on patient's care delivery. One of the widely adopted HL7 standard is the Clinical Document Architecture (CDA) [7] that specifies the encoding, structure and semantics of clinical documents using a XML based mark-up language. Recently, many initiatives have analyzed the importance of designing and implementing a data warehouse starting from XML documents considering the continual growth of representing data using XML documents in different domains [8, 9, 10]. In our vision the main aim of HL7 standards and in particular the CDA [7] can be extended to define a common schema able to represent this information in enterprise data warehouses to be used for secondary purposes.

Aim of this paper is to define a semi-automatic approach to extract information from XML document structured using the CDA standard and transform them to be included in a data warehouse schema. To perform this task an EXtensible Stylesheet Language Transformation (XSLT) document [11] is defined to provide an output XML document that can be easily stored in the data warehouse logical schema. This approach is based on a conceptual framework already described in details in a previous publication using a first-order logic [12]. Next paragraph describes the main steps of this framework that maps the CDA components with the conceptual model concepts. The third paragraph describes how the conceptual framework has been implemented highlighting the generation of the XSLT document. After that, to demonstrate the feasibility of this approach a case study

providing an example of the transformation task is proposed. Final remarks are given in the conclusion paragraph.

This study is part of the Smart Health 2.0 national project that aims to develop a regional healthcare infrastructure based on HL7 standards. It also intends to explore the use of Electronic Health Record (EHR) for secondary purposes in a clinical governance framework to assess the quality of care from the structural, organizational, financial and professional points of view [13].

2 Conceptual Mapping from CDA Schema to Dimensional Model

2.1 Data Warehouse Model

The data warehouse conceptual modelling can be formalized using the dimensional model as depicted in Figure 1.

The core of this schema is the Fact table that describes the measurements of the performance of a business process using qualitative and/or quantitative attributes called measures. It is surrounded by independent Dimensions each one modelled using an independent denormalized table or a normalized hierarchy. In the first case the model is called star schema while in the second representation is called snowflake schema. Fact along with its relevant measures as well as Dimensions represent the concepts of the dimensional model to be mapped with the CDA elements described in the following paragraph.

2.2 CDA Model

CDA Release 2 Level 3 records clinical observations and services in a mark-up structured standard document based on the six backbone classes of the HL7 Reference Information Model (RIM) [5]: Act, ActRelationship, Participation, Entity, Role and RoleLink. As highlighted in Figure 2, these classes as well as their relationships are used to define two main components of the CDA document [12]:

CDA Backbone defined by the Act specializations and their relationships. For instance, the Act ClinicalDocument

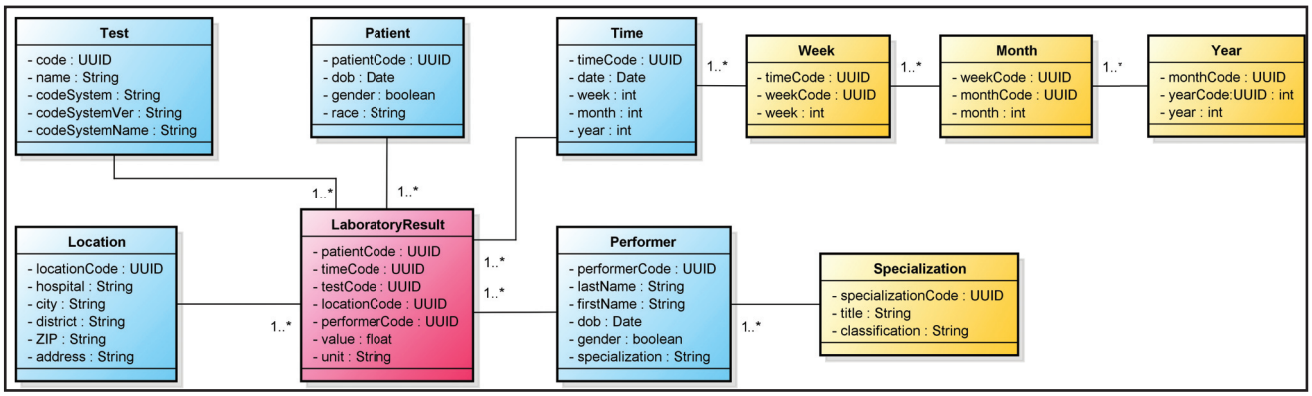


Figure 1: Example of a dimensional representing a snowflake schema composed by a Fact LaboratoryResult related to five dimensions: Test, Patient, Time, Location and Performer.

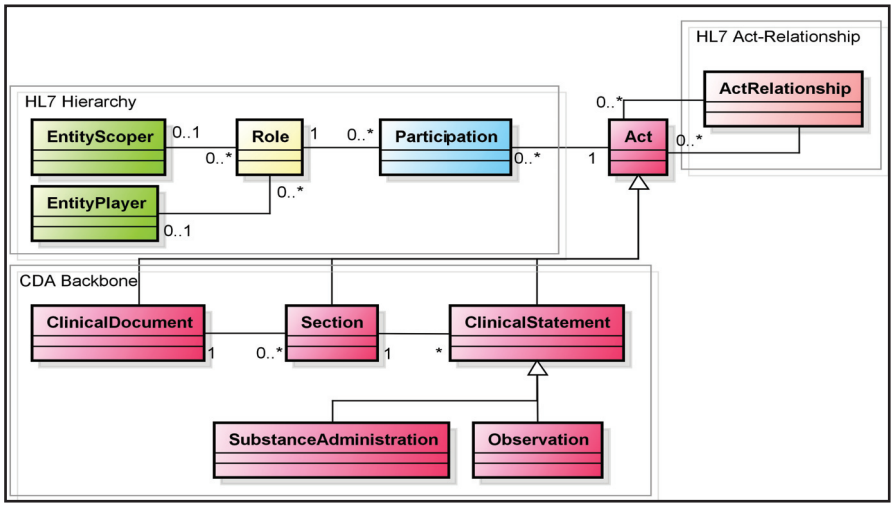


Figure 2: High level class diagram of the CDA schema modelled using the HL7 RIM core classes. The two main components of the CDA are also highlighted: 1) HL7 Hierarchy composed by the triple <Participation, Role, Entity> related to the Act class; 2) CDA Backbone defined by Act specializations and their relationships.

that represents the entry point (i.e. root) of the CDA document is composed by a set of Sections each one collecting one or more events modelled using the Act classes of the ClinicalStatement choice, such as Observation, SubstanceAdministration.

HL7 Hierarchy that describes subjects and objects involved in the process as well as the role played by them within the action using the n-ple <Participation, Role, Entity Player, Entity Scoper> [14]. For instance, the hierarchy <recordTarget, patientRole, Patient, Organization> represents the patient involved in the events documented in the CDA. Each HL7 Hierarchy is related with a specific Act of the CDA Backbone that describes the action performed or scheduled. A portion of the CDA schema highlighting three HL7 hierarchies (i.e. recordTarget, performer and participant) and the CDA Backbone is shown in Figure 3 using the HL7 message information model notation.

The described HL7 Hierarchy and CDA Backbone as well as the relevant complex attributes represent the components of the CDA schema to be mapped with the dimensional model concepts introduced in the previous paragraph.

2.3 Conceptual Framework

In this paragraph we describe a conceptual framework to map the CDA components (HL7 Hierarchy, CDA Backbone) with the dimensional model concepts (Fact, Dimension). To perform this mapping it is necessary that the designer have already identified the business process to be modelled as well as the level of detail to be captured (i.e. what an individual row of the Fact table represents). This is an important aspect given that the granularity of the dimensional model influences the identification of both the Dimensions to be modelled and the attributes and measures to be captured. This decision has to take also into account the granularity of data contained in the CDA document that generally captures atomic data, such as value of vital signs observed during a laboratory test.

2.3.1 Identify the Fact

As already mentioned a Fact describes the relevant event to be analysed trough qualitative and quantitative measures that represent the performance of the business process and that could be analysed using statistical methods. In the CDA these information are collected in specific attributes of the stereotype Act of the RIM that represents "measurement of healthcare business processes". For this reason in our approach the Acts that define the CDA Backbone can be considered as suitable candidates to identify the Fact of the dimensional model depending on the purpose of the analysis to be carried out and on the indicators to be developed. Examples of Act that can describe relate actions and events that constitute health care services are reported in Table 1 where examples of business processes and measures are reported.

Once the Fact has been determined, its attributes are analysed to define measures that represent a qualitative or

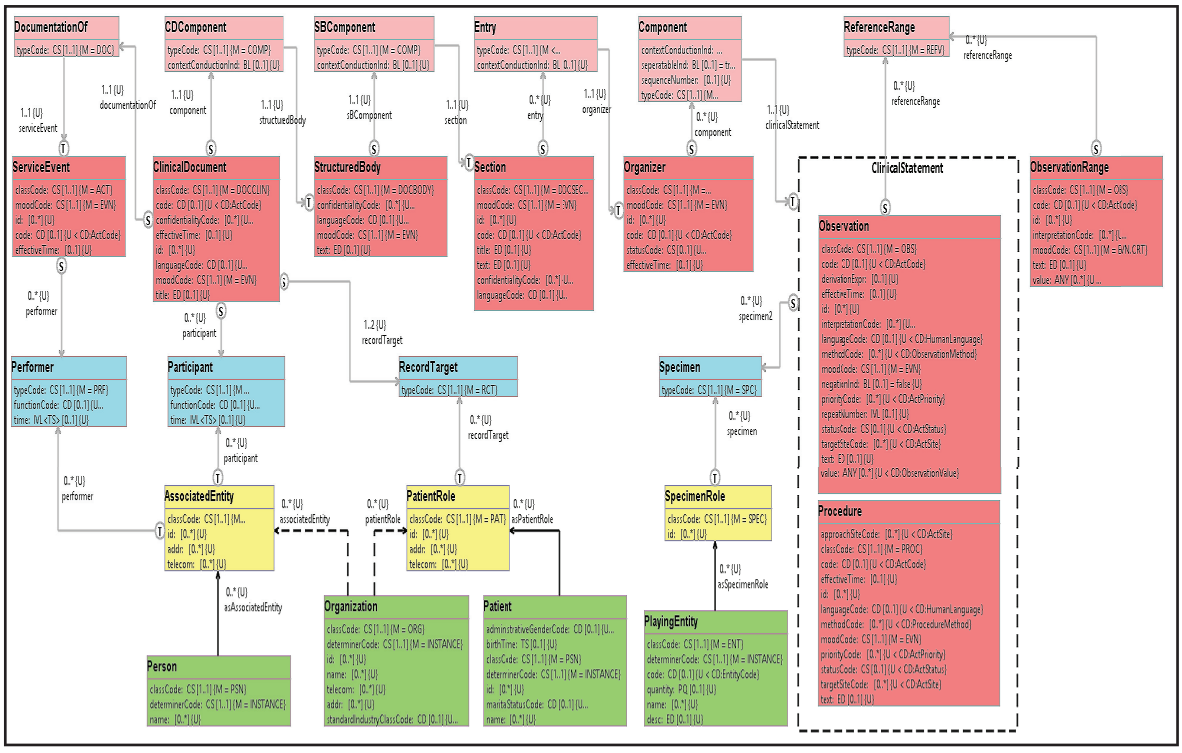


Figure 3: Portion of the of the CDA message model showing the CDA backbone and three HL7 Hierarchies: Performer, Participant and RecordTarget.

