Accessing Patient Data for Chronic Disease Management

DP Hansen, MJ Lawley, D Vickers, J Ryan-Brown, K Harrap, M O'Dwyer, K Melki and S McBride

The Australian E-Health Research Centre, CSIRO ICT Centre, Brisbane, Australia

Abstract

An increasing number of patients are living with chronic disease, leading to an increase in the amount of data which is captured in their on-going treatment. With this data spread across many different databases, there are varied and complex issues with making it accessible as required. These include issues of privacy and security as well as the data-centric issues of syntactic (how is it captured) and semantic (what does it mean) consistency. With such complex data it is essential to provide user interfaces which make it easy for end users to find and understand their data, be they health professionals or the patient themselves. In this paper we discuss the development of a patient data access system which uses a semantic layer between the original data sources and a novel interface to the clinical data which enables easy access to summary data and the full record.

Background:

An increasing amount of data is being collected during the treatment of patients. This is particularly the case for patients with chronic diseases, due both to the number of interactions they have with the health system and the increasingly complex mechanisms of capturing data. Patient data is held in a myriad of different formats and databases, and, due to the long term nature of chronic disease care, consistent and reliable access to this data is both problematic and necessary. Health systems world wide are now adopting clinical terminologies to provide a standard for describing clinical data. The expectation is that collecting data using a single terminology will increase the level of understandability and completeness of the captured data. However, such terminologies are necessarily large and complex, providing both wide coverage of clinical findings and related concepts and allowing for relationships between clinical concepts to be described.

Currently many health systems are working to adopt SNOMED CT™ (American College of Pathologists, 2006) as the clinical terminology to use when collecting health data. The uptake of this terminology will require many modifications in the workflow and work practices of health systems. Information systems will need to be changed to capture data using the new concepts, instead of existing terminologies, standards and local data sets. As well, extensions to the terminology will need to be made to ensure that the required clinical descriptions are available to use.

While much effort is being expended to ensure data is available in a consistent manner for search and retrieval, the actual interface presented to end-users must encourage use. Improving the human-computer interface is one of the largest challenges for Health Information Technologies today (Sittig et al, 2007). An intuitive interface based on the intrinsic nature of the overall data relationships, and allowing different levels of depth in accessing the data would be desirable.

Methods:

Much patient data is captured in heterogeneous data sets, often within different organisational units. Problems in accessing data from these distributed data sets include that they are held by more than one custodian and that the patient identifiers used are different.

To address these access issues, our data linking and integration tool, HDI (Hansen, 2007), offers the ability to integrate health information across organisational boundaries without needing to surrender that information to a centralised database/warehouse. These features provide support to clinicians and data custodians who have legal, ethical and organisational obligations to the health information they control. This tool has thus far primarily been used to access data for secondary use purposes, such as reporting, clinical and medical research. However, the tool is equally capable of being used in a primary setting, since it meets the requirements of privacy and security of health information. The HDI Remote Domain (Ryan-Brown 2007, Ryan-Brown 2007a) feature set in particular enables collaborating organisations to share health information in a manner which supports the legal, ethical and organisational obligations of the data custodians, ensuring the security of data as it travels across Internet Protocol networks by using Public Key Infrastructure (PKI) technologies to encrypt network connections.
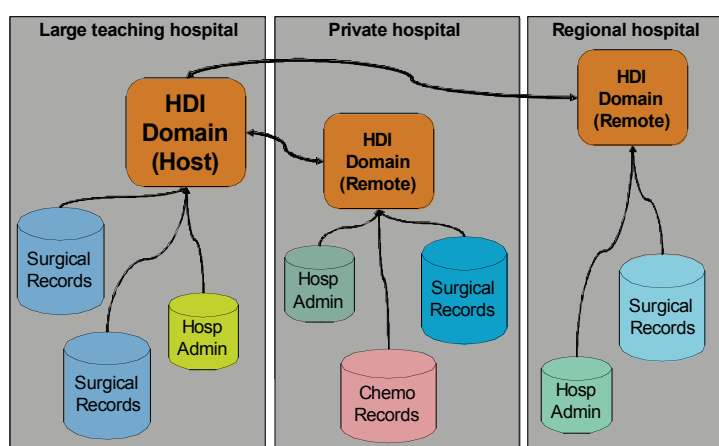

