



GPCG funded IM/IT Projects 2001

The General Practice Computing Group and the Department of Health and Ageing selected a number of general practice informatics projects for funding in 2001. The following document is a report from one of these projects. All reports can be found on the GPCG Website at <http://www.gpcg.org/projects>

Disclaimer

The General Practice Computing Group (GPCG) presents the report of the following project for the purpose of disseminating GP informatics information free of charge for the benefit of stakeholders and the public. It has been prepared by a third party and does not necessarily represent the views of the GPCG or the Department of Health and Ageing (DoHA). DoHA holds intellectual property rights over this document.

All reports are covered by the GPCG disclaimer, which can be found at <http://www.gpcg.org/disclaimer>

FINAL REPORT: THE OACIS-GEHR TRANSFORMATION PROCESS

DSTC, Flinders University et al.

**Performed for the project entitled:
Hospital to GP communication between non-GEHR and GEHR-compliant systems**

RFT 31/0001 Project No: GPCG 2000/01
RFT Project Title: "Trials of IM/IT clinical integration activities within the health sector"

Applicant details

Flinders University of South Australia
BEDFORD PARK SA 5042
Through Flinders Consulting Pty Ltd
ACN: 058 894 456
ABN: 70 058 894 456

Contact details

Dr Sam Heard
Director, GP Education and Research Unit
NT Clinical School, Flinders University
PO Box 41326
Casuarina, NT 0811

Tel: 08 8922 7939
Fax: 08 8922 7928
Mob: 0417 838 808
Email: sam.heard@flinders.edu.au

Table of Contents

1. Introduction.....	4
2. Overview of Transformation Process	5
3. Data Extraction from OACIS	7
3.1 Extracting the Raw Data	
3.2 A Description of the Extracted Data	
3.3 Problems and Issues	
4. Pre-processing the OACIS Data.....	11
4.1 Overview of Pre-processing Process	
4.2 Generic XML Format	
4.3 Problems and Issues	
5. Defining the GEHR Archetypes.....	19
5.1 Overview	
5.2 Terminology	
5.3 Archetype Specialisation	
5.4 GEHR Archetype Descriptions	
5.5 GEHR Archetype Editor	
5.6 Problems and Issues	
6. Mapping OACIS Data into Archetypes.....	37
6.1 Overview	
6.2 Mapping Definition	
6.3 Problems and Issues	
7. Using Mapping to Translate XML Data.....	53
7.1 Overview of Process	
7.2 XSLT	
7.3 GEHR-compliant XML	
7.4 Problems and Issues	
8. Importing and Viewing GEHR-XML	66
8.1 Overview of Process	
8.2 Viewing the Data	
8.3 Problems and Issues	
9. Conclusions.....	72

1. Introduction

The project is addressing the problem of transforming data from current clinical systems to an EHR format - specifically the Good Electronic Health Record format. The trial involves the transfer of clinical data collected over a period of time in the South Australian OACIS database to the GEHR kernel, to enable generic access and safe automatic processing. This is being performed through an intermediary XML format to enable a generic approach suitable for application in other situations.

The data to be transformed is representative of the information held in the OACIS system and involves pathology and radiology results. These are the results most commonly required for inclusion in a discharge referral. The formation of GEHR extracts in XML format will preserve the medico-legal information and maintain the context of reports and any updates. A set of archetypes suitable for accepting the data into the GEHR kernel has been developed, in line with HL7 and LOINC, the National Health Data Dictionary, and the GP Minimum Dataset.

All participants have been active in the project - Flinders University, the DSTC, Ocean Informatics and the SA Department of Human Services. The project is led by Flinders Consulting Pty Ltd, the commercial agency of Flinders University.

The objectives of the project were:

- to develop and document a methodology for transforming clinical data from a non-GEHR hospital clinical system to GEHR format data suitable for use in GEHR-compliant clinical applications used in general practice;
- to demonstrate the transformation of common pathology and radiology data to a form suitable for use in a patient discharge referral for GPs from the South Australian OACIS hospital system;
- to investigate the requirements for transforming all the data in the OACIS hospital system; and
- to ensure that the discharge referral data contains all the data required by Standards Australia IT/14/6/6 HL7 discharge referral draft standard.

The OACIS database in fact only holds laboratory results and radiology reports. To make these accessible in general practice, transformation into GEHR transactions preserves the ability to access these reports in the future from other environments. The archetypes developed to store this information are consistent with HL7 standards and are designed to cope with any data utilising the LOINC code set.

2. Overview of Transformation Process

The process of transforming clinical data from a non-GEHR hospital clinical system to GEHR format data suitable for use in GEHR-compliant clinical applications used in hospitals or general practice, is summarised in *Figure 1*.

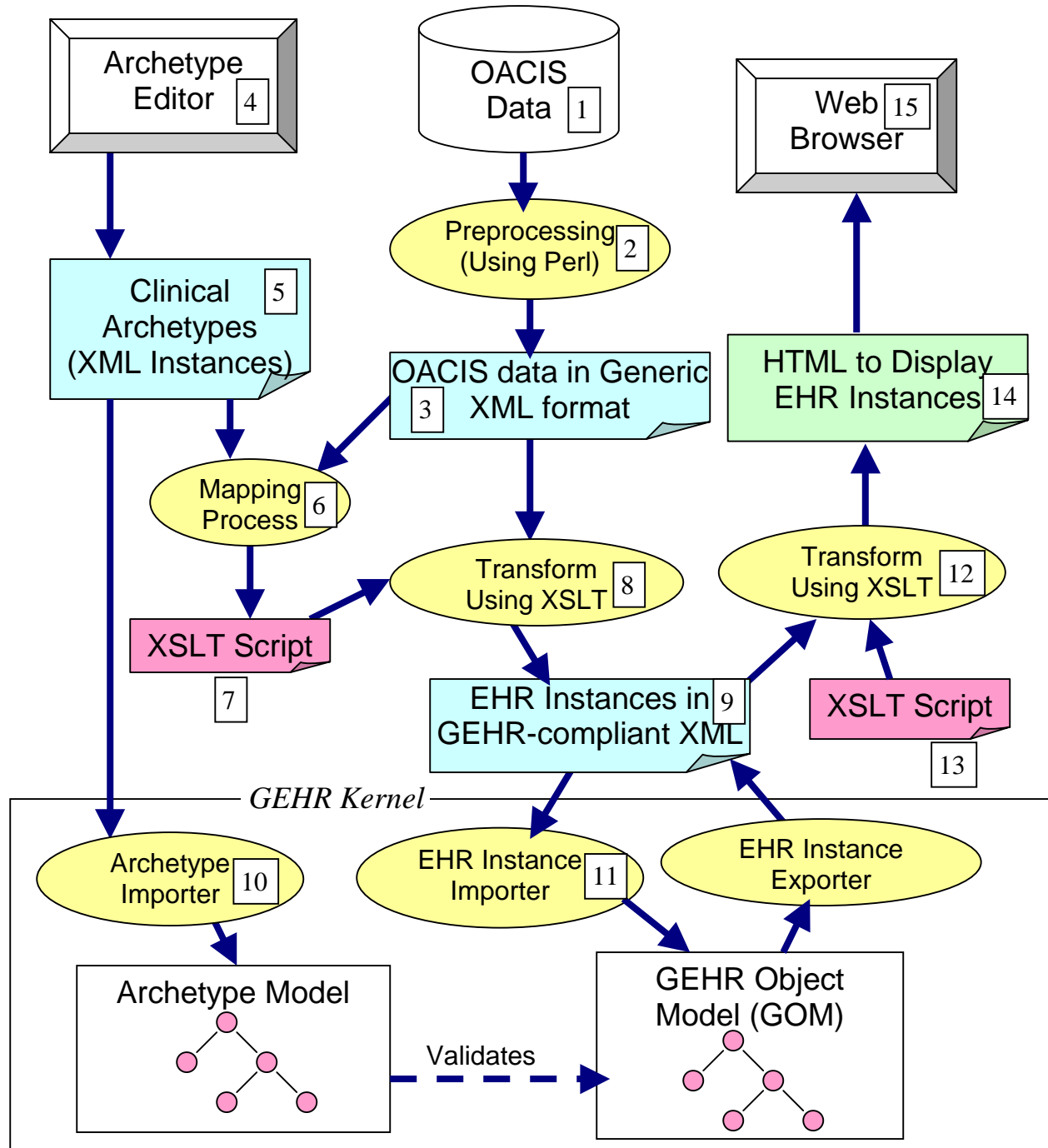


Figure 1: Overview of OACIS-GEHR Transformation

In particular, this diagram shows the main components used to transform the data extracted from the OACIS clinical system into the GEHR kernel.

In this process, raw data, which is extracted directly from the OACIS system (1), is pre-processed using the Perl language (2), to generate a generic XML format that corresponds directly to the extracted data (3). In parallel, standard clinical models (archetypes) are designed for the type of clinical data being processed. These clinical models are edited in an archetype editor (4), which can generate a formal XML definition of the model (5).

Given the generic XML from the OACIS system, and a standard clinical archetype into which this data should be transformed, a mapping process is performed to define the relationships between the OACIS data and the fields in the target clinical archetype (6). From this mapping process, an XSLT script (7) is generated which is able to automatically transform (8) the generic XML data into GEHR-compliant data conforming to the specified clinical archetypes (9).

After importing and initialising the required archetype definitions (10), the GEHR-compliant XML data can be imported and stored in the GEHR kernel (11), for use within a clinical application. To view the data stored in the GEHR kernel, a generic GEHR-XML to HTML transformation (12) is performed (using an XSLT script (13)) and the HTML formatted data (14) is viewed through a standard web browser (15).

For the purposes of this project, this process has been demonstrated on a sample of biochemistry, microbiology, haematology and radiology data from the OACIS system. A more detailed description of the process is included in subsequent sections.

3. Data Extraction from OACIS

3.1 *Extracting the Raw Data*

Approximately 15,500 rows of data were extracted from the OACIS system, covering biochemistry, haematology, microbiology and radiology test results. Each row of extracted data represented one 'batch' of test results.

This data extraction was performed by consultants, who work directly with the OACIS system. The extraction process involved running a batch query to aggregate the data for each suite of test results into a single, comma-delimited row of data fields. In doing so, the internal data structures of the OACIS system were kept private, while the data content was made available for the project. The patient details were altered so that no patients could be identified.