Table 1: Example of Act classes that can be used to represent a Fact table of the dimensional model.

CDA class	Description	Example of processes	Measures
Act	General event that is being done, has been done, can be done, or is intended or requested to be done.	To be used when the other more specific classes aren't appropriate.	N/A
Encounter	An interaction between a patient and healthcare participant(s) to provide service(s) or assessing the health status of a patient.	Specialist and MMG visits	lengthOfStayQuantity (quantity of time when the subject is expected to be or was resident at a facility as part of an encounter)
Observation	action performed in order to determine an answer or a result value	vital signs, clinical results in general and also diagnoses, findings, symptoms,	value (data determined by the observation) interpretationCode (a qualitative interpretation of the observation)
Procedure	An event whose immediate and primary outcome (post-condition) is the alteration of the subject physical condition	conservative procedures such as reduction of a luxated join, including physiotherapy such as chiropractic treatment	N/A
SubstanceAdministration	The act of introducing or otherwise applying a substance to the subject.	Chemotherapy protocol; Drug prescription; Vaccination record	doseQuantity (amount of the therapeutic agent), rateQuantity (the speed with which the substance is dispensed)

quantitative evaluation of the business process. For instance, the Act Observation comprises two measures described by the attributes value and interpretationCode. They represent, respectively, a quantitative and qualitative measure of the event observed. In the RIM numerical information are collected in Act class attributes modelled with quantity (i.e. QTY) or physical quantity data type (i.e. PQ), whereas qualitative analysis are specified using coded data types (e.g. CV, CE, CD).

2.3.2 Identify the Dimensions

In this paper dimensions are determined based on the Zachman framework [15] that provides a systematic information representation starting from the following questions related with the investigated event: who (persons), what (the fact), when (the time), where (the place), why (the reason) and how (the manner). To identify suitable candidates to derive dimensions we start analysing the two main structural components of the CDA document related with the Fact class: 1) Acts that captures the meaning and purpose of each association with the main event as well as additional actions to determine, for instance, why the event has been performed or the criteria used to evaluate the event outcome; 2) HL7 Hierarchy that describes the functions of subjects and objects involved in a specific process, identifying for instance, who performed it (i.e. performer), for whom it was done (i.e. subject), where it was done (i.e. location). This information is captured through the attribute typeCode of the Participation class that specifies its meaning and purpose using a controlled vocabulary defined by HL7.

Table 2 summarizes examples of the different components of the CDA that can be used to identify a dimension of the schema, reporting the type and the name of the component as well as its description and the related Act class of the backbone.

Moreover, there are attributes of the Fact that can be specifically used to define a degenerate dimension, that is not modelled using its own table. A generic Act of the RIM contains several attributes that can be mapped in a degenerate dimension such as, code that classifies the particular kind of Act and statusCode that specifies state of the Act (e.g. active, cancelled). Another important attribute is the effectiveTime that describes time/date when the event took place. This approach of representing dimensions as keys of the Fact table often occurs when the dimensional model captures atomic information with a high-level of granularity transaction.

2.3.3 Refinement of the Dimensional Model

The design of a dimensional model based on the CDA elements results in a high-level normalized data model that is typically adopted in transactional database where an high volume of transactions (insert, update, delete) is performed. Conversely, in a data warehouse environment a highly normalized schema may create inefficiencies in the retrieval as well as in the aggregation of data due to the necessity of executing a large number of joins, which greatly increases response times [16]. Denormalizing relations reduces the number of physical tables that need to be accessed to retrieve the desired data by reducing the number of joins needed to derive a query [17]. For this reason the star schema is typically adopted to model data in analytical databases where a low volume of transactions (insert, update, delete) is performed with complex queries to be executed.

The denormalization is mainly applied to the HL7 Hierarchies and is performed by collapsing the attributes

Table 2: Example of suitable dimensions derived from the CDA components.

HL7 component	Name	Example
	Clinical Document	
HL7 Hierarchy	recordTarget	Patient involved in the event
	performer	Physician/Practitioner that carried out the event
	responsibleParty	Participant with legal responsibility
	location	Healthcare facility where the event occurred
	participant	Other involved and not mentioned participants
CDA Backbone	ServiceEvent	The main event being documented
	EncompassingEncounter	Primary encounter documented
	Section	
HL7 Hierarchy	subject	Target of the entries recorded in the document
	Clinical Statement	
HL7 Hierarchy	performer	Physician/Practitioner that carried out the event
	specimen	Part of entity typically the subject target of the observation
	participant	Other involved and not mentioned participants
	Observation	
CDA Backbone	ObservationRange	Range of values for a particular observation
	SubstanceAdministration	
HL7 Hierarchy	consumable	Substance consumed during the administration.

of the classes Entities and Role in the Participation class. However, healthcare business processes can require the adoption of many-to-many relationships to represent multiple records of a specific dimension associated with the Fact table. For instance, when different practitioners deliver care to an individual over different distinct time intervals or when a specialist visit is performed due to multiple diagnosis. In these cases the hierarchy cannot be fully denormalized and a bridge class should be used to model the many-to-many relationship between the Fact and the hierarchy [18]. An example of the application of the two denormalization methodologies is depicted in Figure 4 where the attributes of PatientRole, Patient and Organization are all collapsed in the recordTarget Participation class and the Performer is used as a bridge table to map the many-to-many relationship between the Act Observation and the healthcare providers involved in the provision of a service.

Another important step to be performed to refine the dimensional model is to resolve complex data types. In fact, several attributes of the CDA are coded using a complex data type that consists in a set of fields used to describe the value along with its properties. For instance, the attribute code of the class Observation is coded using the Concept Descriptor (CD) data type that contains eight attributes to model the code of the particular kind of Observation carried out as well as the information about the coding system used to represent it. A possible solution to represent a complex data type is to store each property in a single column of the relevant table excluding properties that are not needed for the business process analysis. For instance a CD can be mapped using only two attributes: code and codeSystem to store the code of the event occurred and the system used to represent it. Moreover, different attributes of the RIM assume multiple values, such as the interpretationCode that specifies a set of rough qualitative interpretation of an Observation based on a HL7 nomenclature (e.g. "is decreased", "is below alert threshold", "is moderately susceptible"). These attributes can be modelled either creating a separate table to store each instance or capturing only a single value, such as the first reported in the document.

3 Conceptual Framework Implementation

The workflow to transform and load data stored in a CDA document in the data warehouse is shown in Figure 5 highlighting two main sub-processes.

In the first part of the conceptual framework XSLT document is created taking into account the node chosen to represent the Fact of the dimensional model. Moreover, the relevant CDA schema is considered to identify RIM stereotype of each element as well as the cardinality of each relationship, while the data type schema specifies the cardinality and the type of data of each attribute of a specific node. This task is performed by the XSLT definition engine that is further described in the next paragraph.

In the second part of the workflow the XSLT document is used to process a CDA document represented using the XML format in order to produce an output XML document that can be further managed to be mapped into a relational, object-relational or XML-native database. In this perspective, different XML data warehouse architectures have been proposed in the literature to represent complex data as XML documents, such as XCube [19], X-Warehousing [20], XML-OLAP [21] to be physically integrated into an Operational