Figure 1. *A typical example of patient data spread amongst many data sources, and the HDI installation topology which would be used to provide access to these data sets.*

Figure 1 shows a use case of data collected during the course of treatment for a patient with colorectal cancer, although the scenario will be similar for most patients receiving on-going treatment for a serious illness in Australia's fragmented health system. It is more than likely that the patient will receive initial tertiary care in a large hospital, then will have follow up treatment in one or more other health facilities. This means that the record of their treatment will be spread between multiple different data sets.

Once the data is made available from the heterogeneous data sets, the problem becomes one of semantics, i.e how are the data related to one another. If the data were collected using the same terminology, then the relationships between them would be known. However, it will be considerable time before there is enough data collected across the health system to make this a simple problem to address. There is then the problem of including new or more specific clinical descriptions to the existing terminology.

To address the issues of non-standard data and non-existent concepts, we have implemented a mapping tool to allow for existing data sets to be mapped onto the standard ontology. Since the largest standard clinical ontology is SNOMED CT™, we are using this as the basis for our extended clinical ontology. We take advantage of the extension mechanisms provided by SNOMED CT™ to publish local extensions based on the data in our linked data sets. While many extensions are built from existing terminologies or from scratch, extensions build from existing data sets may provide a faster way to obtain more complete coverage of the concepts used to capture clinical information. Previous studies have shown (Wade 2006) that mapping existing data to SNOMED CT™ is possible, although with incomplete coverage and consequently many current efforts are focused on building extensions for clinical concepts and areas which are not already covered.

The mapping tool we have built is described in detail elsewhere (Lawley, 2008). In short it provides a software environment to specify if a data item in an existing database is equivalent, or a variant of, an existing concept in the ontology. An expression editor allows post-coordinated expression to be built, to allow for complex combinations of existing concepts to be built. To check for consistency of this ad-hoc extension, we use an in-house classification engine, "snorocket". Snorocket (Lawley, 2007) is a new, fast, implementation of the Dresden algorithm (Baader, 2007) for classifying ontologies.

With HDI providing a link between a data item and a patient, and the mapping tool creates a link between a data item in the database and the clinical ontology, it is then possible to query across the data as if it was collected using a single clinical terminology. This allows for complex queries to be asked which take full advantage of the relationships encoded in the ontology.
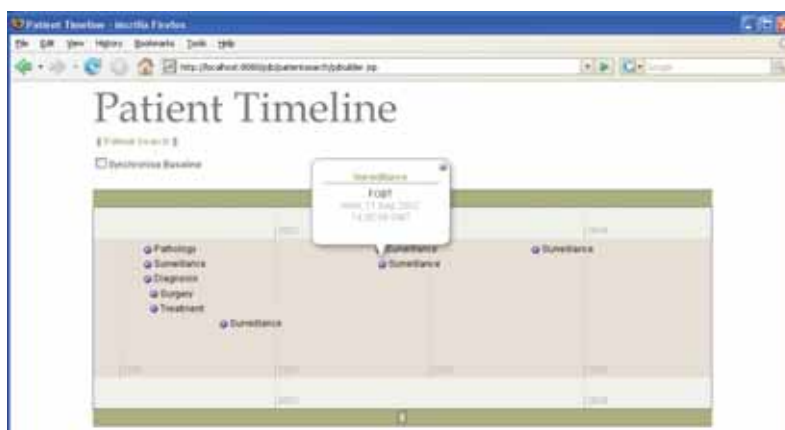

Figure 2. *A patient timeline for patient's diagnosis, treatment and follow-up for colorectal cancer.*