3.2 *A Description of the Extracted Data*

Fifteen Thousand five hundred records were extracted and processed to GEHR format. The break up of the records in the data set was as follows:

Total number of records:	15545
Radiology type records:	1370
Radiology Examination records:	1370
Biochemistry type records:	9839
GENERAL BIOCHEMISTRY records:	1963
COMPLETE BLOOD EXAM records:	343
BIOCHEMICAL ANALYSIS records:	7533
Microbiology type records:	4187
MICROBIOLOGY - CSF' records:	5
FUNGAL MICROBIOLOGY records:	53
URINE - MICROBIOLOGY records:	3625
MICROBIOLOGY records:	504
Other records: (not included)	149
'SYPHILIS SEROLOGY' records:	146
Records without master panel names:	3

In the following subsections, the comma-delimited format in which the raw data was received is described.

3.2.1 Biochemistry and Haematology

Biochemistry and Haematology data were both extracted from OACIS system in the same format. General information about both the patient and the batch of tests appears first, followed by a set of test results, followed by a textual comment.

The first ten data fields of each row have the following meaning:

Field Number	Value
1	LastName^FirstName
2	MRN (Master Record Number)
3	Birth date
4	Gender
5	Collection date time
6	Requested date time
7	Received date time
8	Laboratory Number
9	Laboratory Name
10	Master panel name eg BIOCHEMISTRY

This is followed by any number of test results, for example “Sodium^141^mmol/L^F^137^145” where:

Component	Value in example
Name of result	Sodium
Result value	141
Units	mmol/L
Result Status	F
Normal Low	137
Normal High	145

The test results are followed by a comment, in the format “COMMENT^comment text”. Note that the units for the Normal and High range are assumed to be the same as the result value.

For example, a complete biochemistry row extracted from OACIS looks like:

Duncan^Paul,001054,3/08/1923,M,10/06/1997,10/06/1997,10/06/1997 09:15:00,
2193686CHEM,RAH-LAB,BIOCHEMICAL ANALYSIS,Sodium^141^mmol/L^F^137^
145, Potassium^4.1^mmol/L^ F^3.1^4.2,Chloride^93^mmol/L^F^100^109,Bicarbonate^26^
mmol/L^F^22^32,Anion Gap^ 26^mmol/L^F^7^17,Urea^14.7^mmol/L^F^2.7^7.2,
Creatinine^0.469^mmol/L^F^0.050^0.120,Urea/Creat^31^^F^35^80,COMMENT^PLASMA

3.2.2 Microbiology

Microbiology data was extracted from the OACIS system in a similar format, with general information appearing first, followed by one or more organism results, followed by one or more structured comments.

The first ten fields of an extracted, microbiology row were structured in the same way as that for biochemistry: namely as “LastName^FirstName, MRN, Birth Date, Gender, Collection

date time, Requested date time, Received date time, Laboratory Number, Laboratory Name, Master Panel Name”.

The first ten fields are followed by any number of result data fields, for example “Pseudomonas aeruginosa^F^Amikacin^R^PS” where:

Component	Value in example
Organism Name	Pseudomonas aeruginosa
Organism result status	F
Antibiotic	Amikacin
Sensitivity	R
Sensitivity Status	PS

Note that the organism is included with each sensitivity result as it is possible to have multiple organisms and each sensitivity refers to only one specific organism.

This is followed by a comment in the format COMMENT, “Comment Type”^“Comment Text” where “Comment Type”^“Comment Text” is a repeating segment. The comment type can be thought of as a heading.

For example a complete microbiology row extracted from OACIS looks like:

```
Bloggs^Arthur,1/04/1960,2/08/2000 14:00:00,2/08/2000 14:00:00,2/08/2000
14:00:00,GENERAL MICROBIOLOGY, Pseudomonas aeruginosa^F^Imipenem^S^PS,
Pseudomonas aeruginosa^F^Aztreonam^S^PS, Pseudomonas aeruginosa^F^Ticar/Clav^R
^PS, Pseudomonas aeruginosa^F^Ceftazidime^S^F, Pseudomonas aeruginosa^F^
Meropenem^S^F, Pseudomonas aeruginosa^F^Amikacin^R^PS, Pseudomonas aeruginosa^
F^Ciprofloxacin^S^F, Pseudomonas aeruginosa^F^Piperacillin^R^F, Pseudomonas
aeruginosa^F^Tobramycin^R^F, Pseudomonas aeruginosa^F^Gentamicin^R^F,
COMMENT,"SPECIMEN"^^"PUS SWAB", "Comment"^^", "Cells"^^"+++ polymorphs +
epithelial cells", "ORGANISMS"^^"+++ Gram positive cocci", "Negative Culture"^^",
"Comment"^^", "Growth"^^"++"
```

3.2.3 Radiology

Once again, the Radiology data extracted from the OACIS system follows a similar format, with general information appearing first, followed by one or more structured comments.

The first nine fields of an extracted, radiology row were structured in the same way as that for biochemistry and microbiology, except that “Received date time” did not appear. Therefore the first nine fields have the following meaning: “LastName^FirstName, MRN, Birth Date, Gender, Collection date time, Requested date time, Laboratory Number, Laboratory Name, Master Panel Name”

These nine fields are followed by a comment in the format COMMENT, “Comment Type”^“Comment Text” where “Comment Type”^“Comment Text” is a repeating segment. For example a complete radiology row extracted from OACIS looks like:

Duncan^Paul,001054,3/08/1923,M,12/05/1997 14:37:00,12/05/1997,97-027728,RAH-RAD,Radiology Examination,COMMENT ,""^"" ,""^""Reported by WALLACE,M" ,""^""Verified 12/05/1997 16:48" ,""^"" ,""^""~CLINICAL DETAILS:~ Day 4 post inguinal hernia repair. CRF - polyarteritis" ,""^""nodosa. Dry cough febrile." ,""^"" ,""^""CHEST 27728 (20): There is slight cardiomegaly with bilateral mid zone" ,""^""fibrosis. Overall appearances are similiar to those of 7/5/1997."

3.3 Problems and Issues

The ‘form’, in which the project team chose for the data to be extracted, was a separate record for each client in a text format. This approach suits moving to a ‘real time’ implementation in the future, which is a likely scenario in the government’s envisioned Health Connect environment. Alternative approaches may be more suitable for a data download on mass – with relational tables and preservation of ‘foreign keys’. This would maximise the ability to pre-process the data to preserve as much ‘information’ in the data as possible.

All copies of the messages received into the OACIS system are kept and so records may in fact form interim reports received from a laboratory. The ‘result status’ data item is designated as ‘F’ if it was the final report and ‘I’ if the interim report. Within GEHR, versioned transactions allows ‘over-writing’ of interim results with final results by a versioning process – interim results are then kept and the ability to restore an EHR to its state on any particular day is maintained. However, it was discovered that final results are occasionally updated within the OACIS system by further ‘final’ results. For this reason the Laboratory Name and Laboratory Number were checked as this unique number relates to a particular report. If a non-unique value was found then a new version of the result was saved.

One of the biggest issues faced with the data extracted from the OACIS system was the inconsistency of the data. For example, the content of the data tended to vary from lab to lab (and even within the same lab). Some labs structured the data using the format requested by the OACIS team, and others placed most of the content in the comment fields. Some labs put all comments in the comment text, and did not include comment types. The comma-delimited format was therefore used in a fairly non-standard way, to allow for variation in the structure and data content that was present. This made processing of the data slightly more complex.

We were also faced with the difficulty of commas appearing in the comment text (making it impossible to tell where a comment stopped and the next one started). To solve this problem, double quotation marks were put around each comment type and comment text. To allow for double quotations appearing in the text of a comment, two consecutive double quotes were used to represent a single character in the text. This solution allowed us to satisfactorily identify comment text in the extracted data.

4. Pre-processing the OACIS Data

4.1 Overview of Pre-processing Process

Once the raw, comma-delimited data was extracted from the OACIS system, this data was put through a pre-processing process to massage the data into an XML format, which could be manipulated using an XSLT processor. This pre-processing was performed using the Perl language, due to its simplicity for basic data processing. The generic XML format generated in this pre-processing step corresponds directly to the data extracted from the OACIS system.

In a live transformation wrapper for OACIS to GEHR, it may be possible to merge the extraction and pre-processing steps into one process. However, this was not possible in the project, as it was necessarily for these two steps to be performed by different groups of people.

4.2 Generic XML Format

4.2.1 Biochemistry and Haematology

Given the biochemistry or haematology data row extracted directly from the OACIS system:

```
Duncan^Paul,001054,3/08/1923,M,10/06/1997,10/06/1997,10/06/1997 9:15:00,
2193686CHEM,RAH-LAB,BIOCHEMICAL ANALYSIS, Sodium^143^mmol/L^F^
137^145, Potassium^4.3^mmol/L^F^3.1^4.2, Glucose^4.4^mmol/L Fasting^F^3.8^5.5,
COMMENT^ Plasma
```

the generic XML generated in the pre-processing step would look like:

```
<record>
  <row-type>B</row-type>
  <name> <lastname>Duncan</lastname> <firstname>Paul</firstname> </name>
  <MRN>001054</MRN>
  <dob>3/08/1923</dob>
  <sex>M</sex>
  <collection-time>10/06/1997</collection-time>
  <requested-time>10/06/1997</requested-time>
  <received-time>10/06/1997 9:15:00</received-time>
  <lab-number>2193686CHEM</lab-number>
  <lab-name>RAH-LAB</lab-name>
  <master-panel>BIOCHEMICAL ANALYSIS</master-panel>
  <data>
    <result-name>Sodium</result-name>
    <result-value>143</result-value>
    <units>mmol/L</units>
    <status>F</status>
    <normal-low>137</normal-low>
    <normal-high>145</normal-high>
```

```

</data>
<data>
  <result-name>Potassium</result-name>
  <result-value>4.3</result-value>
  <units>mmol/L</units>
  <status>F</status>
  <normal-low>3.1</normal-low>
  <normal-high>4.2</normal-high>
</data>
  <result-name>Glucose</result-name>
  <result-value>4.4</result-value>
  <units>mmol/L</units>
  <action>Fasting</action>
  <status>F</status>
  <normal-low>3.8</normal-low>
  <normal-high>5.5</normal-high>
</data>
  <comment>Plasma</comment>
</record>

```

Note that the units field from the raw data (e.g. “mmol/L Fasting”) is decomposed into a units field (“mmol/L”) and a challenge action field (“Fasting”) in the pre-processing stage, in order to separate the individual components of data required in the future transformations, and therefore simplifying future processing. Also note that a row-type field was added to the beginning of the row to provide easy identification of the subsequent structure of the row. As discussed in Section 4.3 (Problems and Issues) this improves the genericity of the process.