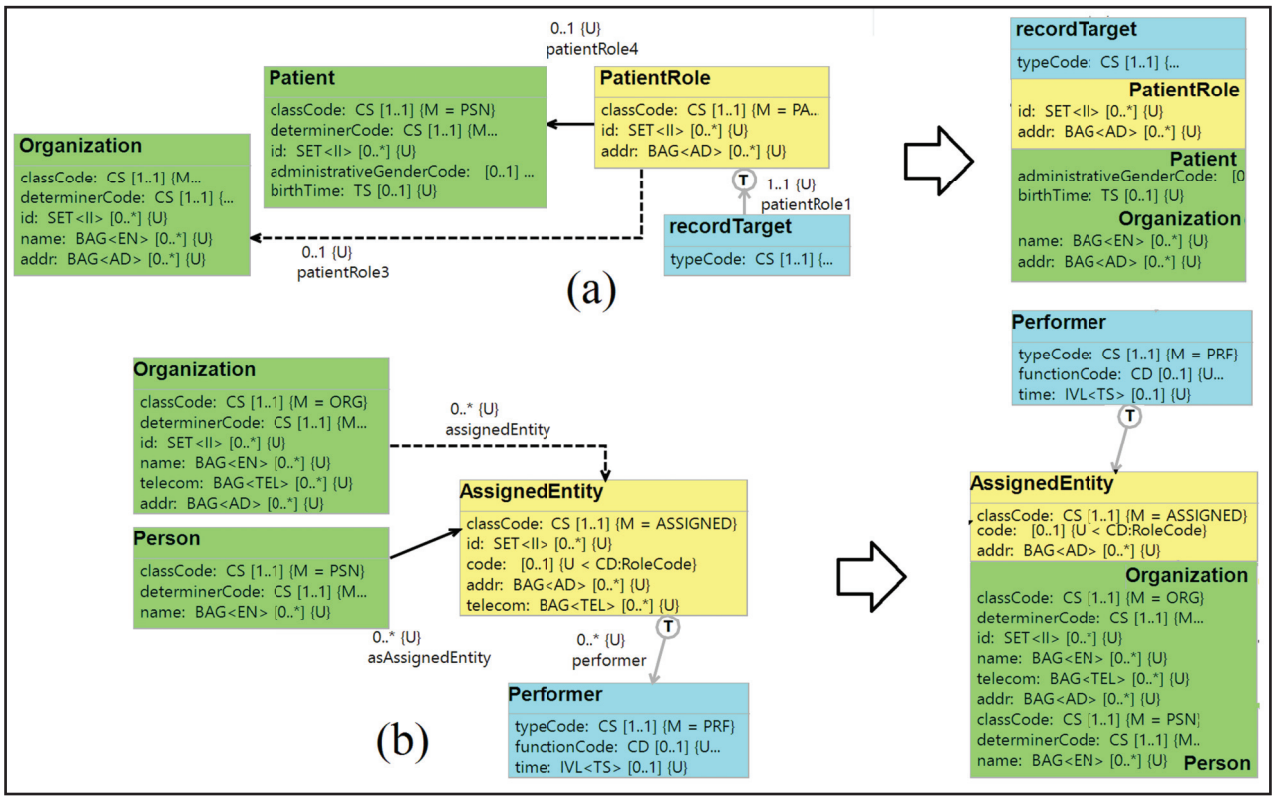


Figure 4: Examples of denormalization of two HL7 Hierarchies: a) recordTarget where attributes of Entities and Role are collapsed in the Participation class; b) performer where attributes of the Entities are collapsed in the Role assignedEntity and the Participation models a bridge to represent a many-to-many relationship with the Fact.

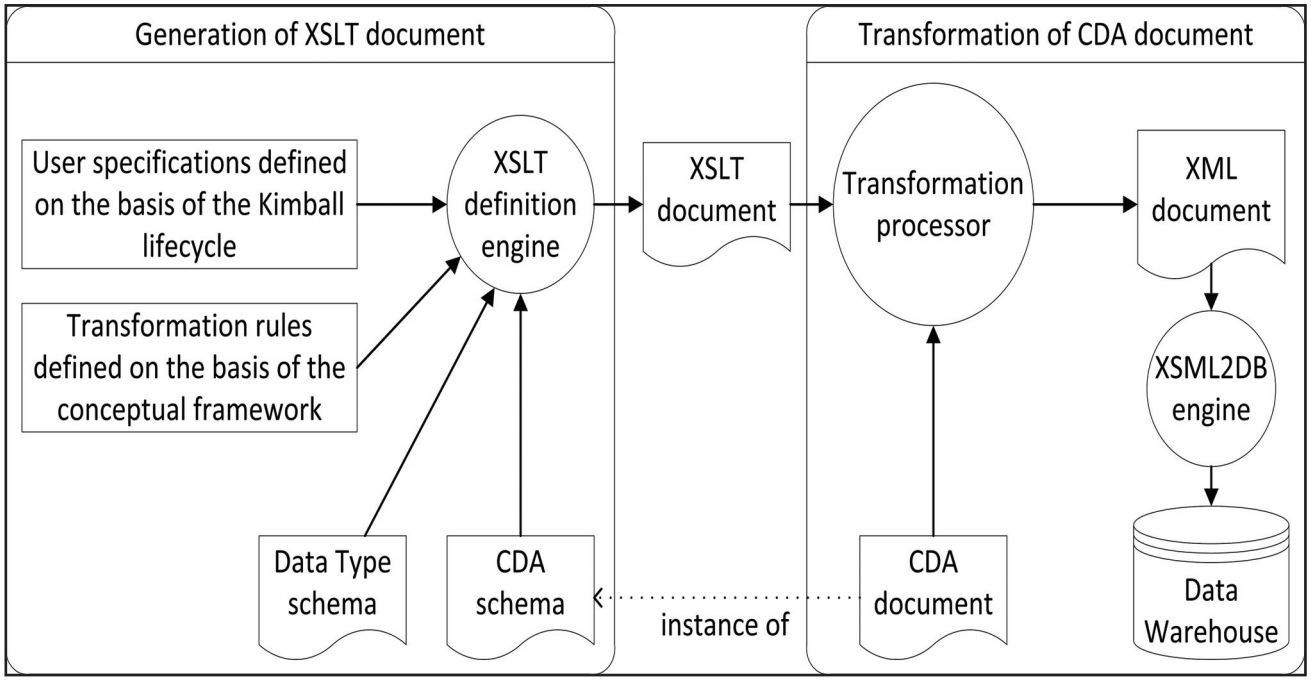


Figure 5: Transformation process to load a CDA document in a Data Warehouse.

Data Store (ODS) and further analyzed using statistical and business intelligence methodologies. These representations converge toward a unified model that differ in the number of XML documents used to store facts and dimensions [22]. In this paper transformed XML documents are organized on the basis of X-Warehousing architecture where each XML embed the facts stored in the original CDA document as well as their related dimensions. This transformation is performed by a XSLT processor, such as the Open Source SAXON XSLT engine developed by Saxonica Limited.

Note that, to comply with the privacy regulations the original CDA document must be anonymized. However, this activity has not been discussed in the paper given that it has to be applied to the CDA before applying the proposed conceptual framework.

3.1 Generation of the XSLT Document

Figure 6 reports the four main components (i.e. templates) of the XSLT document.

Each template that composes the XSLT document is identified by a specific pool using the BPMN notation. It highlights the different activities to be executed to transform a CDA structured document in a XML document.
In particular:

1. Main. As highlighted in Figure 7 it finds all the nodes that match with the class chosen by the data warehouse designer to represent the Fact table of the dimensional model (e.g. Observation). Starting from each node it navigates the XML document in both directions: each ancestor is explored by the Examine Ancestor Node template, while each child is analyzed by the Examine Node template.
2. Examine Ancestor Node. It includes the node passed as input in the transformed document considering its resolved attributes. Moreover, each child is analyzed by the Examine Node template.
3. Examine Node. It checks if the stereotype of the node received as input is a Participation. In this case the node is passed to the Denormalize Hierarchy, otherwise it is included in the output document along with its resolved attributes. Moreover, each child is recursively analyzed by this template to be included in the output document. Once all children have been analyzed the tag of the relevant node is closed.
4. Denormalize Hierarchy. As shown in Figure 8 starting from a participation node the 4-ple <Participation, Role, Entity Player, Entity Scoper > is analyzed and a denormalized node is reported taking into account the multiplicity of the relationship between the participation and the act class. If the multiplicity is 1-to-1 the complex attributes of role and entity nodes are resolved by the Resolve Data Type function and collapsed in the output schema as children of the participation node using the function Collapse

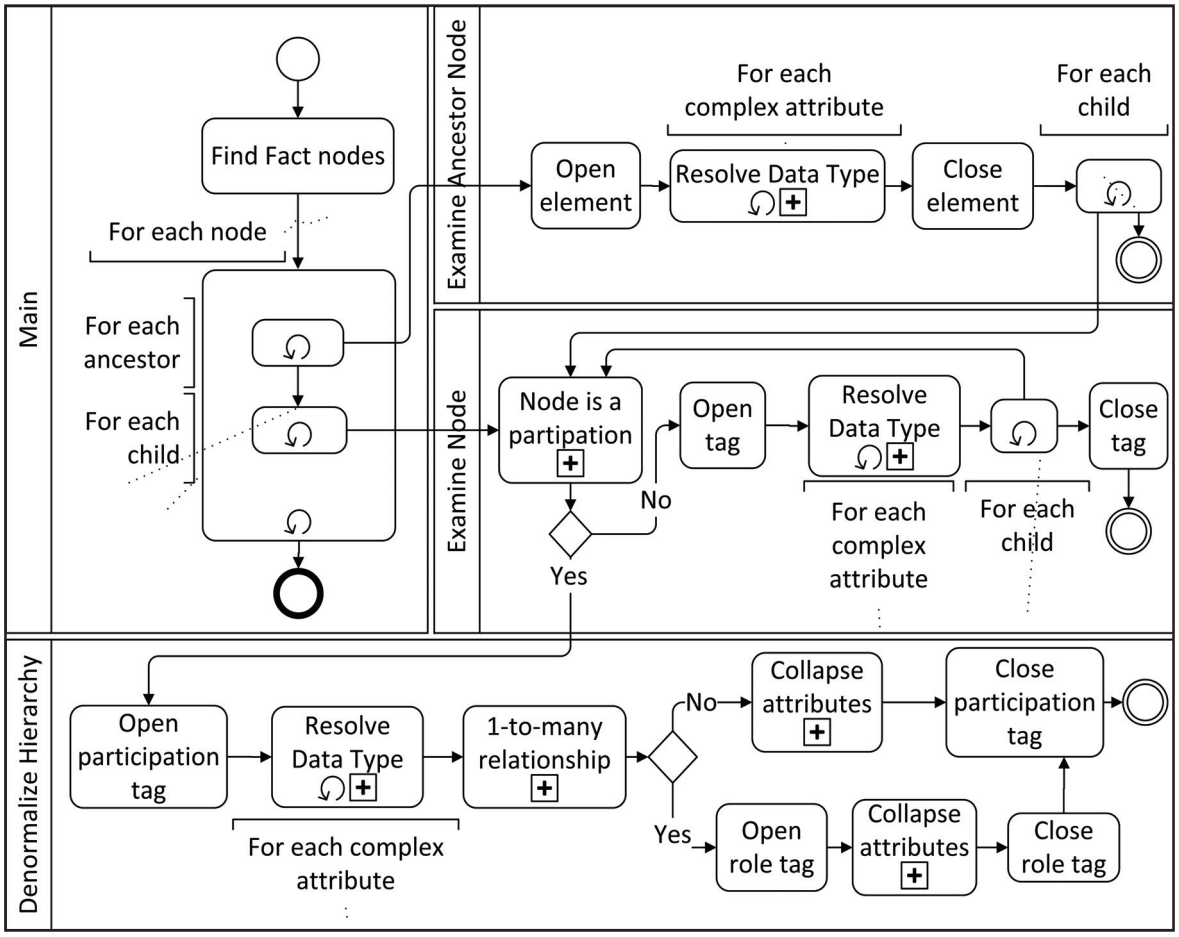


Figure 6: Business process to generate the XSLT document.

```
<xsl:template match="//observation">
    <xsl:element name="{name(.)}">
        <!-- examines the children of the fact node -->
        <xsl:for-each select="*">
            <xsl:call-template name="examineNode">
                <xsl:with-param name="node" select="." />
            </xsl:call-template>
        </xsl:for-each>
        <!-- examines all the ancestor of the fact node -->
        <xsl:variable name="ancList" select="ancestor-or-self::*"
        <xsl:for-each select="$ancList">
            <xsl:if test="parent::*">
                <xsl:call-template name="examineAncestorNode">
                    <xsl:with-param name="node" select="../." />
                    <xsl:with-param name="toRemove" select="name(.)" />
                </xsl:call-template>
            </xsl:if>
        </xsl:for-each>
    </xsl:element>
</xsl:template>
```

