Importing Clinical Data into Electronic Health Records - Lessons Learnt from the First Australian GEHR Trials

Linda Bird[1], Andrew Goodchild[1] and Sam Heard[2]
[1] DSTC, Level 7, GP South, University of Queensland, St Lucia, Qld, 4072. Email bird@dstc.edu.au and andrewg@dstc.edu.au
[2] Director, General Practice Education and Research Unit, Northern Territory Clinical School, Flinders University, PO Box 41326, Casuarina, NT 0811. Email sam.heard@flinders.edu.au

Abstract

One of the many features of an Electronic Health Record (EHR) is that it can be used as an accumulator of clinical information from a number of other clinical applications. In this paper, we present the outcomes of two General Practice Computing Group (GPCG) funded trials (run under the "Trials of IM/IT clinical integration activities within the health sector" RFT). Both of these trials explored using an EHR as an accumulator of clinical information by investigating the process of exporting clinical information from existing applications into a 'Good Electronic Health Record' (GEHR) based repository. Once the clinical information was uploaded into the GEHR repository, the data could then be validated, viewed and queried. The two projects that will be discussed are:

- The OACIS project (formally titled "Hospital to GP communication between non-GEHR and GEHR-compliant systems"), which involved the transfer of approximately 15,500 microbiology, biochemistry, haematology and radiology test results collected over a period of time in the South Australian OACIS database into a GEHR-based EHR repository.
- The GP Software Integration project (formally titled "Shared diabetes care in general practice"), which involved the transfer of medication lists, therapeutic precautions, problem lists and events for diabetes patients collected over a period of time in Medical Director and Locum Script into a GEHR-based EHR repository.

Background:
Currently general and specialist practices are embracing a variety of practice health record systems, including Medical Director, Locum Script, Medical Spectrum and several other less well-known systems. In addition to this many hospitals are using quite different systems for managing their clinical data. Each of these systems varies in their underlying technology and approach for structuring health records. In order to be able to exchange information between these systems and not build a one-to-one gateway or interface engine between every vendor it is desirable that a single common approach be adopted for structuring health records.

There are many possible approaches to this problem, including using standards such as HL7, CEN 13606 and CorbaMED. However, a common problem exists with these approaches, which is that they do not provide a simple future-proof solution for standardizing the ever increasing variety of clinical information structures such as clinical tests, notes, care plans and lists such as allergies, medications, problems etc.

The Good Electronic Health Record (GEHR) is one approach, which addresses this issue. GEHR uses a two level modelling approach, where all information is described using a generic health record model that enables a wide variety of health information to be stored and then the structure of that information is further constrained by an archetype. An archetype is a constraint model that limits the structure of certain kinds of information such as clinical tests, notes, care plans, etc. Thus the archetype can be used to ensure that only data of a certain structure and hence quality can be added to the record. Furthermore, as the archetypes are decoupled from the underlying health record model, new archetypes can be added over time allowing a health record system to evolve without substantial changes.

Lessons Learnt:
The two trials discussed in this paper, namely the OACIS project (formally titled "Hospital to GP communication between non-GEHR and GEHR-compliant systems") and the GP Software Integration project (formally titled "Shared diabetes care in general practice"), explored GEHR as a possible approach for a common EHR model for data extracted from clinical systems. Both trials also explored the use of XML as a syntax for encoding and exchanging GEHR records and archetype definitions.

In performing these trials, we learnt a number of valuable lessons with respect to importing data from an existing clinical system into an EHR, including:

- The benefits of using GEHR.
- The benefits and difficulties of an XML based approach to extracting clinical data from an existing clinical system and importing it into a GEHR based Electronic Health Record;
- How to identify data quality issues that can arise in exported data and how these might be addressed; and
- How a formal approach can be used for modelling clinical archetypes to accept information extracted from existing systems.

The rest of this paper will discuss each of these topics in more detail.