Each piece of data can henceforth be referenced using its XML path. For example the patient’s firstname (Paul) could be referenced as “/record/name/firstname”.

4.2.2 Microbiology

Given the microbiology data row extracted directly from the OACIS system:

```

Baxter^Glen,001197,7/08/1934,M,13/10/1998,13/10/1998,13/10/1998 16:49:00,
MB187130UR, QEH-LAB, URINE - MICROBIOLOGY, EC^F^AMOXYCILLIN^S^F, EC^
F^AUGMENTIN^S^SR, EC^F^CEPHALOTHIN/CEPHALEXIN^S^F, , COMMENT
,"SPECIMEN SITE"^"Midstream", "WHITE BLOOD CELLS (N<10)"^">1000", "ANTI-
MICROBIAL AGENTS"^"Detected", "signature : "^"VANIK
DERHARROUTIAN", "DIPSTIX"^"D:D1/1+ D:D2/2+", "CLINICAL&LAB"^"RENAL TX"

```

the generic XML generated in the pre-processing step would look like:

```

<record>
  <row-type>M</row-type>
  <name><lastname>Baxter</lastname><firstname>Glen</firstname></name>
  <MRN>001197</MRN>
  <dob>7/08/1934</dob>

```

```

<sex>M</sex>
<collection-time>13/10/1998</collection-time>
<requested-time>13/10/1998</requested-time>
<received-time>13/10/1998 16:49:0013/10/1998 16:49:00</received-time>
<lab-number>MB187130UR</lab-number>
<lab-ID>QEH-LAB</lab-ID>
<master-panel>URINE - MICROBIOLOGY</master-panel>
<data>
  <organism-name>EC</organism-name>
  <organism-result-status>F</organism-result-status>
  <antibiotic>AMOXYCILLIN</antibiotic>
  <sensitivity>S</sensitivity>
  <sensitivity-status>F</sensitivity-status>
</data>
<data>
  <organism-name>EC</organism-name>
  <organism-result-status>F</organism-result-status>
  <antibiotic>AUGMENTIN</antibiotic>
  <sensitivity>S</sensitivity>
  <sensitivity-status>SR</sensitivity-status>
</data>
<data>
  <organism-name>EC</organism-name>
  <organism-result-status>F</organism-result-status>
  <antibiotic>CEPHALOTHIN/CEPHALEXIN</antibiotic>
  <sensitivity>S</sensitivity>
  <sensitivity-status>F</sensitivity-status>
</data>
<comment>
  <field type="SPECIMEN SITE">Midstream</field>
  <field type="WHITE BLOOD CELLS (N<lt;10)">>1000</field>
  <field type="ANTI-MICROBIAL AGENTS">Detected</field>
  <field type="signature :">VANI<sup>K</sup> DERHARROUTIAN</field>
  <field type="DIPSTIX">D:D1/1+ D:D2/2+</field>
  <field type="CLINICAL&LAB">RENAL TX</field></comment>
</record>

```

Note that, once again, a row-type field was added to the beginning of the row to identify the structure of the row to follow, and improve the genericity of the process (as discussed in Section 4.3 – Problems and Issues). Also note that the individual comments were represented as “fields” within a “comment”, with the comment type as an attribute of field, and the comment text as the body of the field element.

4.2.3 Radiology

Given the radiology data row extracted directly from the OACIS system:

Duncan^Paul,001054,3/08/1923,M,12/05/1997 14:37:00,12/05/1997,97-027728,RAH-RAD,Radiology Examination,COMMENT ,""^"" ,""^""Reported by WALLACE,M",""^""Verified 12/05/1997 16:48",""^"" ,""^""~CLINICAL DETAILS:~ Day 4 post inguinal hernia repair. CRF - polyarteritis",""^""nodosa. Dry cough febrile.",""^"" ,""^""CHEST 27728 (20): There is slight cardiomegaly with bilateral mid zone",""^""fibrosis. Overall appearances are similiar to those of 7/5/1997."

the generic XML generated in the pre-processing step would look like:

```
<record>
  <row-type>R</row-type>
  <name><lastname>Duncan</lastname><firstname>Paul</firstname></name>
  <MRN>001054</MRN>
  <dob>3/08/1923</dob>
  <sex>M</sex>
  <collection-time>12/05/1997 14:37:00</collection-time>
  <requested-time>12/05/1997</requested-time>
  <lab-number>97-027728</lab-number>
  <lab-ID> RAH-RAD </lab-ID>
  <master-panel> Radiology Examination </master-panel>
  <comment>
    <field type="Reported_by">WALLACE,M</field>
    <field type="Verified">12/05/1997 16:48</field>
    <field>~CLINICAL DETAILS:~ Day 4 post inguinal hernia repair. CRF -
      polyarteritis </field>
    <field>nodosa. Dry cough febrile.</field>
    <field>CHEST 27728 (20): There is slight cardiomegaly with bilateral mid zone
      </field>
    <field>fibrosis. Overall appearances are similiar to those of 7/5/1997.</field>
  </comment>
</record>
```

Note that it was decided that the “Reported by” and “Verified” comments from “RAH-RAD” laboratory had a sufficiently standard format to allow the type of these comments to be extracted from the comment text. For example, the comment:

```
""^"" ,""^""Reported by WALLACE,M"
```

was pre-processed into:

```
<field type="Reported_by">WALLACE,M</field>
```

and the comment:

```
""^""Verified 12/05/1997 16:48"
```

was pre-processed into:

```
<field type="Verified">12/05/1997 16:48</field>
```

By extracting the type information in this way, these comment fields can be mapped into specific GEHR fields, rather than into general comments, in the future mapping process.

4.3 Problems and Issues

A number of difficulties arose in this step – most of them stemming from the unreliability of the structure and content of the data. In order to prepare the data for XSLT processing in the following step, it was important for us to separate and identify the individual components of data during this stage.

Fields Requiring Separation

For example, in the biochemistry data if fasting was required before the test, then the keyword “Fasting” was appended to the Units fields of the raw data (e.g. “mmol/L Fasting”). To allow for future transformations using XSLT, it was important to identify and separate these two distinct pieces of information during this pre-processing step. A units field of “mmol/L Fasting” in the input data would therefore be transformed into “<units>mmol/L </units><action>Fasting</action>” during the pre-processing into generic XML.

Identifying the Type and Structure of Each Row

Another important function of the pre-processing step was to simplify the identification of the type of each row (and therefore the expected structure of the row). This was required because Biochemistry, Microbiology and Radiology data rows were intermingled within the same data extraction file, even though each had a slightly different structure, and therefore needed to be processed differently. To make the transformation process as generic as possible, we therefore introduced a row-type to the beginning of each row. This row-type identified for the processor, which structure to expect in the upcoming row, and therefore identifies which method it should use to process that row. In this project, we used the four row-type codes “B”, “M”, “R” and “O” to represent the row-types “Biochemistry”, “Microbiology”, “Radiology” and “Other”. However, in the more general case, a new row type should be introduced for each major category of data, which requires a different data format, or different processing rules. This allows the same overall process to be used irrespective of the structure or type of data being extracted from OACIS.

In order to add a row-type to the beginning of each row, we first needed to define a rule for identifying the row-type of each row. The rule, which was agreed upon was:

```
IF /record/master-panel in {"BIOCHEMICAL ANALYSIS", "GENERAL
BIOCHEMISTRY", "COMPLETE BLOOD EXAM", "FULL BLOOD EXAM"}
THEN row-type = "B"
```

```
ELSIF /record/master-panel in {"MICROBIOLOGY", "URINE -
MICROBIOLOGY", "MICROBIOLOGY - CSF", "FUNGAL MICROBIOLOGY"}
THEN row-type = "M"
```

```
ELSIF /record/master-panel in {"Radiology", "Radiology Examination"}
THEN row-type = "R"
```


ELSE row-type = “O”

This rule, however, had two important prerequisites. Firstly, that we needed to be able to identify the master-panel of each row; and secondly we needed to be able to identify the set of all master-panels, and manually classify them appropriately according to row-type.

Identifying the master-panel of each row would have been simple if the master-panel was always in the same field position in the raw data. However, in the microbiology and biochemistry data rows the master-panel was in the tenth position, and in the radiology data rows the master-panel was in the ninth position. Therefore both the ninth and tenth position had to be checked for the keywords “BIOCHEMICAL ANALYSIS”, “GENERAL BIOCHEMISTRY”, “COMPLETE BLOOD EXAM”, “FULL BLOOD EXAM”, “MICROBIOLOGY”, “URINE – MICROBIOLOGY”, “MICROBIOLOGY - CSF”, “FUNGAL MICROBIOLOGY” and the ninth position had to be checked for the keyword “RADIOLOGY”. If none of these keywords were found in their respective positions, then the row-type of “O” (or “Other”) was chosen.

Secondly, each valid master-panel needed to be classified into a row-type. It was agreed that a live file should be kept of mappings from master-panel to row-types, so that master-panels that hadn’t previously been encountered could be manually classified at run-time, without requiring any changes to the code. This method of defining a run-time configurable row-type mapping rule was adopted.

Data Not Where Expected ... or Not There at All

Another problem arose with data either not being where it was expected to be, or mandatory data not being present at all. For example, there were a couple of rows, which did not have a master-panel name – even though this field was mandatory. The decision was that these rows would not be processed, as they obviously contained ‘dirty’ or incomplete data. Instead, they produced an error that was recorded in the processing log file.

A number of other cases arose with the microbiology data. In particular, the result fields (namely ‘organism-name’, ‘organism-result-status’, ‘antibiotic’, ‘sensitivity’ and ‘sensitivity-status’) in many cases did not contain any data. In some cases, this data was hidden within the comment fields, in a way that was difficult to extract; and in other cases it was absent altogether. As a result, we processed the information that was there - but obviously we could do nothing with data that was not there.