Figure 7: Portion of the XSLT document highlighting the Main template.

```
<xsl:template name="denormalizeHierarchy">
    <xsl:param name="node" />
    <xsl:param name="multiplicity" />
    <xsl:element name="{name($node)}">
        <xsl:for-each select="@*">
            <xsl:attribute name="{name(.)}">
                <xsl:value-of select="."/>
            </xsl:attribute>
        </xsl:for-each>
        <xsl:for-each select="$node/*">
            <xsl:if test="my:isStereotype(., name(./parent::*), 'Role') = 'true'">
                <!-- if is a 1-to-many relationship adds the role tag -->
                <xsl:if test="compare($multiplicity, 'true') = 0">
                    <xsl:element name="{name(.)}">
                        <!-- this template resolves role and
                        entities attributes and collapses them -->
                        <xsl:call-template name="collapsingAttributes">
                            <xsl:with-param name="role" select="." />
                        </xsl:call-template>
                    </xsl:element>
                </xsl:if>
                <xsl:if test="compare($multiplicity, 'true') != 0">
                    <xsl:call-template name="collapsingAttributes">
                        <xsl:with-param name="role" select="." />
                    </xsl:call-template>
                </xsl:if>
            </xsl:if>
            <xsl:if test="my:isStereotype(., name(./parent::*), 'Role') != 'true'">
                <xsl:call-template name="resolveDataType">
                    <xsl:with-param name="node" select="." />
                </xsl:call-template>
            </xsl:if>
        </xsl:for-each>
    </xsl:element>
</xsl:template>
```

Figure 8: Portion of the XSLT document highlighting the Denormalize Hierarchy template.

attributes. Otherwise if the relationship is 1-to-many the hierarchy cannot be fully denormalized and a bridge class is needed. To accomplish this task the attributes of entity nodes are resolved and included in the schema as children of the role node.

Moreover, the following functions have been implemented and used in the above-described templates, identified by a rectangle with the plus sign against the bottom line:

- Resolve Data Type: it analyses a complex attribute and store each property in a single column of the relevant table on the basis of the data type schema. However, attributes that assume multiple values (e.g. value of the Observation class) are modeled creating a bridge table to associate each attribute instance to the relevant node.
- Node is a participation: it checks if a relevant node belongs with a participation stereotype of the HL7 RIM on the basis of the CDA schema.
- 1-to-many relationship: it examines whether the multiplicity of the relationship between the relevant node and its father is 1-to-many on the basis of the CDA schema.
- Collapse attributes: starting from the participation node, this function collects the attributes of both role and entity nodes and collapse them in a single node after resolving data types.

The result of this process is a XML document that can be subsequently pruned and grafted considering the specifications of the user with a particular attention on nodes considered unnecessary for the purpose of the business process analysis.

4 Transformation of a CDA Document: A Case Study

In this paper, the proposed approach is tested on a case study that analyses current and historically relevant vital signs. This information is collected in different specifications of the CDA schema produced by different organizations during different events, depending also on the national implementations. For instance, in Italy this information is stored and exchanged using the Report that collects results based on observations generated by laboratories and the Discharge letter that gathers information relative to the patient's hospitalization.

At international level HL7 has released an implementation guide, the Continuity of Care Document (CCD) [23], to share patient clinical data specifying the structure and semantics of a patient summary clinical document. In this paper the attention will be focused on the vital signs section of the CCD that models individual's clinical findings, such as blood pressure, heart rate, respiratory rate, height, weight, body mass index, head circumference, crown-to-rump length, and pulse oximetry.

For the purpose of our case study we choose the class Observation as a Fact of the dimensional model given that it describes an "action performed in order to determine an answer or a result value". This is the starting point to transform the CDA document in a XML document to be loaded in the data warehouse as reported in the example depicted in Figure 9, where the main template that implements the function to visit the XML tree is based on the proposed methodology. Navigating the tree in a child-parent direction each Observation node will include its ancestors with relevant attributes, such as organizer, section and ClinicalDocument. Moreover, both children of the ClinicalDocument node (i.e. recordTarget and documentationOf) are included in the model as children of the Observation node, along with their children. Subsequently, the tree is parsed in a parent-child direction and the only child of the Observation node (i.e. referenceRange) is included in the model. During these activities each attribute is analyzed and resolved through the template Resolve Data Type taking into account the HL7 data type they are belonging to and also considering if they are multi- or single-valued attribute. This task will be better analyzed in the following when the denormalization of HL7 hierarchies is addressed.

5 Conclusion

The paper presents a systematic approach to extract clinical information from CDA documents and to transform them in a XML document to be loaded in a data warehouse for secondary purposes. It is based on a conceptual framework that maps the primitives of the CDA schema with the concepts of the dimensional model. The transformation procedure proposed is based on the widely diffused XSLT style sheet language. It analyses the original XML document structured on the basis of the CDA schema to derive the Fact as well as the relevant measures and dimensions of the data mart schema without specific user requirements, thus representing the original information on the basis of the snowflake schema. The result of this transformation is a XML document

This approach will be further tested on a wider set of clinical documents based on different CDA specifications, such as discharge report forms, prescription of pharmacological products and specialist visits, patient summary. This semi-automatic procedure will be applied on

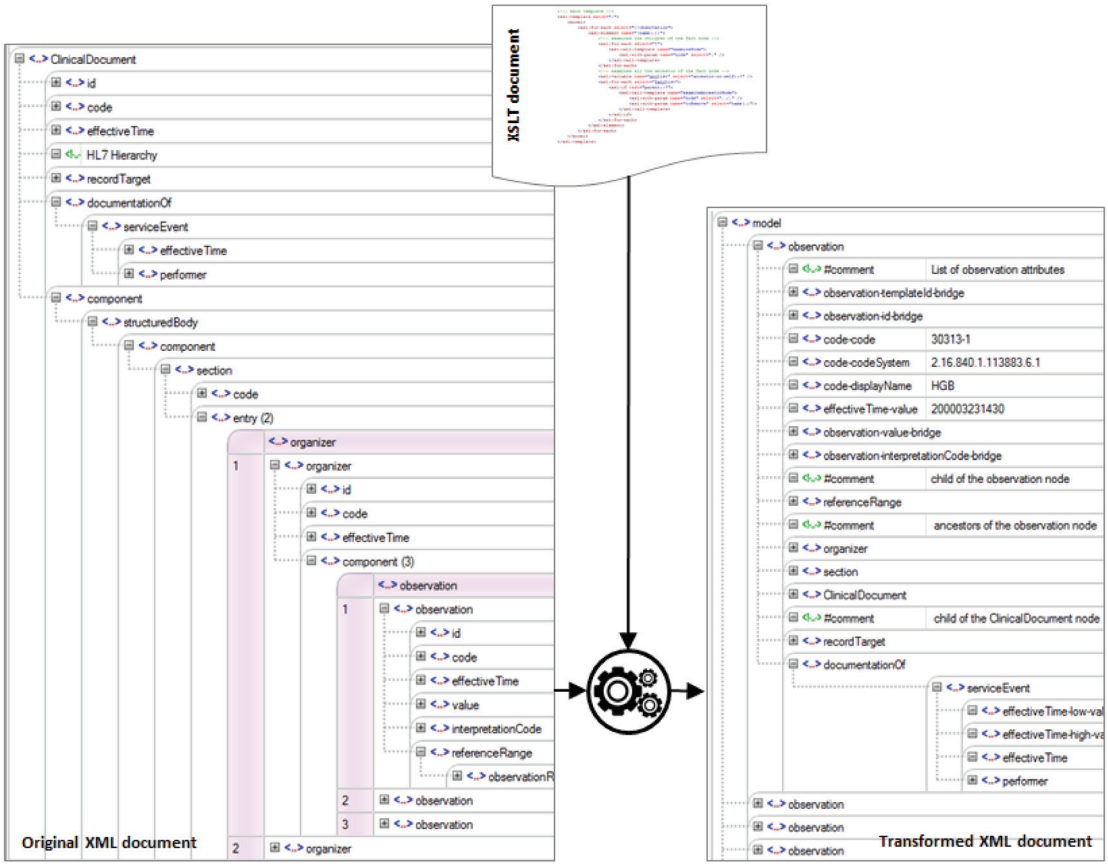


Figure 9: Transformation of the original XML document structured based on the HL7 CDA standard schema in a dimensional model oriented XML document based on the XSLT document.

the Smart Health 2.0 national project that aims to develop a regional healthcare infrastructure based on HL7 standards. It will be used to develop a dashboard to assess the quality of healthcare service provided in the framework of continuity of care. Starting from a set of selected quality indicators this approach will enable to extract data from CDA documents stored in the Electronic Health Record (EHR) and semi-automatically transform and store them in a data warehouse for secondary purposes in a clinical governance framework.

References

[1] Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper. J Am Med Inform Assoc. 2007; 14: 1-9.