Benefits of GEHR:
In performing both trials, we demonstrated that a GEHR based framework for exchanging clinical information is practical for two main reasons:

1. GEHR provides a uniform way of sharing electronic health records, which standardizes how the health record is structured and managed and what contextual information needs to be recorded. In the trials we demonstrated that data from existing GP desktop software and hospital software could be successfully converted to fit within the GEHR framework.
2. GEHR allows a large and evolving variety of clinical information to be recorded within it, by using a mechanism called archetypes. In the trials we demonstrated that the GEHR archetypes are a flexible approach to support a wide variety of clinical information for existing GP desktop software and hospital software.

XML-based Data Extraction:
In these trials we used XML as the primary format for exchanging EHR data and XML based tools, such as XSLT, for manipulating EHR data. The process used for transforming clinical data from a non-GEHR system into GEHR-format EHR data, can be summarised as follows:

1. Raw data is extracted directly from the source clinical system, such as OACIS or Medical Director/Locum;
2. This data is pre-processed to generate a generic XML format that corresponds directly to the extracted data;
3. In parallel, standard clinical models (called 'archetypes') are designed for the type of clinical data being processed;
4. Given the generic XML from the source clinical system (produced in Step 2), and a standard clinical archetype into which this data should be transformed (produced in Step 3), a mapping process is performed to define the relationships between the fields in each.
5. From the mapping process performed in Step 4, an XSLT script is generated which is able to automatically transform the generic XML data into GEHR-compliant data conforming to the specified clinical archetypes.
6. The resulting XML-formatted EHR data is then imported into a GEHR based EHR repository.

For the purposes of these projects, this process was demonstrated on a sample of pathology and diabetes care data – however, the process was specifically developed using a generic approach, to ensure that it could be applied to a diverse range of input clinical data.

The approach taken in these two projects appears to work quite well for handling most kinds of data. However, a number of issues, relating to how the transformation mapping between the source clinical data and the EHR formatted data is defined, were uncovered, including:

- In performing the mapping from clinical data to EHR data, the data sometimes seemed to be 'grouped' in different ways, which required specialised processing. For example, in the OACIS project's microbiology archetype, antibiotics are grouped by organism name (in the 'Culture' group'), but in the original OACIS data each antibiotic used was contained in a separate test result (leading to repeating organism-names).
- In the OACIS project, in particular, much of the information was hidden within the comment fields. It was decided that it was only viable to map those comments with a recognised comment type into specific fields of the EHR extract. The other comments were mapped into the general 'comments' field. Unfortunately, this meant that quite a bit of content that could easily be manually interpreted could not be individually identified and categorised. So while an automated procedure could have been developed to extract the individual components of the comment text, it was felt that this process was too liable for error due to the unpredictability and non-standardised nature of the data.
- Some test results recording in the data were repeated in multiple rows (due either to 'interim' test results becoming 'final', or to additional information being added to previously 'final' test results). In these cases, the rows were mapped to different 'versions' of the same test.
- For the purpose of the OACIS project, a term set mapping file, which defined the mappings from OACIS terms to coded term sets to be used in the GEHR extract, was created and defined for use within the translation script. However, it may be better for a number of reasons to directly reference a live term set server. This would provide more up-to-date term translations, and be more scalable for very large term sets.
- It was quickly discovered that the process of writing the XSLT translation scripts was undesirably labour intensive, and that a more efficient approach would be to automate the writing of the translation scripts based on the mappings between the source and target data structures.

It was also demonstrated in the two trials that a generic GEHR-XML to HTML transformation can be developed (using XSLT) to allow the EHR formatted data to be viewed through a standard web browser. The advantage of this approach is that any GEHR-compliant EHR extract can be displayed for the user in a presentable way, even if the clinical archetypes used have not previously been seen. Clinical archetypes, which the viewer has knowledge of, however, can be displayed in a more specialised manner if preferred. It was also concluded that because a significant amount of clinical information was bound within textual fields (such as 'comments'), which could not necessarily be separated into individual, structured fields, it would be advisable that a text-indexing service be used over the resulting EHR data to enable more efficient searching.

Data Quality Issues:

One of the biggest issues faced with transforming data into a common EHR format is the inconsistency of the structure and content in the current data being collected for clinical purposes. This was evident in a number of ways:

- Some of the extracted data did not include all mandatory data fields. For example, Medical Director was missing 12 fields and Locum was missing 31 fields that are mandatory in the archetypes. There were two main lessons learnt from this. Firstly, that mandatory fields in archetypes should be kept to an absolute minimum; and secondly that there needs to be better coordination with vendors to allow more complete data to be collected.