In another case, the microbiology result data included the string “Escherichia coli^ F^Specimen Type:^S^F”. In this case the text “Specimen Type:” appears in the position that we would expect an antibiotic name. However, as this is clearly not an antibiotic name, we obviously have some “dirty” or incorrect data in some rows.

Lastly, there was an issue with how much content appeared in the comment fields, usually in a very unstructured and ad-hoc way. Some labs (e.g. “QEH-LAB”) tended to use the comment-type and comment-text fields appropriately, making it slightly easier to identify the keywords, representing the types of data content present. However, other labs (e.g. “RAH-LAB”) tended to not use the comment-type field, and instead put the whole comment into the

comment text fields – sometimes combining information into the one comment, and other-times separating common information into multiple fields. Because there was no predefined keywords or structure, it was virtually impossible to process some of the content-rich information present in these unstructured comment fields.

Pre-Processing the Microbiology Comment Fields

As previously mentioned, one of the functions of the pre-processor was to identify individual pieces of information that should be able to be processed by the XSLT processor in the upcoming step. Therefore, we needed to ascertain what information could sensibly and legally be extracted from the sometimes ad-hoc microbiology comment fields.

It was decided that the comment type would be represented as an XML attribute value for ease of future processing, and that the comment text would remain unprocessed due to its ad-hoc and inconsistent nature.

For example, the comment:

```
COMMENT ,"SPECIMEN SITE"^^"Midstream", "ANTI-MICROBIAL AGENTS"^^"Not
detected", "CULTURE"^^"<1000 organisms/ml", "signature :""^^"CAROL LANG(9.5)",
"DIPSTIX"^^"D:D0/0 D:D2/2+"
```

would be translated into:

```
<comment>
<field type="SPECIMEN SITE">Midstream</field >
< field type="ANTI-MICROBIAL AGENTS">Not detected</field >
< field type="CULTURE">&lt;1000 organisms/ml</field >
< field type="signature :">CAROL LANG(9.5)</field >
< field type="DIPSTIX">D:D0/0 D:D2/2+</field >
</comment>
```

Notice in this example that the comment text “<1000 organisms/ml” was translated into “< 1000 organisms/ml” as the less-than sign “<” is not a legal character in XML. Character translations that were required included: “<” was mapped to “<”, “>” was mapped to “>”, “&” was mapped to “&” and “” (ie a double quote in the text) was mapped to “"”.

Note that the comment:

```
COMMENT,""^^"SPECIMEN URINE", ""^^", ""^^"MICROSCOPY", ""^^" POLYMORPHS
20 per uL", ""^^" RED BLOOD CELLS 0 , ""^^" EPITHELIAL CELLS 10  \\"
\\"", ""^^", ""^^"CULTURE A T", ""^^" M M", ""^^"
P P", ""^^" : : :", ""^^"< 10,000 per ml. Strep. faecalis S S"
```

would be simply be translated into:

```
<comment>
```

```

<field>SPECIMEN URINE</field> <field></field> <field>MICROSCOPY</field>
<field> POLYMORPHS    20    per uL</field>
<field> RED BLOOD CELLS  0    </field>
<field> EPITHELIAL CELLS 10    </field>
<field></field> <field>CULTURE                A T</field>
<field>                M M</field> <field>                P P</field>
<field>                : : </field> <field>&lt; 10,000 per ml. Strep. faecalis S S</field>
</comment>

```

While an automated procedure could have been developed to extract the individual components of microscopy information from the comment text, it was felt that this process was too liable for error due to the unpredictability of the data. It was also agreed that without a pre-defined standard representation of these concepts by the individual labs involved, that manual user validation would always be required to confirm any processing decisions made, in order to avoid any potential medico-legal problems arising.

Note, however, that two comment types (namely “Reported by” and “Verified”) were extracted from the comment text for radiology data. It was felt that these comments were of a sufficiently standard format to warrant the individual text processing required for extracting type information, to allow future mappings to occur of these specific fields.

5. Defining the GEHR Archetypes

5.1 Overview

Once the raw data extracted from the OACIS system was interpreted and understood, it was possible to consider which generic GEHR archetype was most appropriate to standardise the representation of the data. A GEHR archetype is a clinical definition, which conforms to the basic building blocks defined in the GEHR Object Model (GOM). For the purposes of this project three new archetypes (and one specialisation archetype) were developed to represent the OACIS data using a GEHR record structure: namely Biochemistry, Microbiology, Radiology and Blood Lipids (a specialisation of Biochemistry).

GEHR archetypes provide a standardisation mechanism that will allow automatic processing of information within EHRs. The idea then is that if a clinician working with any clinical application wants to know the last 5 haemoglobin readings for a particular patient, then this can be determined safely. For this to be the case a standard mechanism for storing results is required. The level of standardisation is yet to be determined but this project has given us the opportunity to begin to assess this with a large clinical data set.

The question facing the development team is how many archetypes are required to deal with the results that are being downloaded and what is the best way to represent this information within the generic GEHR architecture. While the OACIS dataset is large, the most comprehensive work in this area has been by the team led by Dr Clem McDonald at the Regenstrief Institute¹. Their work has been with HL7 messages and has led to the development of LOINC – Logical Observation Identifiers Names and Codes.

The LOINC users manual provides a wealth of information about the variety of clinical information that might be received in a comprehensive health information system and as such has been the main information source for the development of the GEHR archetypes in this area. Australian HL7 message standards have also been a resource.

For radiology, the DICOM standard was also used as a resource. This standard relates predominantly to the storage of digital images but also contains clinical information. This has informed the more complex radiology archetype.

The clinical archetypes are very important. For example, in the section “Fields Requiring Separation” in section 4.3 above the instance of the word “fasting” appended to the units data in some blood glucose levels is raised. This appended word relates to the challenge the patient underwent before the blood sample was taken – that is to say the patient had no calories for 12 hours prior to the test. This state has enormous implications for the interpretation of a blood glucose level and will mean the difference between the diagnosis of diabetes or a normal result! There are many similar examples in clinical practice. An example is a hormone level after administration of a stimulating hormone – the level achieved may be quite abnormal without the result being significant.

¹ www.regenstrief.org/loinc

Although ‘fasting’ was the only clinical challenge we had to deal with in this trial, the archetype developed will deal with many other challenges including administration of substances. Further, the timing of a test may be critical, such as a hormone level at a particular point in a woman’s menstrual cycle. All factors identified that relate to pathology have been incorporated into the archetypes.

The result is two archetypes for biochemistry – one generic and one dealing with time series data. The later is only required when there is rapid sampling over an extended period and is not required to deal with any of the OACIS data. These biochemistry archetypes will deal with pathology results and haematology results. These are content archetypes and mean that multiple panels can be entered within one transaction.

A further archetype is required for microbiology, which is more complex. The archetype has been designed to receive all HL7 messages of this type and so some of the features are not utilised in this trial. This archetype is also a content archetype and a number can be entered in any transaction.

The radiology archetype is more complex. The health service transaction involving radiology is much more of a clinical interaction than the other areas studied in this trial. Clinical history is relayed back to the EHR as well as new clinical information derived by the radiologist. The radiologist may have carried out a procedure. It is also envisaged that a clinician receiving the data may require access to the images at a later date. This archetype contains links that will allow this. The radiology archetypes are therefore a full set of archetypes relating to a complete transaction – an event transaction, organiser and a number of content archetypes which are not unique to radiology but a more generally applicable in a clinical setting.

5.2 Terminology

Terminology is a complex issue throughout the health care domain at this time². The GEHR architecture has a potentially large impact on the requirements for terminology as all information is held within archetypes, which are themselves based on a standard set of building blocks. Importantly, while this does place some constraints on the EHR to allow automatic processing, these are minimal.

The providers of terminology at present in the Australian health domain are ICD 10, ICPC Plus, Docle, and more recently SNOMED. SNOMED and the UK NHS Clinical Terms groups are cooperating to produce what will be known as SNOMED CT – it is likely that this controlled nomenclature will have a wider uptake than the two separate terminologies at present. In Patholog, the work of the LOINC team is the most comprehensive. However, working within an EHR architecture changes everything!

Archetypes are uniquely identified and so to identify information within them a small set of terms is required. LOINC by comparison operates in the message domain space and faces a

² Slee V, Slee D, Schmidt HJ “The Endangered Medical Record: Ensuring its Integrity in the Age of Informatics”
Tringa Press, Minnesota USA 2000 ISBN 0-9615255-2-5

very different problem. Let us consider the term SODIUM for example. Within a biochemistry archetype only one sample is possible – so all values relate to a single sample. If more than one sample is taken, then the results will appear in a separate archetype. The way the information has been gathered – the measuring device etc – is recorded within the archetype protocol – a standard information construct with GEHR. To identify a SODIUM value within the biochemistry archetype, the simple term SODIUM will suffice.

For LOINC the situation is far more complex. A sodium measurement may be of urine, blood, gastric contents, it may be measured manually or automatically on various pieces of equipment. LOINC utilises a separate code for every possible combination of measurement scenarios – so that there are very many codes for SODIUM measurements. The same applies for POTASSIUM etc.

Further, it has always been the aim of GEHR to remain language independent and independent of particular coding schemes. For this reason, the Unified Medical Language System³ (UMLS) is appealing. The terminology is available in many languages and concept codes can be interpreted within many classification systems. For this reason the GEHR labels have been represented with UMLS concept unique identifiers (CUIs). It is important to recognise that with the archetype structures it would not be possible to use the LOINC termset without increasing the complexity of the archetypes by many orders of magnitude. It is also important to understand that HL7 messages with LOINC codes can be mapped into the GEHR archetype unambiguously. Once this mapping is completed – and it is a relatively simple exercise – the process can be utilised in any setting.

5.3 Archetype Specialisation

GEHR Archetypes can be specialised – that is to say, a generic archetype like biochemistry can be specialised to a specific test such as Blood lipids or Glucose tolerance test. This may be required locally where standardisation of a particular test result is required. Specialised archetypes satisfy all the constraints of the ‘parent’ and add more of their own. Blood lipid archetype is now expressed as a specialisation of the generic biochemistry archetype. Such specialised archetypes will be beneficial for mapping import of HL7 messages or for local standardisation. The data entered in this manner can be associated with the specialised or the generic archetype within the EHR itself and always be processed on the basis of the generic archetype – which forms the basis for sharing information.