[2] Elkin PL, Trusko BE, Koppel R, Speroff T, Mohrer D, Sakji S,et al. Secondary use of clinical data. Stud Health Technol Inform. 2010; 155: 14-29.

[3] Wickramasinghe N, Schaffer JL. Creating knowledge-driven healthcare processes with the intelligence continuum. Int J Electron Healthc. 2006; 2: 164-174.

[4] World Wide Web Consortium (W3C). XSL Transformations (XSLT), Version 2.0, W3C Recommendation 23 January 2007. Available from: http://www.w3.org/TR/xslt20/

[5] Schadow G, Mead CN. The HL7 Reference Information Model under scrutiny. Stud Health Technol Inform. 2007; 124: 151-156.

[6] Benson T. SNOMED CT. Principles of Health Interoperability HL7 and SNOMED. London: Springer; 2010.

[7] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, et al. HL7 clinical document architecture, release 2. J Am Med Inform Assoc. 2006; 13: 30-9.

[8] Sen S, Ghosh R, Paul D, Chaki N. Integrating XML Data Into Multiple Rolap Data Warehouse Schemas. In Int J Soft Eng & Appl. 2010; 3: 197-206.

[9] Kavitha P, Vydehi S. Query Processing of XML Data Warehouse Using XML Pattern Matching Techniques. Int J Eng Res Technol. 2014; 3.

[10] Golfarelli M, Rizzi S, Vrdoljak B. Data warehouse design from XML sources. Proceedings of ACM International Workshop on Data Warehousing and OLA. Atlanta, GA, USA. 40-47. 2001.

[11] World Wide Web Consortium (W3C). XSL Transformations (XSLT), Version 2.0, W3C Recommendation 23 January 2007. Available from: http://www.w3.org/TR/xslt20/

[12] Pecoraro F, Luzi D, Ricci FL. Designing a Data Warehouse Dimensional Model Based on the HL7 Clinical Document Architecture. Proceedings of International Conference on Health Informatics (HEALTHINF). Lisbon, Portugal, 284-292. 2015.

[13] Pecoraro F, Luzi D, Ricci FL. A Clinical Data Warehouse Architecture based on the Electronic Healthcare Record Infrastructure. Proceedings of 7th International Conference on Health Informatics (HEALTHINF), Angers, France. 2014.

[14] Luzi D, Pecoraro F, Mercurio G, Ricci FL. A medical device Domain Analysis Model based on HL7 Reference Information Model. Proceedings of International Conference on Medical Informatics in a United and Healthy Europe (MIE). Sarajevo, Bosnia and Herzegovina. 2009: 162-166.

[15] Inmon WH, Zachman JA, Geiger JG. Data stores, data warehousing and the Zachman framework: managing enterprise knowledge. New York: McGraw-Hill, Inc.; 1997.

[16] Thomas H, Datta A. A conceptual model and algebra for on-line analytical processing in decision support databases. Inform Sys Res. 2001: 12(1): 83–102.

[17] Hoffer JA, Venkataraman R, Topi H. Modern database management. New Jersey: Prentice Hall, Upper Saddle Review; 2002.

[18] Eggebraaten TJ, Tenner J W, Dubbels JC. A health-care data model based on the HL7 Reference Information Model. IBM J Res Dev. 2006; 46: 5-18.

[19] Hümmer W, Baur A, Harde G. XCube: XML for data warehouses. Proceedings of the 6th ACM international workshop on Data warehousing and OLAP. 2003: 33-40.

[20] Boussaid O, Messaoud RB, Choquet R, Anthoard S. X-warehousing: an XML-based approach for warehousing complex data. Adv Data Inform Sys. 2006; 39-54.

[21] Park BK, Han H, Song IY. XML-OLAP: A multidimensional analysis framework for XML warehouses. Data Warehousing and Knowledge Discovery. 2005; 32-42.

[22] Boussaid O, Messaoud RB, Choquet R, Anthoard S. X-wacoda: An xml-based approach for warehousing and analyzing complex data. Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction. 2009; 38-54.

[23] Health Level Seven, Inc. HL7 Implementation Guide. CDA Release 2 – Continuity of Care Document (CCD). Ann Arbor: HL7; 2007.

European Journal for Biomedical Informatics is dealing with a range of topics which is included under the scopes of biomedical informatics. The focus of the journal is on the major aspects of implementation of information science in Medical sciences, which includes Biomedicine, Ontology, Genomics, Biometrics, Biostatistics, Health informatics, Medical informatics, Clinical informatics, Semantics Interoperability, Informatics applications in Imaging.

Original contribution in terms of research, review, short-communication, case reports, are welcomed from the authors, who are engaged in medical or academic research. The articles selected for publication, must have to go through critical process of peer-review as per journal norms. A well-constructed editorial board is the higher most determining authority for the journal, which is formed by highly experienced subject experts in the field of biomedical informatics. Section editors and guest editors are also engaged there to maintain topic specific proper evaluation of research manuscripts.

List of Published Special Issues

Issue	Type	Name
2016/1	event-based	Interoperability is more than just technology(HL7 IHIC 2016)
2015/3	topic-based	Apps for Medicine, Health and Home Care - Elements of Safety and Effectiveness
2015/2	event-based	Concept, Models and Implementations for Innovative Interoperable eHealth Solutions (HL7 IHIC 2015)
2014/2	topic-based	Using Information to Improve the Quality of Care in Type II Diabetes in Primary Care
2013/3	topic-based	Sharing Knowledge and Tools for Decision Support in Biomedicine and Healthcare
2013/2	topic-based	Applications of Computational Techniques in Infectious Diseases and Clinical Research
2012/5	topic-based	Semantic Interoperability in Biomedicine and Health Care III.
2012/4	event-based	Standards and Solutions for eHealth Interoperability (part 2)
2012/3	event-based	Standards and Solutions for eHealth Interoperability (part 1)
2012/2	topic-based	Supporting eHealth Applications by Legal Systems in Europe
2011/2	topic-based	Education in Biomedical Informatics and Bioinformatics
2011/1	event-based	Semantic Interoperability in Biomedicine and Health Care II.
2010/1	event-based	Semantic Interoperability in Biomedicine and Health Care

Advancing HL7 v2 to New Heights: A Platform for Developing Specifications, Test Plans, and Testing Tools

Robert Snelick

National Institute of Standards and Technology (NIST), Gaithersburg, USA

Abstract

Development of HL7 v2 data exchange interface specifications has long been problematic, plagued with ambiguous and inconsistent requirement specifications. This situation leads to potential misinterpretation by implementers, thus limiting the effectiveness of the specification and creating artificial and unnecessary barriers to interoperability. Likewise, the ability to test implementations effectively for conformance to the specifications is hindered. The current approach of specification development and test plan creation relies on word processing tools, meaning implementers and testers must read and interpret the information in these documents and then translate it into machine-computable requirements and test assertions. This approach is error prone—a better methodology is needed. We present a set of productivity tools in an integrated platform that allow users to define and constrain HL7 v2 specifications and to develop test plans that result in machine-computable artifacts. A testing infrastructure and framework subsequently uses these artifacts to create conformance testing tools automatically. We present and demonstrate the utility of a platform for developing specifications, writing test plans, and creating testing tools. The value proposition of this end-to-end methodology is explained for authors writing HL7 v2 specifications, for developers implementing interfaces, and for testers creating validation tools.

Keywords

Conformance; Healthcare Data Exchange Standards; Healthcare Information Systems; Interoperability; Specification Development Tools; Testing Tools

Correspondence to:

Robert Snelick

National Institute of Standards and Technology, 100 Bureau Drive Stop 8970, Gaithersburg, MD 20899, USA.

E-mail: robert.snelick@nist.gov

EJBI 2017; 13(1):97-104

received: June 08, 2017

accepted: July 19, 2017

published: October 10, 2017

1 Introduction

For 30 years, HL7 (Health Level 7) Version 2 (v2) has been the predominant standard used for the exchange of healthcare administrative and clinical data. Healthcare information systems use the HL7 v2 protocol to develop standardized interfaces to connect to and exchange data with other systems. HL7 v2 covers a broad spectrum of domains including Patient Administration, Laboratory Orders and Results, and Public Health Reporting. The base HL7 v2 standard [1] is a framework that contains many message events, and for each event it provides an initial template (starting point) that is intended to be constrained for a specific use case. The application of constraints to a message event is referred to as profiling [2, 3]. For example, the VXU V04 (Unsolicited Vaccination Record Update) message event is a generic template for communicating information about a patient's immunization related events. The base message template is composed of mostly optional data elements. For a given use case, e.g., Send Unsolicited Immunization Update for the US Realm [4], the message template is "profiled". That is, elements can be constrained to be required, content can be bound to a set of pre-coordinated codes, and so on. The base message event (e.g., VXU V04) that has been constrained for a particular use (e.g., submitting immunization events) is referred to as a conformance profile[1]. An implementation guide is a collection of conformance profiles organized for a workflow (e.g., submitting, acknowledging, querying, and responding to/for immunization events). In this example, four conformance profiles exist, each with different message events; one for submitting an immunization event, for sending an acknowledgment, for querying for an immunization history, and for providing an immunization history. To date, HL7 v2 implementation guides have been created using word processing programs, which has resulted in ambiguous and inconsistent specification of requirements. This practice has hindered consistent interpretation among implementers, which has created an unnecessary barrier to interoperability.

[1]Also, referred to as a message profile.

We present an end-to-end methodology and platform for developing specifications (implementation guides), writing test plans, and creating testing tools in the HL7 v2 technology space [5]. The platform includes three key foundational components:

- A tool to create implementation guides and conformance profiles
- A tool to create test plans, test cases, and associated test data
- A testing infrastructure and test framework to build testing tools

A key to the approach is that the "normal" process of creating implementation guides, test plans, and testing tools is "reversed". Instead of creating requirements using a natural language and subsequently interpreting the requirements to create test plans and test assertions, the requirements are captured with tools that internalize the requirements as computable artifacts.