One such set of complex questions might revolve around the progression of a disease in a certain cohort of patients. With the growth of chronic diseases, understanding the progression of disease in a particular patient will become increasingly important. One way of viewing this progression is through a timeline of "health events" which display data from multiple databases against a timeline of when the health event which the data describes occurred. We have used the semantic layer provided by mapping existing data to a single clinical ontology to understand how different events in the patient data are related. In the case of colorectal cancer, we have modeled an interaction with the health service as a surveillance event, diagnostic event or for primary or secondary treatment. Figure 2 shows the events which have been captured for a single patient, in a timeline. Each event in the patient record is shown in summary in the timeline, but the full original record is available by drilling down through each of the events to the original data source.

Results:

A simple example of a patient data access system using the above concepts is now presented. In our example, we use synthetic data about patients receiving clinical treatment for the diagnosis and treatment of colorectal cancer (CRC), which like many cancers is seen as a chronic disease requiring interventions over a number of years. We use the HDI software (Hansen, 2007) to access the distributed data from six different databases. Our mapping tool is then used to map the data items in the linked data set to the existing ontology and an ad-hoc extension is created. We then use the relationships encoded in the extended ontology in querying the data. These queries take the form of finding events in the patient data which are linked to diagnosis and treatment of CRC. Example queries include: "find me all patient events which have a code of 73761001 or one of its children". This will find events which map to a colonoscopy (73761001) or one of its children (for example a sigmoidopscopy, 24420007). This we map to a "surveillance event" in the patient timeline. We build similar queries for other types of patient events, such as surgery or follow up treatment. The events are then presented in the patient timeline.

Discussion:

In this paper we have discussed some of the tools which we are currently building which will enable existing data sets to make use of current developments in data collection. While many information systems are now collecting data using standard terminologies it will still be considerable time until a majority of patient data is collected in this way. This does not then solve the issues of expanding medical knowledge requiring new expressions to fully capture their meaning. The system presented allows for legacy data and data representing new clinical knowledge to be used seamlessly with data collected using any current standard. Taking advantage of these developments will improve access to patient data, for clinicians, the health system and the patient.

References:

SNOMED Clinical Terms. College of American Pathologists, 2006. http://www.snomed.org.

G. Wade and S.T. Rosenbloom. 'Experiences Mapping a Legacy Interface Terminology to SNOMED CT' In Proceedings of the SMCS 2006 - Semantic Mining Conference on SNOMED CT, 2006. http://www.hiww.org/smcs2006/proceedings/9WadeSMCS2006final.pdf

Dean F. Sittig, Adam Wright, Jerome A. Osheroff, Blackford Middleton, Jonathan M. Teich Joan S. Ash, Emily Campbell, David W. Bates. 'Grand Challenges in Clinical Support', Journal of Biomedical Informatics, 2007 (in press)

M.J. Lawley, 'Exploiting Fast Classification of SNOMED CT for Query and Integration of Health Data', KRMed 2008, Pheonix, Arizona

F. Baader, C. Lutz, and B. Suntisrivaraporn. 'CEL - a polynomial-time reasoner for life science ontologies' In U. Furbach and N. Shankar, editors, Proceedings of the 3rd International Joint Conference on Automated Reasoning (IJ-CAR'06), volume 4130 of Lecture Notes in Artificial Intelligence, pages 287{291. Springer-Verlag, 2006.

Hansen, D. P.; Pang, C. & Maeder, A. (2007), 'HDI: Integrating Health Data and Tools', Journal of Soft Computing 11(4), 361-367

J. Ryan-Brown, N. Keynes, K Harrap, D. Vickers and S. McBride, 'Opening a Closed System – Implementing a Remote Communication Mechanism over an Existing Product', Australian Software Engineering Conference, 2007.

J. Ryan-Brown and S. McBride, 'Integrating Custodial Controlled data for collaborative Research' HIC 2008, submitted.

M.J. Lawley, D. Vickers and D. Hansen, 'Converting ad hoc terminologies to SNOMED CT extensions', HIC 2008, submitted.