5.4 GEHR Archetype Descriptions

In the following subsections we define the GEHR content archetypes into which the OACIS data was mapped. In addition to these archetypes, simple organiser and transaction archetypes were also used to encapsulate the content data - however, these supporting archetypes have not been included in this report.

³ <http://www.nlm.nih.gov/research/umls/umlsmain.html>

5.4.1 Demographics

The Demographics archetype used for this project is defined as follows:

Archetype		Demographics	
Identification			
GEHR identifier	gehr.content.define.patient_identity		
Type	Content		
Concept	Patient identity		
Description	Contains basic demographic details to be recorded about a patient.		
Use	Use to define the subject of an EHR		
Misuse	To be determined		
Structure			
Content Root			
Content Type	DEFINITION_CONTENT		
Hierarchical Proposition			
Form	List		
Hierarchical Group (Root)			
1. Hierarchical Value (0..*)			
Name	"EHR Source Patient ID"		
TYPE=PLAIN_TEXT			
2. Hierarchical Value (0..*)			
Name	"First Name"		
TYPE=PLAIN_TEXT			
3. Hierarchical Value (0..*)			
Name	"Last Name"		
TYPE=PLAIN_TEXT			
4. Hierarchical Value (0..*)			
Name	"Date of Birth"		
TYPE=DATE			
5. Hierarchical Value (0..*)			
Name	"Sex"		
TYPE= PLAIN_TEXT {"male", "female", "not specified"}			

5.4.2 Biochemistry and Blood Lipids

The Biochemistry archetype used for this project is defined as follows:

Archetype		Biochemistry	
Identification			
GEHR identifier	gehr.cont.observe.biochemistry		
Type	Content		
Concept			
Description	A generic archetype for laboratory results		
Use			
Misuse			
Structure			
Content Root			
Content Type	OBSERVATION_CONTENT		
Subject	{Self, Fetus, Newborn, Donor, CONTROL}		
Hierarchical Proposition			
Form	Hierarchy		
Hierarchical Group (Root)			
1. Hierarchical Value (0..*)			
Name	"Sample"		
TYPE=TERM_TEXT (See LOINC table (10))			
2. Hierarchical Group (0..*)			
Name	TERM_TEXT (Battery Name)		
2.1 Hierarchical Value (1..*)			
Name	TERM_TEXT (??)		
TYPE={QUANTITY,QUANTITY_RATIO}			
3. Hierarchical Value (0..*)			
Name	TERM_TEXT (??)		
TYPE={QUANTITY,QUANTITY_RATIO}			
4. Hierarchical Value (0..*)			
Name	"Comment"		
TYPE=PLAIN_TEXT			

5. Hierarchical Group (0..*)	
Name	“Context”
5.1 Hierarchical Group (0..*)	
Name	“Challenge”
5.1.1 Hierarchical Value (0..*)	
Name	“Action”
TYPE=PLAIN_TEXT (See LOINC Table 6 + many others!)	
5.1.2 Hierarchical Value (0..1)	
Name	“Route”
TYPE=TERM_TEXT (See LOINC Guide table 5)	
5.2 Hierarchical Value (0..1)	
Name	“Sequence number”
TYPE=NUMERIC	
5.3 Hierarchical Value (0..1)	
Name	TERM_TEXT (??)
TYPE=PLAIN_TEXT	
Protocol	
Hierarchical Proposition	
Form	List
Hierarchical Value (0..*)	
Name	“Laboratory”
TYPE=PLAIN_TEXT	
Hierarchical Value (0..*)	
Name	“Laboratory Id”
TYPE=PLAIN_TEXT	
Hierarchical Value (0..*)	
Name	“Adjustments
TYPE=PLAIN_TEXT (See LOINC 2.2.3)	
Hierarchical Value (0..*)	
Name	“Method”
TYPE=PLAIN_TEXT (See LOINC Table 13)	
Hierarchical Value (0..*)	
Name	“Time Aspect”

TYPE=PLAIN_TEXT (See LOINC Table 8)	
Hierarchical Value (0..*)	
Name	“Requested Time”
TYPE=DATETIME	
Hierarchical Value (0..*)	
Name	PLAIN_TEXT (New value)
TYPE=ANY	

5.4.3 Microbiology

The Microbiology archetype used for this project is defined as follows:

Archetype	Microbiology
Identification	
GEHR identifier	gehr.cont.observe.microbiology
Type	Content
Concept	Laboratory findings and comments for microbiology
Description	A generic archetype for laboratory results from Microbiology
Use	This archetype is for storing microbiology information that is returned from a laboratory system – either transformed from an HL7 message or as part of a transaction in a GEHR extract.
Misuse	Yet to be determined
Structure	
Content Root	
Content Type	OBSERVATION_CONTENT
Subject	{Self, Donor, Fetus}
Hierarchical Proposition	
Form	Hierarchy
Hierarchical Group (Root)	
1. Hierarchical Value (0..1)	
Name	“Battery Name”
TYPE=TERM_TEXT	

2. Hierarchical Group (0..1)	
Name	“Specimen Collection”
2.1 Hierarchical Value (0..1)	
Name	“Sample”
TYPE=PLAIN_TEXT	
2.2 Hierarchical Value (0..1)	
Name	“Site”
TYPE=PLAIN_TEXT	
2.3 Hierarchical Value (0..1)	
Name	“Method”
TYPE=PLAIN_TEXT	
3. Hierarchical Group (0..1)	
Name	“Macroscopic Findings”
3.1 Hierarchical Value (0..1)	
Name	“Colony Count”
TYPE=QUANTITY	
3.2 Hierarchical Value (0..1)	
Name	“Description”
TYPE=PARAGRAPH	
4. Hierarchical Group (0..1)	
Name	“Microscopy”
4.1 Hierarchical Value (0..*)	
Name	PLAIN_TEXT (Name of organism)
TYPE=QUANTITY (Count/Low power field... or Concentration)	
4.2 Hierarchical Value (0..1)	
Name	“Comments”
TYPE=PARAGRAPH	
5. Hierarchical Group (0..1)	
Name	“Culture”
5.1 Hierarchical Value (0..*)	
Name	“Method”
TYPE=PLAIN_TEXT	

5.2 Hierarchical Value (0..*)	
Name	“Findings”
TYPE=PLAIN_TEXT	
5.3 Hierarchical Group (0..*)	
Name	“Micro-organism Identified
5.3.1 Hierarchical Value (0..1)	
Name	“Micro-organism”
TYPE=PLAIN_TEXT {SNOMED - subset}	
5.3.2 Hierarchical Group (0..1)	
Name	“Therapeutic susceptibilities”
5.3.2.1 Hierarchical Group (0..*)	
Name	PLAIN_TEXT (Antibiotic names)
5.3.2.1.1 Hierarchical Value (1..*)	
Name	“Susceptibility description”
TYPE=PLAIN_TEXT {“Susceptible”, “Resistant”}	
5.3.2.1.2 Hierarchical Value (0..*)	
Name	“Susceptibility”
TYPE=QUANTITY	
5.3.2.1.3 Hierarchical Value (0..1)	
Name	“Method”
TYPE=PLAIN_TEXT {AGAR DIFFUSION, MIC, MLC, SBT, GRADIENT STRIP} See LOINC table 16	
6 Hierarchical Value (0..*)	
Name	“Overall Comment”
TYPE=PLAIN_TEXT	
Protocol	
Hierarchical Proposition	
Form	List
Hierarchical Value (0..1)	
Name	“Microscope Type”
TYPE=PLAIN_TEXT {Light, Electron microscopy}	
Hierarchical Value (0..1)	
Name	“Stain”
TYPE=PLAIN_TEXT {GRAM, ACID FAST, DRY MOUNT, INDIA INK, TRICHROME, GIEMSA}	

Hierarchical Value (0..*)	
Name	“Laboratory”
TYPE=PLAIN_TEXT	
Hierarchical Value (0..*)	
Name	“Laboratory Id”
TYPE=PLAIN_TEXT	

Hierarchical Value (0..*)	
Name	“Adjustments
TYPE=PLAIN_TEXT (See LOINC 2.2.3)	
Hierarchical Value (0..*)	
Name	“Method”
TYPE=PLAIN_TEXT (See LOINC Table 13)	
Hierarchical Value (0..*)	
Name	“Time Aspect”
TYPE=PLAIN_TEXT (See LOINC Table 8)	
Hierarchical Value (0..*)	
Name	“Requested Time”
TYPE=DATETIME	
Hierarchical Value (0..*)	
Name	PLAIN_TEXT (New value)
TYPE=ANY	

5.4.4 Radiology

Four archetypes were used to represent the radiology data from the OACIS system: a radiology report transaction archetype, a radiology report organiser archetype, a radiology details content archetype and a clinical information content archetype.

The Radiology report transaction archetype used for this project is defined as follows:

Archetype		Radiology Report	
Identification			
GEHR identifier		gehr.trans.event.radiology	
Type		Transaction, event	

Concept	A radiology report
Description	A radiology report in response to a referral from a clinician reporting back to that clinician
Use	
Misuse	
Compound	gehr.trans.event.radiology \ gehr.org.radiology_report \ gehr.cont.xxxx
Organisers	
Organiser (1)	
Name	gehr.org.radiology_report
Other Context	
Hierarchical Proposition	
Form	List
1. Hierarchical Value (0..*)	
Name	“Radiology Service”
TYPE=PLAIN_TEXT	
2 Hierarchical Value (0..*)	
Name	“Verified By”
TYPE=DEMOGRAPHIC (Demographic of clinician verifying report)	

The Radiology report organiser archetype used for this project is defined as follows:

Archetype	Radiology Report
Identification	
GEHR identifier	gehr.org.radiology_report
Type	Organiser
Concept	Organiser for a radiology report
Description	
Use	This organiser is to be used in gehr.trans.event.radiology
Misuse	
Compound	gehr.trans.event.radiology \ gehr.org.radiology_report \ gehr.cont.subj xxx or gehr.cont.obj.xxx or gehr.cont.inst.xxx