Figure 1 illustrates a high-level overview of the methodology. Domain experts develop use cases, determine the message events that correspond to the interactions in the use cases, and then proceed to define the requirements. Using the methodology, they accomplish these tasks by entering this information into the Implementation Guide Authoring and Management Tool (IGAMT). During this process, the domain experts constrain the message events according to the requirements needed by the use case. Section 2 will elaborate more on this process and on the details of how the requirements are constrained. The output of IGAMT is a set of artifacts that are represented in Word, HTML, and XML formats. The complete implementation guide, including the narrative and messaging requirements, can be created in IGAMT and then exported in Word or HTML. Such formats are suitable for ballot at standards development organizations such as HL7 or IHE (Integrating the Healthcare Enterprises [6]). In May 2017, two HL7 v2 implementation guides that were generated by IGAMT were submitted for ballot. Each conformance profile can be exported as XML[2]. The XML format contains all the messaging requirements in a machine-computable representation, which is the most important aspect of IGAMT, since the XML conformance profiles have many uses including a computable definition of the message interface, message validation, test case and message generation, and source code generation.

The XML conformance profiles can be imported into the Test Case Authoring and Management Tool (TCAMT). TCAMT is used to create targeted test cases for interactions (profiles) defined in the implementation guide. The output is an additional set of constraints in an XML format. The entirety of the output generated from IGAMT and TCAMT is called a "resource bundle"[3].

The NIST platform includes a testing infrastructure of common utilities used for testing, such as a message validation engine, along with a testing framework that provides various testing tool components, such as a communication framework and a profile viewer. Testing Tool instances are then created using both the testing infrastructure and framework components as well as the resource bundle output generated from IGAMT and TCAMT.

The NIST platform allows end users to create conformance testing tools by means of a set of productivity tools. This streamlined approach can greatly reduce today's problems with conformance test tools. These problems include: tools often don't exist, they are expensive to build, they are difficult to update in a timely fashion, they are not adaptable for local refinements, and their time to market is lengthy. Additionally, the platform provides value through enforcing consistent and rigorous rules for requirements specifications.

The remainder of this paper explains the NIST platform in more detail in the context of how it can be applied in real-world use case settings. We first describe how IGAMT is used to define and constrain conformance profiles. One important aspect is the application of recently developed methods and best practices for requirements specification. Additionally, a brief overview of the validation process is given. Next, an explanation of how a set of targeted test cases are created in TCAMT is provided. In Section 4 we discuss a testing infrastructure and framework components. Next, an overview of the resulting test tools and how they are created is presented. Finally, there is a discussion on how the platform supports testing capabilities beyond the scope of the HL7 v2 interoperability specification. One goal of this paper is to inform the reader about the ease with which HL7 v2 implementation guides, test cases, and testing tools can be created using the NIST platform compared to the current laborious methods used today.

2 IGAMT

IGAMT [5] is a tool used to create HL7 v2.x implementation guides that contain one or more conformance profiles. The tool provides capabilities to create both narrative text (akin to a word processing program) and messaging requirements in a structured environment. Our focus in this paper is on the messaging requirements.

IGAMT contains a model of all the message events for every version of the HL7 v2 standard. Users begin by selecting the version of the HL7 v2 standard and the message events

[2]The XML format is defined by NIST and is publicly available but is not yet standardized. NIST intends to propose the format to HL7 for adoption. Additionally, there is no relationship between this format and other HL7 profiling formats such as the Templates Implementable Technology Specification (ITS) standard and FHIR.

[3]Is not related to a resource bundle in FHIR (Fast Healthcare Interoperability Resources).

they want to include and refine in their implementation guide. For example, the message events VXU^V04, ACK, QBP^K11, and RSP^K11 are used to create eight conformance profiles in the immunization implementation guide [4]. Each message event is profiled (constrained) to satisfy the requirements of the use case. The QBP and RSP message types are used more than once to specify different uses.

Rules for building an abstract message definition are specified in the HL7 message framework, which is hierarchical in nature and consists of building blocks generically called elements [1]. These elements are segment groups, segments, fields, and data types (i.e., components and sub-components). The requirements for a message are defined by the message definition and the constraints placed on each data element. The constraint mechanisms are defined by the HL7 conformance constructs, which include usage, cardinality, value set, length, and data type. Additionally, explicit conformance statements are used to specify other requirements that can't be addressed by the conformance constructs. The process of placing additional constraints on a message definition is called profiling. The resulting constrained message definition is called a conformance profile (also referred to as a message profile). An example of a constraint is changing optional usage for a data element in the original base standard message definition to required usage in the conformance profile.

IGAMT provides, in a table format user interface, the mechanisms to constrain each data element at each level in the structure definition. The rows of the table list the data elements according to the structure definition being constrained (segments, fields, and data types). The columns list the conformance constructs that can be constrained for a data element, including the binding to a value set. Figure 2 shows a screen capture of the navigation and the segment profiling panels. On the left-hand side, the user can select the object to edit. The right-hand side displays the list of fields in

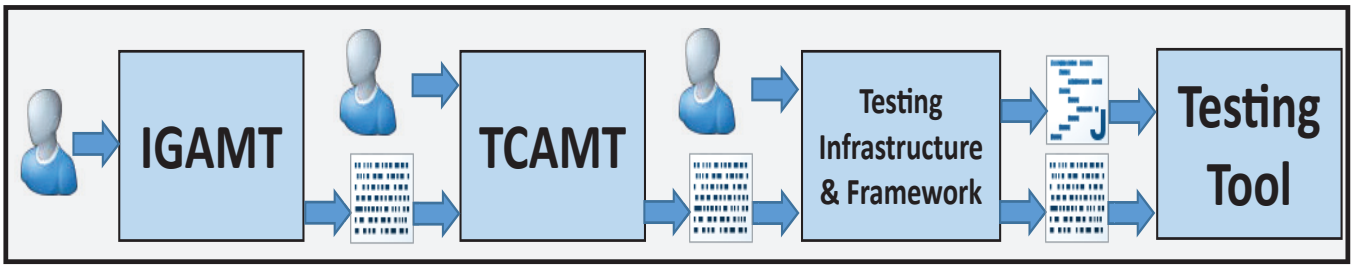


Figure 1: NIST HL7 v2 standards development and testing platform overview.

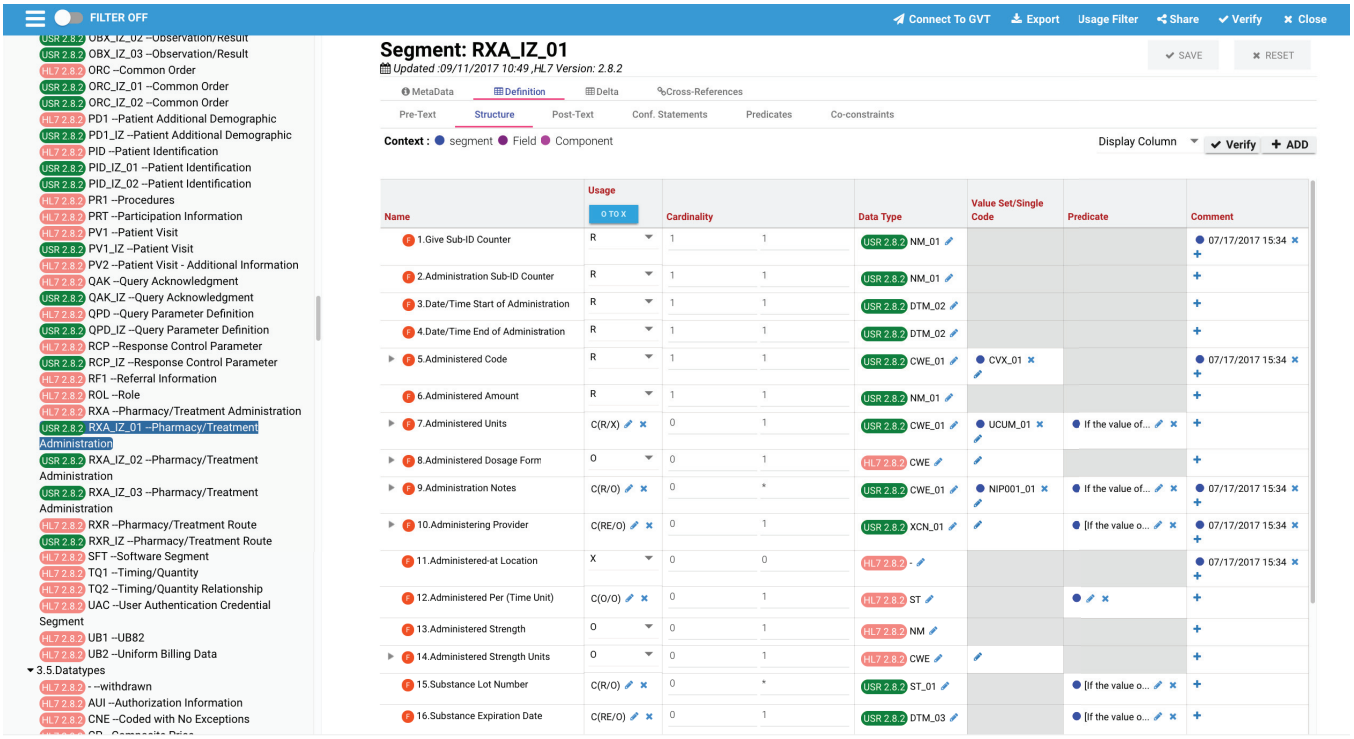


Figure 2: IGAMT screen capture: navigation and segment profiling view.

the segment and the requirements that can be specified for the field.

One key philosophy of IGAMT is the capability of creating and reusing building block components. These lower level building blocks can be used to create higher level constructs efficiently. The building blocks include data type flavors, segment flavors, and profile components. A base data type can be constrained for a given use; the resulting data type is called a data type flavor (or data type specialization). A given base data type may have multiple data type flavors. These flavors can be saved in libraries and reused as needed. A similar process applies to creating segment flavors.