- Some rows contained data placed in the incorrect fields.
- Some fields contain data in a single field that should be spread over two or more fields and required separation before further processing was possible.
- Some data is stored in an encrypted form. For example, the progress notes in Medical Director are encrypted.
- HL7 messages feeding the OACIS system have often used a generic comment field to place key data.
- Different rows of data (each representing a 'batch' of test results) were structured differently, depending on the type of tests performed. However, determining the type of test performed (and therefore the structure of the row) involved mapping the valid list of possible 'mater-panel' names to a row type (in this case 'microbiology', 'biochemistry' or 'radiology') and therefore to a row structure. This was complicated by the fact that the 'master-panel' name appeared in a different field position in different row types.

The data quality issues that are evident from such work draw attention to the safety aspects of transformation of data to a new format. While some of the errors appeared consistent and could be 'massaged' to preserve their information content, this was thought to be unsafe and inappropriate. The archetypes developed for the trial had comment fields that could be used to contain the unstructured text – preserving clinical 'readability' and not compromising safety.

It would be possible to get more information from the database if a thorough study of the extent and nature of data errors was studied and the safety issues of 'adhoc' transformation investigated.

Designing Archetypes to Accommodate Existing Clinical Data:
In trying to capture clinical information from an existing information system we modelled the information using archetypes. In doing so, we discovered that naively modelling archetypes, by copying the structure that exists in information systems, leads to archetypes that are not reusable. By accessing clinical standards in use such as HL7 messages, LOINC, the National Health Data Dictionary and the GP Minimum Dataset a better solution became available.

The problem of 'finding' archetypes is clear and is predominantly a clinical one. In this project we took an approach to finding archetypes, which was based on a number of basic archetype modelling principles, such as:

- Archetypes should be generalized as far as possible to allow multiple types of clinical information to be conveyed. For example, in the OACIS trials we constructed a generic biochemistry archetype, which can be used for the vast majority of biochemical test results. These general archetypes can then be specialized into a number of specific situations such as lipid studies and other common batteries.
- Archetypes for physical examinations should be based on the approach to measurement rather than parts of a body system that are to be examined together or have anatomical proximity – which becomes an issue of how the examination findings are organised. For example, the testing of sensation is common in most parts of the body – which may be an arm or leg or a dermatome. This may be performed as part of a neurological system examination, the examination of a limb or as part of the assessment of a laceration in another part of the body.

In addition to these results, we found that medical terminologies, such as LOINC, which are highly pre-coordinated, are not very suitable for use in the archetypes developed. For a number of reasons, terminologies, which are not highly pre-coordinated, such as UMLS and GALEN, are more suitable. Finally, some common tests (e.g. Glucose Tolerance Test) still pose a challenge to the current GEHR model and require specific archetyping. These issues are being addressed in the current openEHR model under development in conjunction with CEN.

Conclusions:

The importing of large clinical datasets into a GEHR-based Health record, such as was done in the two trials referred to in this paper, is the first evidence we have that our approach to transforming EHR data can deal with real data of a reasonably diverse nature. This paper concludes, however, that it is very important that the issues uncovered and lessons learnt during this trial are fed into future work in this area.

References:

1 The Good Electronic Health Record Website: http://www.gehr.org/

2 *OACIS-GEHR Transformation Process*. A report to the GPCG. Available at: http://www.gpcg.org/publications/docs/projects2001/GPCG_Project2_01.pdf

3 *Shared Diabetes in General Practice*. Available at: http://www.gpcg.org/publications/docs/projects2001/GPCG_Project5_01.PDF

Acknowledgements:

The work reported in this paper has been funded by a number of organizations, including the General Practice Computing Group and the Cooperative Research Centres Program through the Department of the Prime Minister and Cabinet of the Commonwealth Government of Australia. In addition to these organizations, we would like to thank the South Australian Health Commission and HCN for providing assistance in these projects. Finally, we would like to thank Zar Zar Tun (DSTC) and Kris Evans (DSTC) for their time in implementing the trials.