Organisers	
1. Organiser	
Name	“Report
Concept	The root organiser for the report
1.1 Organiser (1..*)	
Name	PLAIN TEXT (Study name)
Concept	The name of the study
Contains	gehr.cont.observe.radiology_details
1.1.1 Organiser (0..*)	
Name	“Clinical Information”
Concept	Clinical details supplied by referring clinician
Contains	gehr.cont.subj.clinical_information or gehr.cont.subj.*
1.1.2 Organiser (1..*)	
Name	“Findings”
Concept	Clinical findings supplied by reporting clinician
Contains	gehr.cont.obj.* or gehr.cont.obj.bodypartmeasure or gehr.cont.subj.clinical_information
1.1.3 Organiser (0..*)	
Name	“Conclusion”
Concept	Assessment or conclusion supplied by reporting clinician
Contains	gehr.cont.subj.{assessment,summary,conclusion} gehr.cont.inst.*

The Radiology details content archetype used for this project is defined as follows:

Archetype	Radiology Details
Identification	
GEHR identifier	gehr.cont.observe.radiology_details
Type	Content
Concept	Radiological logical information apart from clinical information
Description	Details to be determined
Use	

Misuse	
Structure	
Content Root	
Content Type	OBSERVATION_CONTENT
Subject	{Self, Donor, Fetus}
Hierarchical Proposition	
Form	Hierarchy
Hierarchical Group (Root)	
1. Hierarchical Value (0..1)	
Name	“Investigation Type”
TYPE=TERM_TEXT (from LOINC)	
2. Hierarchical Group (0..*)	
Name	“Image Set”
2.1 Hierarchical Value (0..1)	
Name	PLAIN TEXT (Image Name)
TYPE=GEHRL (Link to Image)	
2.2 Hierarchical Value (0..1)	
Name	“Method”
TYPE=PLAIN_TEXT	
2.3 Hierarchical Value (0..1)	
Name	“Comment”
TYPE=PARAGRAPH	
Protocol	
Hierarchical Proposition	
Form	List
Hierarchical Value (0..*)	
Name	“Service”
TYPE=PLAIN_TEXT	
Hierarchical Value (0..*)	
Name	“Laboratory ID”
TYPE=PLAIN_TEXT	
Hierarchical Value (0..*)	
Name	“Adjustments

TYPE=PLAIN_TEXT (See LOINC 2.2.3)	
Hierarchical Value (0..*)	
Name	“Method”
TYPE=PLAIN_TEXT (See LOINC Table 13)	
Hierarchical Value (0..*)	
Name	“Time Aspect”
TYPE=PLAIN_TEXT (See LOINC Table 8)	
Hierarchical Value (0..*)	
Name	“Requested Time”
TYPE=DATETIME	
Hierarchical Value (0..*)	
Name	PLAIN_TEXT (New value)
TYPE=ANY	

The clinical information content archetype used for this project is defined as follows:

Archetype	Clinical Information
Identification	
GEHR identifier	gehr.cont.subj.clinical_information
Type	Content
Concept	Clinical information that are to be passed as a summary for the purposes of a test being carried out or a report being returned.
Description	
Use	
Misuse	
Structure	
Content Root	
Content Type	SUBJECTIVE_CONTENT
Subject	{fetus, child, sibling, mother, father, VALID NAMES}
Provider	{HCP, Other clinician}
Hierarchical Proposition	
Form	Hierarchy

Hierarchical Group (Root)	
1. Hierarchical Group (0..1)	
Name	PLAIN_TEXT
1.1. Hierarchical Group (0..*)	
Name	PLAIN_TEXT
1.1.1 Hierarchical Value (0..*)	
Name	PLAIN TEXT
TYPE=PARAGRAPH (Any)	
2 Hierarchical Value (0..*)	
Name	“Comment”
TYPE=PLAIN_TEXT	
Protocol	
Hierarchical Proposition	
Form	List
Hierarchical Value (0..*)	
Name	PLAIN_TEXT (New value)
TYPE=ANY	

5.5 GEHR Archetype Editor (Clinical Model Builder)

To create a formal representation of the constraints defined in each of these archetypes, it is recommended that an archetype editor (or ‘clinical model builder’) be used to provide a user friendly means of building archetypes. Based on the archetype designed, the archetype editor should be able to generate an archetype representation (in this case XML), which can be directly imported into the GEHR kernel to initialise the appropriate archetypes.

The archetype editor used for this project is one that has been independently developed by the DSTC, and is illustrated in the following screen shots.

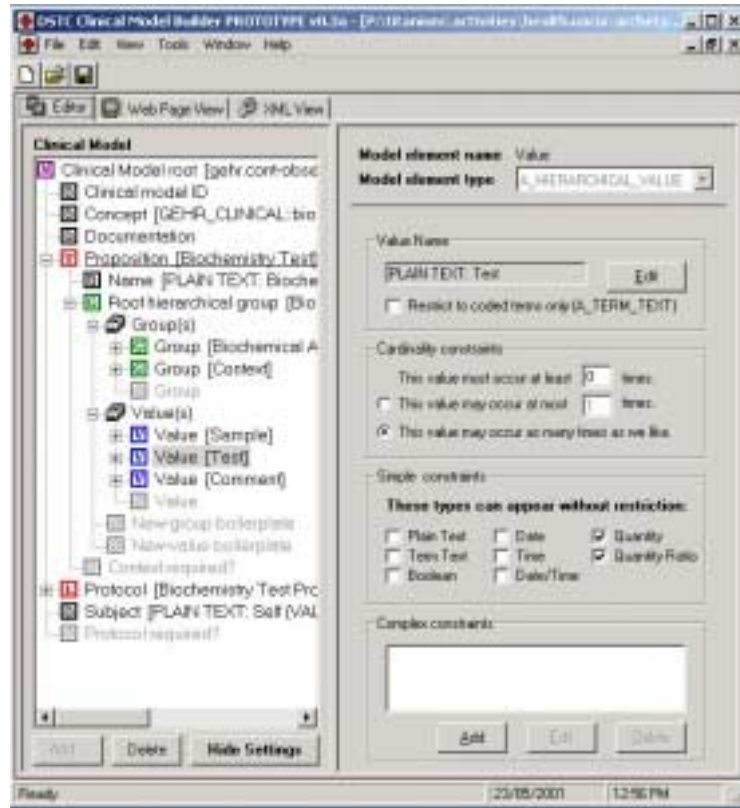


Figure 2: Editing an Archetype in the 'Clinical Model Builder'

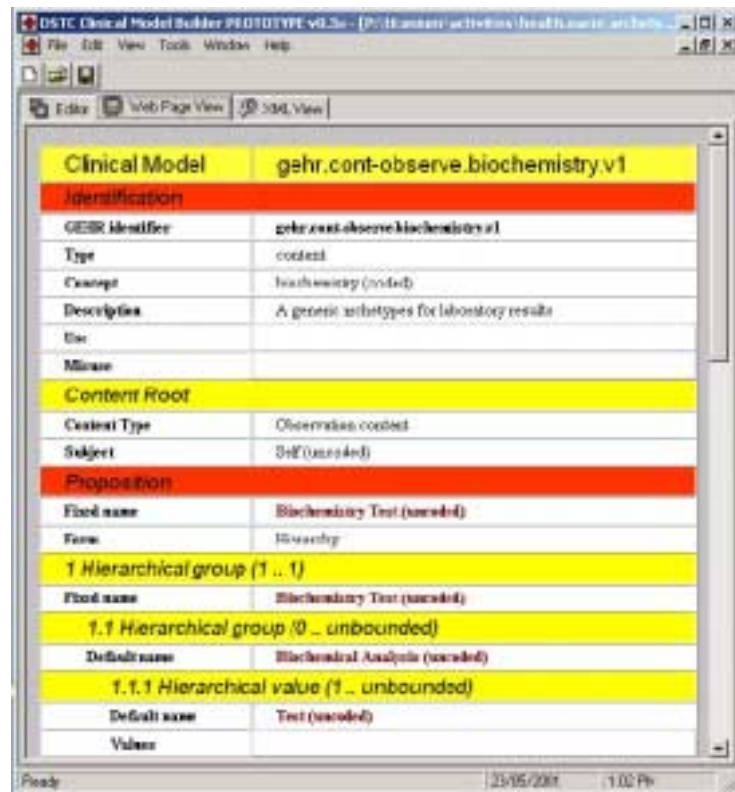


Figure 3: The Web Page View of an archetype in the 'Clinical Model Builder'

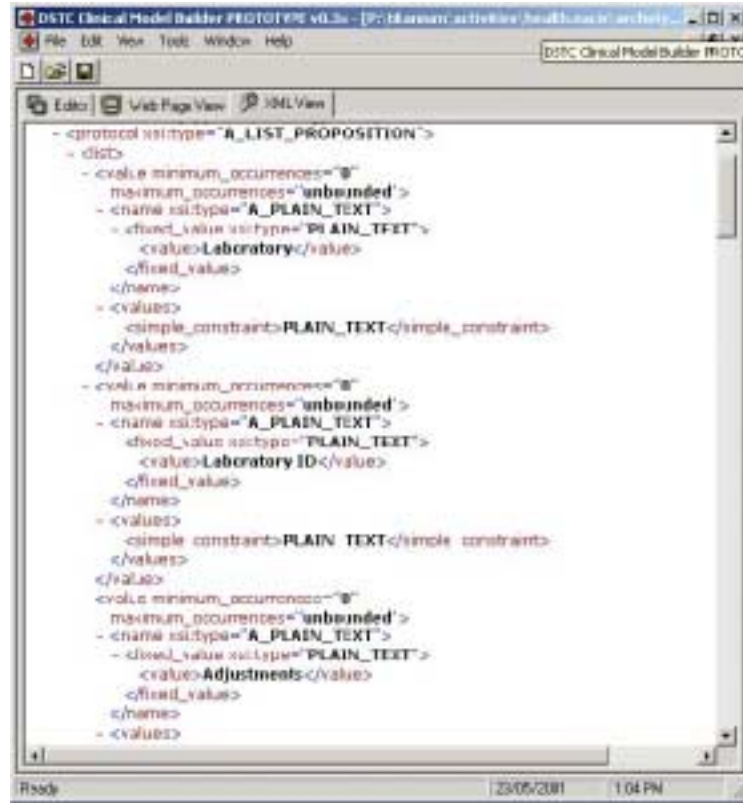


Figure 4: The XML representation of an archetype in the 'Clinical Model Builder'

Figure 2 shows the Biochemistry archetype being created and edited in the archetype editor. In Figure 3, the read-only web-page view of the archetype is shown, for easier comprehensibility and archetype web-publishing. Finally, Figure 4 shows part of the Biochemistry archetype's XML representation, which is automatically generated by the archetype editor. This XML format, which conforms to the archetype constraint classes defined in the GEHR kernel, can be directly imported into the GEHR kernel to enable the initialisation of the relevant archetypes

5.6 Problems and Issues

It is important to recognise that GEHR archetypes are an experimental construct at present but are being well received by the international community. Tom Beale, who has overseen the test implementation of this approach, is to be the guest at the HL7 annual Board retreat this year. However there is still considerable scepticism (and rightly). Importing a large clinical dataset, such as has been undertaken in this trial, is the first evidence we have that our approach can deal with real data of a reasonably diverse nature. It is also clear that the ability to store information, rather than data, is considerably greater within GEHR record rather than the current relational data approach. There is no doubt that it is possible to store highly complex data within traditional relational databases – the issue is that maintaining and sharing this information become virtually impossible and the performance overhead becomes potentially overwhelming. There are good reasons to pursue the GEHR approach! A major issue is that it will not be clear that the archetype approach is a good one until we have had to

deal with many and varied clinical applications. Each one takes time and may impact on the specific archetype used for mapping. The generic biochemistry archetype will be required to remain relatively stable over time with an upgrade path that is backwardly compatible. Inevitably, this will not be without its problems.