A profile component represents a subset of requirements that can be combined with other profiling building blocks. One such example is the definition of a profile for submitting immunizations. The Centers for Disease and Control and Prevention (CDC) creates a national level profile, however, individual states may have additional local requirements that can be documented in a profile component. Only the delta between the national and local requirements is documented in the profile component. Combining the national level profile and the state profile component yields a complete (composite) profile definition for a given state. Another example is for the case of sending laboratory results and reportable laboratory results to public health. The use cases are very similar. The reportable laboratory results have additional requirements; therefore, a profile should be created for sending laboratory results, followed by a profile component for reportable laboratory results. A composite profile for the public health use case can be created by combining the profile and the profile component. This design principle provides a powerful and effective approach for leveraging existing profiles and profile components [2].

A utility for creating and managing value sets is also provided. Specific value sets can be created and bound to data elements. For example, a base HL7 v2 table can be cloned and modified ("constrained") to create a value set for a specific use, thus enabling more granular value set bindings [2]. Instead of binding an entire HL7 v2 table to an element (typical practice), a value set containing only codes relevant to that element for a particular use is specified. Using this approach, multiple value sets are derived from a single HL7 v2 table, which provides clear requirements for implementers. Mechanisms for creating value set libraries are provided to promote reuse.

2.1 Improved Requirements Specification

In the effort to create conformance test tools for the Office of the National Coordinator (ONC) certification in support of the US Centers for Medicare and Medicaid Services (CMS) Meaningful Use (MU) program, it quickly became apparent that the HL7 v2 specifications named in the ONC rule were ambiguous, under-specified, and inconsistent. This made it difficult to create rigorous, comprehensive, and meaningful test tools and test cases to adequately validate vendor implementations for the ONC stated goal of enabling interoperability. If implementers can interpret and implement requirements in different ways, interoperability is impeded. To improve this situation, NIST worked closely with the specification authors and other stakeholders to gain clarity and subsequently co-published addendums and errata. This effort revealed deficiencies in the mechanisms for specification of requirements and approaches for creating implementation guides. As a remedy, new and improved methods for specifying requirements emerged along with a set of best practices. IGAMT incorporates these methods and encapsulates, automates, and simplifies how the requirements are specified. Table 1 provides a list of the most important methods, concepts, and best practices for improved specifications (beyond current practices).

2.2 IGAMT Message Model and Validation Process

IGAMT has an internal model of all HL7 v2 messages for each version of the standard (Figure 3). HL7 v2 publishes the standard in human readable text documents. Message definitions and accompanying structures are *codified* into a data base, which is available from HL7. IGAMT reads the data base and converts the message definitions into the IGAMT message model. The message model is the anchor on which all IGAMT functions and features are based. IGAMT reveals the model via a graphical user interface (GUI) where the user can constrain the message as needed. The user interface displays panels for the Message, Segment, Data Type, Value Set, Profile Components, Condition Predicates, and Conformance Statements. IGAMT exports the constrained message definition (a profile) as an XML profile instance. IGAMT ensures that the XML profile instance adheres to the rules of the Profile Schema. Validation is performed by validating a message instance against the constraints defined in the XML Profile. The validation engine interprets the requirements as documented in the XML Profile and makes assertions against the message instance accordingly. A Validation Report is generated. The validation process forms the basis of the conformance test tools.

3 TCAMT

TCAMT [5] is a tool used to create HL7 v2.x test plans that contain one or more (typically many) test cases. Key features in TCAMT include test plan creation (narrative and computable), IGAMT XML profile import, HL7 v2 message

Table 1: Methods, concepts, and best practices for improved specifications.

Concept	Issue	Feature/Improvement
Explicit Condition Predicates	Conditional usage is specified but lacks conditional statement or an explicit conditional statement	Explicit condition predicate with defined format, style, and pre-defined patterns
Condition Predicate True/False Outcomes	Limited True/False outcomes for conditional usage (C and CE only)	Full range of true/false outcomes; for example, C(R/RE) and C(RE/O)
Explicit Conformance Statements	Statements that hinted at being requirements are hidden in narrative sections of the specification	Explicit conformance statements with defined format, style, identification, and pre-defined patterns
Data Type Flavors	Conflated specializations of data type constraints, in-line constraints, un-managed data type flavors	Explicit data type flavor definitions, naming conventions, and style
Data Type Flavor Library	No notion of creating a library of data flavors for reuse by the community at-large	Master set of data type flavors and defined process for user defined flavors; promote consistency and reuse
Segment Flavors	Segments typically are defined to account for requirements for use in more than one message definition—resulting in conflation of requirements	Provide mechanisms to allow specific segment definition via segment flavors, profile components or explicit conformance statements
Profiling Multiple Occurrences	Capability to assign different data type flavors to multiple occurrences to a field element; defined in v2.8	Implemented in IGAMT and in XML profile instance; can vary by "type code", "order", and "one of"
Co-constraints	Missing, inconsistent, or lack of detailed specification of relationship among data element content; typically, in elements OBX-2, OBX-3, and OBX-5	Mechanism to define data element content relationships and dynamic data type flavor mapping for OBX-2 and OBX-5[5]
Value Set Specification	No explicit value set or code table specifications; often the base HL7 or HL7 User table is bound to an element (or elements) with no further constraints	Explicit value set definition creation and value set binding strength
Value Set Profiling	No formal methodology to constrain code systems for specific element binding and use	Explicit value set definition usage indicator for codes and attributes to indicate extensibility and stability
Profile Components	No constructs or methods to define profile building blocks of constraints for reuse	Profile components are introduced to defined a set of arbitrary requirements that when combined with a profile or other profile components create a complete profile (Composite Profile)
Delta Profiles	Complete specifications for closely related use cases	"delta" specifications can be created leveraging the concept of profile components
IG Template	No guidance on what implementation guides should contain	IGAMT incorporates several default templates and export options
Conformance Keywords	Non-existence and inconsistent definition and use of verbs to express requirements	Explicit definition and use of conformance keywords as part of the IG template; based on RFC 2119

[5]For example, based on different codes in OBX-3, different data type flavors of the same base data type can be specified in OBX-2 that indicates the requirements in OBX-5. This enables precise requirements definition.

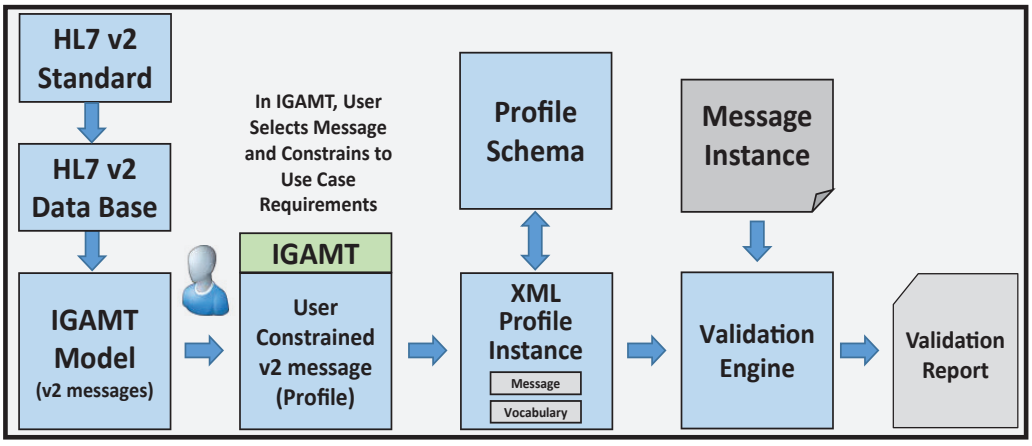


Figure 3: IGAMT message model and validation process.

creation and import, constraint editing, constraint and messaging templates, and multiple export formats. A test case can consist of one or more test steps. A test step can be an HL7 v2.x interaction or a manual step such as visually inspecting the contents of an application's display screen. Each test case and test step can consist of a test description,

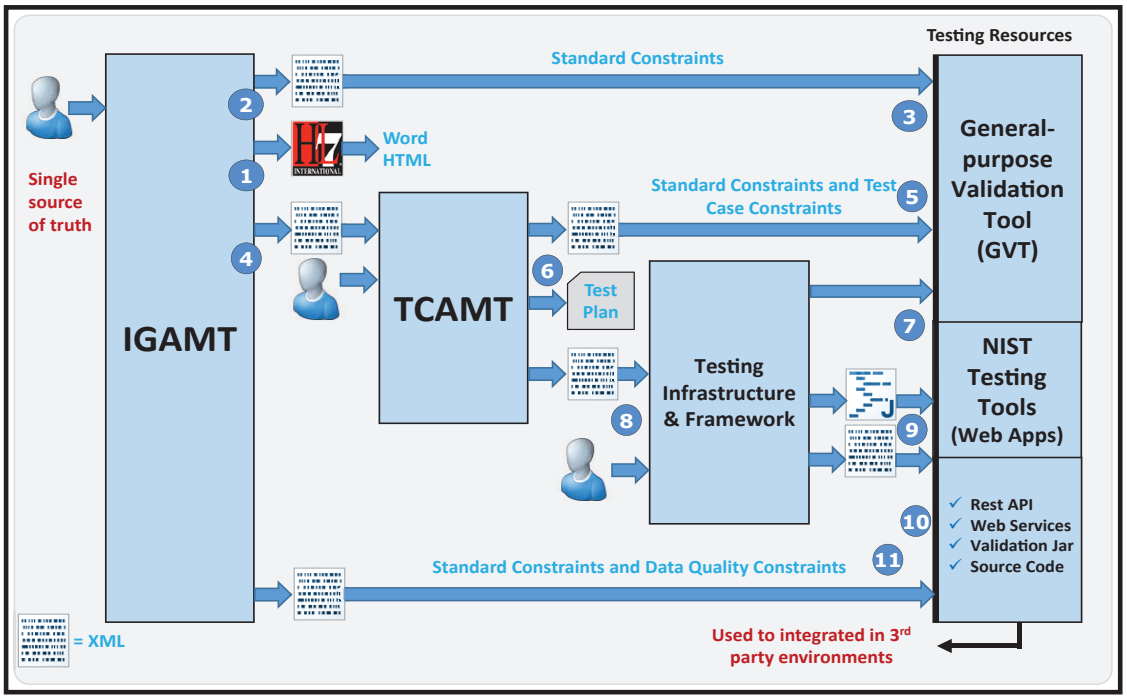


Figure 4: NIST HL7 v2 standards development and testing platform architecture.

pre- and post-conditions, objectives, evaluation criteria, and additional notes and comments. Test steps for an HL7 v2.x interaction contain an HL7 v2 message (with specific data) that aligns with the XML conformance profile created from IGAMT[4].