What is clear is that the GEHR archetype system, based on the lessons learned from implementing LOINC, can easily cope with OACIS data. Unfortunately some of the OACIS data is in unstructured text format that varies a great deal and as such is not available to be expressed within the structure of the GEHR archetype and will be lumped into one of the archetype 'fields'. This makes it available to clinicians but it is not available for automatic processing.

Finally, a text indexing service may be appropriate within the GEHR kernel so that text can be found quickly – this would maximise access to such unstructured textual information. Natural language processing of some use may be available for this purpose in the future!

6. Mapping OACIS Data into Archetypes

6.1 Overview

Once the raw data from the OACIS system has been pre-processed into generic XML (Section 4) and the appropriate clinical archetypes have been defined (Section 5), a mapping process is performed to define the relationships between the OACIS data and the fields in the target clinical archetypes. This mapping process is used to generate the XSLT script used in the subsequent stage to transform the generic XML data into GEHR-compliant XML data conforming to the specified clinical archetypes.

Note that each row of raw data extracted from the OACIS system is mapped into a separate Electronic Health Record (EHR) ‘extract’, represented in XML. Each test result, within a row of OACIS data is mapped into a separate GEHR ‘versioned_transaction’ within the row’s EHR extract. The GEHR EHR extract generated in this process is in a form, which can be transferred from one EHR system to another, and then imported into a GEHR kernel for validation and storage.

6.2 Mapping Definition

The following tables show the relationship between the generic XML data values and the GEHR EHR extract fields into which these values are mapped. An example value that could appear in each EHR extract field is also provided.

Note that fields in the generic XML are referred to using XML paths. For example, ‘/record/name/firstname’ would refer to the value of the generic XML field titled “firstname” (e.g. “Paul”). An artificial ‘Field ID’ is used to indicate the level of nesting of each GEHR field being mapped to. Also note that transactions of different types are identified using the following colour coding scheme:

The first mapping table, shown below, shows the mapping between the generic XML and the GEHR fields which provide general information about the extract itself.

Field ID	GEHR Extract Field	Mapped Generic XML	Example Value
1	Ehr_Extract	-	-
1.1	Ehr_source	“OACIS”	“OACIS”
1.2	Creation_time	[\$Time] {Time of extract creation }	15/03/2001 9:15:00
1.3	Hcp_created_by	[\$User] {Health care administrator who creates extract }	“Zar Zar Tun”
1.4	Destination_ehr_source	“DSTC_1”	“DSTC_1”
1.5	Hca_authorising_acquisition	[\$Auth] {Healthcare administrator who authorises transfer of extract }	“Scott Germann”

The following four tables shows the XML mapping for each of the four GEHR transaction types used in this project: the *Subject* transaction (which describes the patient’s details), the *Biochemistry* transaction, the *Microbiology* transaction and the *Radiology* transaction.

The following table describes the generic XML to GEHR mapping for a *Subject* transaction.

Field ID	GEHR Transaction Field	Mapped Generic XML	Example Value
1.6 (S)	Subject	-	-
S.1	Time_created	[\$Time] {Time of transaction creation}	15/03/2001 9:15:01
S.2	Versions/item	-	-
S.2.1	Audit		
S.2.1.2	Hca_authorising	“Scott Germann”	“Scott Germann”
S.2.1.3	Time	[\$Time] {Time of version creation}	15/03/2001 9:15:02
S.2.1.4	Reason	“Initial Version”	“Initial Version”
S.2.1.5	Ehr_source_id	“OACIS”	“OACIS”
S.2.1.6	Language	“EN”	“EN”
S.2.1.7	Territory	“AUS”	“AUS”
S.2.1.8	Code_set	“ISO8859-1”	“ISO8859-1”
S.2.2	Content {EHR Content}	-	-
S.2.2.1	Content {Organiser Root}	-	-
S.2.2.1.1	Name/value	“Patient Identity”	“Patient Identity”
S.2.2.1.2 (SC)	Content {Definition_Content}	-	-
SC.1	Item	-	-
SC.1.1	Archetype_id	“gehr.trans-persist.patient_identity	“gehr.trans-persist.patient_identity”
SC.1.2	Concept/value	“Patient_Identity”	“Patient Identity”
SC.1.3	Proposition	-	-
SC.1.3.1	Name/value	“Patient Identity”	“Patient Identity”
SC.1.3.2	Root	Xsi:type=”G1_HIERARCHICAL_GROUP”	Xsi:type=”G1_HIERARCHICAL_GROUP”
SC.1.3.2.1	Name/value	“Patient_Identity”	“Patient Identity”
SC.1.3.2.2 (SCC)	Children	-	-
SCC.1	Item	Xsi:type=”G1_HIERARCHICAL_VALUE”	Xsi:type=”G1_HIERARCHICAL_VALUE”
SCC.1.1	Name/value	“EHR Source Patient ID”	“EHR Source Patient ID”
SCC.1.2	Value/value (xsi:type = ”G1_PLAIN_TEXT”)	/record/MRN	“001054”
SCC.2	Item	Xsi:type=”G1_HIERARCHICAL	Xsi:type=”G1_HIERARCHICAL

		HICAL_VALUE”	CHICAL_VALUE”
SCC.2.1	Name/value	“First Name”	“First Name”
SCC.2.2	Value/value (xsi:type = “G1_PLAIN_TEXT”	/record/name/firstname	“Paul”
SCC.3	Item	Xsi:type=“G1_HIERAR HICAL_VALUE”	Xsi:type=“G1_HIERAR CHICAL_VALUE”
SCC.3.1	Name/value	“Last Name”	“Last Name”
SCC.3.2	Value/value (xsi:type = “G1_PLAIN_TEXT”	/record/name/lastname	“Duncan”
SCC.4	Item	Xsi:type=“G1_HIERAR HICAL_VALUE”	Xsi:type=“G1_HIERAR CHICAL_VALUE”
SCC.4.1	Name/value	“Date of Birth”	“Date of Birth”
SCC.4.2	Value/value (xsi:type = “G1_PLAIN_TEXT”	/record/dob	“3/08/1923”
SCC.5	Item	Xsi:type=“G1_HIERAR HICAL_VALUE”	Xsi:type=“G1_HIERAR CHICAL_VALUE”
SCC.5.1	Name/value	“Sex”	“Sex”
SCC.5.2	Value/value (xsi:type = “G1_PLAIN_TEXT”	IF /record/name/sex in {“m”,“M”} THEN “male” ELIF /record/name/sex in {“f”,“F”} THEN “female” ELSE “not Specified”	“male”
S.4.2.2	Is_persistent	“TRUE”	“TRUE”
S.4.2.3	Persistence_status	“Pers_active”	“Pers_active”

The following table describes the generic XML to GEHR mapping for a *Biochemistry* transaction – included within the set of event clinical transactions of an extract.

Field ID	GEHR Transaction Field	Mapped Generic XML	Example Value
1.7 (E)	Event_clinical_transactions	-	-
E.1 (B)	Item {Biochemistry versioned_transaction}		-
B.1	Time_created	[\$Time] {Time of transaction creation}	15/03/2001 9:15:05
B.2	Versions/item	-	-
B.2.1	audit		
B.2.1.2	Hca_authorising	“Scott Germann”	“Scott Germann”
B.2.1.3	Time	[\$Time] {Time of version creation}	15/03/2001 9:15:06
B.2.1.4	Reason	IF any value for data/status in the row is = “I” THEN Reason = “Interim Version”	“Final Version”

		ELSIF all values for data/status = "F" and no subsequent rows have the same (lab-number, mrn) values THEN Reason = "Final Version" ELSE Reason = "Unknown"	
B.2.1.5	Ehr_source_id	"OACIS"	"OACIS"
B.2.1.6	Language	"EN"	"EN"
B.2.1.7	Territory	"AUS"	"AUS"
B.2.1.8	Code_set	"ISO8859-1"	"ISO8859-1"
B.2.2	Content {EHR Content}	-	-
B.2.2.1	Content {Organiser Root}	-	-
B.2.2.1.1	Name/value	"Biochemistry Test"	"Biochemistry Test"
B.2.2.1.2	Content {Observation_Content}	-	-
B.2.2.1.2.1 (BD)	Item		-
BD.1	Archetype_id {For this transaction, assume that /record/row-type = "B"}	IF /record/row-type = "B" THEN Archetype_id = "Gehr.cont.observe.biochemistry" ELSIF /record/row-type = "M" THEN Archetype_id = "Gehr.cont.observe.microbiology" ELSIF /record/row-type = "R" THEN Archetype_id="Gehr. Cont.observe.radiology" ELSE Send row to error log	"Gehr.cont.observe.biochemistry"
BD.2	Concept/value	"Biochemistry Test result"	"Biochemistry Test Result"
BD.3	Proposition (xsi:type="G1_TREE_PROPOSITION")	-	-
BD.3.1	Name/value	"Biochemistry Test Result"	"Biochemistry Test Result"
BD.3.2	Root (xsi:type="G1_HIERARCHICAL_GROUP")	-	-
BD.3.2.1	Name/value	"Biochemistry Test Result"	"Biochemistry Test Result"