Targeted test cases are critical for assessing the capabilities of a system. TCAMT allows domain experts to create test cases (that include example messages) for certain scenarios and capabilities. Test cases provide context, which expands the scope of testing. Without context, a validation tool cannot test a message exhaustively to all requirements specified in the implementation guide. For example, elements with "required, but may be empty (RE)" usage, elements with "conditional usage (C)", or elements with cardinality greater than "1" cannot be assessed without targeted tests. A message that is validated against the requirements of a conformance profile without any provided context is called "context-free testing". A message that is validated against the requirements of a conformance profile and with a provided context is called "context-based testing" [2]. The test cases provide context, and TCAMT is a tool that allows users to create the test cases.

A key design component in TCAMT is its use of the XML profiles created in IGAMT as a foundation. The message definition defined in the profile provides the foundation such that data associated with each message element of interest can be specified. TCAMT also allows the user to enter additional assertion indicators based on what they want to test. For example, for an element with a usage of "RE", the user can provide data that are expected to be entered into the sending system for the element and can select an assertion indicator. There are several assertion indicators that could be selected, for example, "presence". In this case, if the user provides test data and selects the indicator of "presence", a constraint is generated by TCAMT and is provided to the validation. For elements with "RE" usage, the element must be supported by the system-under test (SUT), but in a given message instance the element may not be populated. For this construct, the tester wants to ensure that the implementation has, in fact, included support for the element.

In a context-free environment, the absence of data in a message is not a conformance violation for elements with "RE" usage. However, in the example test case described above, data were provided and a presence constraint was specified. Now, when a message created for this test case is validated, the additional constraint triggers an assertion for the presence of data for this element. This method is one way to determine support for the element.

Via TCAMT, the user can create an unlimited number of test cases and test a broad spectrum of requirements. Other constraint indicators can be used to test for specific content or for the non-presence of an element. Additionally, test data can be provided to trigger conditional elements. In other instances, support for certain observations may need to be ascertained. In such cases, test data for specific observations (e.g., in an immunization forecast, the vaccine group, earliest date to give, and due date) can be provided, requiring the message instance to contain an OBX segment

[4]Not necessarily conformant data; invalid data may be used in the testing process

for each observation. The test case might be set up to expect certain LOINC[6] codes to ensure each observation (capability) is implemented by the system. TCAMT provides the mechanisms to conveniently and consistently create test cases. Output from TCAMT provides the additional constraints that are interpreted by the validation engine.

4 Testing Infrastructure and Framework

NIST has built an HL7 v2.x testing infrastructure and framework to aid in the process of creating conformance testing tools. The testing infrastructure provides a set of services utilized by the test tool framework to build specific instances of tools. A test tool can be built for a specific need or to be a general-purpose tool to handle multiple implementation guides and profiles. The latter tool is a web application where a user can upload implementation guides, conformance profiles, and test plans to "create" a test tool. The test tool is "built-on-the-fly" and can be generated as a by-product "for free" once the XML profile and associated artifacts have been created (in IGAMT and TCAMT). This process allows domain experts to "build" the test tool. Alternatively, the framework can be leveraged, customized, and installed locally. Using the framework, developers can choose to create customized, specific, or general-purpose web application conformance test tools, and they can access the validation via web services or incorporate validation via a JAR (Java Archive) file or source code. Regardless of the use, the platform can significantly improve the quality of implementation guides, assist in the creation and maintenance of test plans, expedite the stand-up of a validation tool, and, overall, reduce the cost and time of the entire process.

Figure 4 shows in more detail the end-to-end methodology and platform. A key design principle is that there is a single source of truth in the creation of implementation guides and test plans. Modifications are made in one place and are propagated to associated services, utilities, and tools. IGAMT is a tool used by domain expert authors to define requirements for interface specifications. Human readable (1) and machine computable (2) artifacts are exported. A context-free conformance test tool is automatically generated when the IGAMT XML profiles are loaded in the general-purpose validation tool (3). At this level, validation is based on the technical requirements defined in the profile. No context is associated when validating the message instance against the requirements defined in the profile. This type of validation is called context-free testing.

Point (4) shows the XML Profile as input into TCAMT. Test scenarios provide a context, that is, a real-world story with associated data. Additional constraints are generated from having context. The profile and context constraints are loaded into the general-purpose validation tool to create a context-based validation tool automatically (5). Point (6) indicates a human readable export of the Test Plan.

Point (7) indicates that the testing infrastructure and framework components are used as the basis for the general-purpose validation tool. The general-purpose validation tool is itself a tool that takes as input the resource bundle (XML Profile, TCAMT constraint file, etc.) to automatically generate a conformance test tool. Points (8) and (9) indicate the process by which developers can leverage the testing infrastructure and framework to create customized conformance test tools. Point (10) indicates that validation can be accessed via other methods that allow a user to integrate it into their local environments. The platform provides access to the tool validation via REST and web services. Additionally, the validation JAR and source code are available. Point (11) indicates that additional constraints can also be included that go beyond the scope of typical interface requirements. These can include data quality business rules, for example, ensuring that a vaccine dose reported is consistent in terms of the manufacturer, lot number, and date given. More on this topic is given in Section 6.

5 Conformance Test Tools

As shown, conformance testing tools are built using the testing infrastructure and framework, the IGAMT-produced conformance profiles, and the TCAMT-produced test plan. Testing tools are web-based applications that can support both context-free and context-based validation [5]. In addition to performing message validation, the tools provide a browse-able view of the requirements for each conformance profile. In the context-based mode, the test story, test data, and an example message are provided for each test step.

In the context-free mode, the user simply selects the conformance profile to validate against and then imports the message. The validation is performed automatically and a report is given. In the context-based mode, the user selects the test step and imports the message to validate. The test tool sets the validation to the conformance profile linked to the test step, performs the validation, and provides a report. In both modes, a tree structure of the message is shown on the left panel of the validation screen and can be used to inspect the content of individual data elements.

Test plans can be executed in non-transport mode and transport mode. Non-transport mode provides an interface to upload (cut/paste or load file) a message into the validation edit box. Transport mode allows an application to connect to the test tool to exchange messages interactively. The test tool can act as an initiator or responder as directed by the test plan. Various transport protocols are supported including MLLP and SOAP. Test Cases can also include manual test steps in addition to automated test steps that contain an HL7 v2 message exchange.

6 Requirements beyond the Interface Specification

The intent of HL7 v2 is specifically scoped to defined requirements for exchanging data between applications. The specifications typically do not impose any requirements on how the data are processed. Other specifications, in conjunction with the interface specification may specify such requirements (e.g., IHE integration profiles and functional requirements specifications). In real world settings, exchange partners need to account for more than just conformance to the exchange requirements. Data quality, business rules, and functional requirements are necessary to satisfy the desired outcome of the use case scenario. Mechanisms to define such requirements, and testing support that can verify that the complete workflow is implemented as intended, are beneficial.

The generic constraint generation utility in IGAMT can be used to create data quality constraints. Certain business rules can be applied to a message to determine if it meets the requirements necessary for incorporation by the receiver. A simple data quality rule for reporting an immunization record is that the date of administration must be after the date of birth. This constraint likely is never given in an HL7 v2 interface specification, however, data quality rules such as these are important at the local level. Users can create these rules in IGAMT that provide additional validation (point (11) in Figure 4).

TCAMT can be used to create test cases to test functional requirements. For example, a scenario can be crafted in which three different immunization records for the same patient are created from different providers and sent to an immunization information system (IIS). A subsequent query to the IIS to return a complete immunization history can be performed. The response message can be examined to see if the consolidated record contains the expected combined immunization history. TCAMT provides the capability to create such a scenario and the additional content validation constraints. Testing for invalid (or negative) test case scenarios also can be created. The platform provides the capabilities for the tester to create unlimited test scenarios using convenient and powerful tooling.

7 Conclusion

We presented an end-to-end methodology and platform for developing standards, writing test plans, and creating testing tools in the HL7 v2 technology space. The platform includes three key foundational components: (1) a tool to create implementation guides and conformance profiles; (2) a tool to create test plans, test cases, and associated test data; and (3) a testing infrastructure and test framework to build testing tools. Requirements are captured in IGAMT and exported as conformance profiles. TCAMT is used to create a set of test cases based on the conformance profiles. A conformance test tool is created by combining the validation and associated artifacts with the testing infrastructure and framework.

Acknowledgements

I'd like to thank the NIST development and analysis team: H. Affo, W. Jung Yub, H. Tamri, I. Mellouli, A. El Ouakili, C. Rosin, N. Crouzier, M. Lefort, and S. Martinez; reviewers S. Taylor and F. Oemig, and finally C. Newman for providing user feedback on the tool suite.

References

[1] Health Level Seven International, Inc. Health Level 7 (HL7) Standard Version 2.7, ANSI/HL7, January, 2011, http://www.hl7.org.

[2] Oemig F, Snelick R. Healthcare Interoperability Standards Compliance Handbook. Basel: Springer International Publishing Switzerland; 2016. ISBN 978-3-319-44837-4.

[3] Snelick R, Oemig F. Principles for Profiling Healthcare Data Communication Standards. Software Engineering Research and Practice (SERP13), WORLDCOMP'13 July 22-25, 2013, Las Vegas, NV, 2013.

[4] Health Level Seven International, Inc. HL7 Version 2.5.1 Implementation Guide for Immunization Messaging; Release 1.5, October 1, 2014. http://www.cdc.gov/vaccines/programs/iis/technical-guidance/downloads/hl7guide-1-5-2014-11.pdf

[5] National Institute for Standards and Technology. NIST Resources and Tools in Support of HL7 v2 Standards. http://hl7v2tools.nist.gov/

[6] Integrating the Healthcare Enterprise (IHE). http://www.ihe.net