BD.3.2.2 (BC)	Children	-	-
BC.1	Item (xsi:type="G1_HIERARCHICAL_VALUE")	-	-
BC.1.1.1	Name/value	"Sample"	"Sample"
BC.1.2.1	Value/value (xsi:type="G1_TERM_TEXT")	"Blood – peripheral"	"Blood – peripheral"
BC.1.2.2	Name/primary	-	-
BC.1.2.2.1	Umls_entry/cui	"C0229664"	"C0229664"
BC.2	Item (xsi:type="G1_HIERARCHICAL_GROUP")	-	-
BC.2.1.1	Name/value	/record/master-panel	"BIOCHEMICAL ANALYSIS"
BC.2.2 (BCC)	Children	-	-
BCC.1	Item (xsi:type="G1_HIERARCHICAL_VALUE")	{e.g.: Sodium Test Item}	-
BCC.1.1	Name (xsi:type="G1_TERM_TEXT")	-	-
BCC.1.1.1	value	/record/data/result-name	"Sodium"
BCC.1.1.2	primary	-	-
BCC.1.1.2.1	Umls_entry/cui	[cui] (from Common Biochemistry Elements Lookup Table)	"C0037473"
BCC.1.2	Value (xsi:type="G1_QUANTITY")	-	-
BCC.1.2.1	Value	/record/data/result-value	"143"
BCC.1.2.2	Property	[property] from Units-Property Lookup table	"Concentration"
BCC.1.2.3	Units	/record/data/units	"mmol/L"
BCC.1.3	Reference/normal_range	-	-
BCC.1.3.1	Lower (G1_REAL)	/record/data/normal-low	"137"
BCC.1.3.2	Upper (G1_REAL)	/record/data/normal-high	"145"
BCC.2	Item (xsi:type="G1_HIERARCHICAL_VALUE")	{e.g.: Potassium Test Item: Same as for BCC.1: Sodium Test Item}	-
BCC.3	Item (xsi:type="G1_HIERARCHICAL_VALUE")	{e.g.: Glucose Test Item: Same as for BCC.1: Sodium Test Item}	-
BC.3	Item (xsi:type="G1_HIERARCHICAL_VALUE")	-	-
BC.3.1.1	Name/value	"Comment"	"Comment"
BC.3.2.1	Value/value (xsi:type="G1_TERM_TEXT")	/record/comment	"Plasma"

	= "G1_PLAIN_TEXT"		
BC.4	Item (xsi:type="G1_HIERARCHICAL_GROUP")	-	-
BC.4.1.1	Name/value	"Context"	"Context"
BC.4.2 (BCX)	Children	-	-
BCX.1	Item (xsi:type="G1_HIERARCHICAL_GROUP")		-
BCX.1.1.1	Name/value	"Challenge"	"Challenge"
BCX.1.2	Children	-	-
BCX.1.2.1 (BCA)	Item (xsi:type="G1_HIERARCHICAL_VALUE")	FOR EACH /record/data/action value in current row CREATE a new BCA.1.1 Action item	-
BCA.1.1	Name/value (xsi:type="G1_PLAIN_TEXT")	"Action"	"Action"
BCA.2	Value (xsi:type="G1_TERM_TEXT")	-	-
BCA.2.1	value	/record/data/action	"Fasting"
BCA.2.2	primary	-	-
BCA.2.2.1	Umls_entry/cui	IF /record/data/action = "Fasting" THEN cui = "C0015633" ELSE cui = null	"C0015633"
B.3	Context	-	-
B.3.1	Recorder	"unknown"	"unknown"
B.3.2	Comment/value	Null	Null
B.3.3	Observation_time/value	/record/received-time	"10/06/1997 9:15:00"
B.3.4	Activity_time/value	/record/collection-time	"10/06/1997"
B.3.5	Location	Null	Null
B.4	Subject/value	"Self"	"Self"
B.5 (BP)	Protocol (xsi:type="G1_LIST_PROPOSITION")	-	-
BP.1	Name/value	"Biochemistry Test Protocol"	"Biochemistry Test Protocol"
BP.2	Item (xsi:type="G1_HIERARCHICAL_VALUE")	-	-
BP.2.1.1	Name/value	"Laboratory"	"Laboratory"
BP.2.1.2	Value/value (xsi:type="G1_PLAIN_TEXT")	/record/lab-name	"RAH-LAB"
BP.3	Item (xsi:type="G1_HIERARCHICAL_VALUE")	-	-
BP.3.1.1	Name/value	"Laboratory ID"	"Laboratory ID"
BP.3.2.1	Value/value (xsi:type="G1_PLAIN_TEXT")	/record/lab-number	"2193686CHEM"

BP.4	Item (xsi:type="G1_HIERARCHICAL_VALUE")	-	-
BP.4.1.1	Name/value	"Method"	"Method"
BP.4.2.1	Value/value (xsi:type = "G1_PLAIN_TEXT")	-	-
BP.5	Item (xsi:type="G1_HIERARCHICAL_VALUE")	-	-
BP.5.1.1	Name/value	"Requested Time"	"Requested Time"
BP.5.2.1	Value/value (xsi:type = "G1_DATE_TIME")	/record/requested-time	"10/06/1997"

The following table describes the generic XML to GEHR mapping for a *Microbiology* transaction – included within the set of event clinical transactions of an extract.

Field ID	GEHR Transaction Field	Mapped Generic XML	Example Value
E.2 (M)	Item {Microbiology versioned_transaction}		
M.1	Time_created	[\$Time] {Time of transaction creation}	15/03/2001 9:16:35
M.2	Versions/item	-	-
M.2.1	audit		
M.2.1.2	Hca_authorising	"Scott Germann"	"Scott Germann"
M.2.1.3	Time	[\$Time] {Time of version creation}	15/03/2001 9:15:06
M.2.1.4	Reason	IF last version with same (lab-number, mrn) values THEN Reason = "Final" ELSE Reason = "Interim"	"Final Version"
M.2.1.5	Ehr_source_id	"OACIS"	"OACIS"
M.2.1.6	Language	"EN"	"EN"
M.2.1.7	Territory	"AUS"	"AUS"
M.2.1.8	Code_set	"ISO8859-1"	"ISO8859-1"
M.2.2	Content {EHR Content}	-	-
M.2.2.1	Content {Organiser Root}	-	-
M.2.2.1.1	Name/value	"Microbiology Test"	"Microbiology Test"
M.2.2.1.2	Content {Observation_Content}	-	-
M.2.2.1.2.1 (MD)	Item		-
MD.1	Archetype_id {For this transaction assume that /record/row-	IF /record/row-type = "B" THEN Archetype_id = "Gehr.cont.observe.biochemistry"	"Gehr.cont.observe.microbiology"

	type = "M"} 	ELSIF /record/row-type = "M" THEN Archetype_id = "Gehr.cont.observe. microbiology" ELSIF /record/row-type = "R" THEN Archetype_id= "Gehr. cont.observe. radiology" ELSE Send row to error log	
MD.2.1	Concept/value	"Laboratory findings and omments for icrobiology"	"Laboratory findings and comments for microbiology"
MD.3	Proposition (xsi:type= "G1_TREE_PROPOSITI ON")	-	-
MD.3.1	Name/value	"Microbiology Test Result"	"Microbiology Test Result"
MD.3.2	Root (xsi:type="G1_HIE RARCHICAL_GROUP")	-	-
MD.3.2.1	Name/value	"Microbiology Test Result"	"Microbiology Test Result"
MD.3.2.2 (MC1)	Children	-	-
MC.1	Item (xsi:type="G1_HIE RARCHICAL_VALUE")		-
MC.1.1.1	Name/value	"Battery Name"	"Battery Name"
MC.1.2.1	Value/value (xsi:type= "G1_PLAIN_TEXT")	/record/master-panel	"MICROBIOLOGY"
MC.2	Item (xsi:type="G1_HIERARC HICAL_GROUP")	-	-
MC.2.1.1	Name/value	"Specimen collection"	"Specimen collection"
MC.2.2 (MC2C)	Children	-	-
MC2C.1	Item (xsi:type="G1_HIE RARCHICAL_VALUE")		-
MC2C.1.1.1	Name/value	"Sample"	"Sample"
MC2C.1.2.1	Value/value (xsi:type= "G1_PLAIN_TEXT")	/record/comment/field where [@type] = "SPECIMEN"	"Urine"
MC2C.2	Item (xsi:type="G1_HIERARC HICAL_VALUE")		-

MC2C.2.1.1	Name/value	“Site”	“Site”
MC2C.2.2.1	Value/value (xsi:type=“G1_PLAIN_TEXT”)	/record/comment/field where [@type] = “SPECIMEN SITE”	“Midstream”
MC.3	Item (xsi:type=“G1_HIERARCHICAL_GROUP”)	-	-
MC.3.1.1	Name/value	“Macroscopic findings”	“Macroscopic findings”
MC.3.2 (MC3C)	Children	-	-
MC3C.1	Item (xsi:type=“G1_HIERARCHICAL_VALUE”)	-	-
MC3C.1.1.1	Name/value	“Description”	“Description”
MC3C.1.2.1 (MC3CP)	Value/value (xsi:type=“G1_PARAGRAPH”)	-	-
MC3CP. 1.1	Items/item (xsi:type=“PARAGRAPH_ITEM”)	-	-
MC3CP. 1.1.1.1	Items/item (xsi:type=“PLAIN_TEXT_LIST”)	-	-
MC3CP. 1.1.1.1.1	Value	/record/comment/field where [@type] = “MACROSCOPY”	“Specimen shows macroscopic haematamia”
MC.4	Item (xsi:type=“G1_HIERARCHICAL_GROUP”)	-	-
MC.4.1.1	Name/value	“Microscopy”	“Microscopy”
MC.4.2 (MC4C)	Children	-	-
MC4C.1	Item (xsi:type=“G1_HIERARCHICAL_VALUE”)	-	-
MC4C.1.1.1	Name/value	“Comments”	“Comments”
MC4C.1.2.1 (MC4CP)	Value/value (xsi:type=“G1_PARAGRAPH”)	-	-
MC4CP. 1.1	Items/item (xsi:type=“PARAGRAPH_ITEM”)	-	-
MC4CP. 1.1.1.1	Items/item (xsi:type=“PLAIN_TEXT_LIST”)	-	-
MC4CP. 1.1.1.1.1	Value	/record/comment/field where [@type] = “MICROSCOPY”	“No bacteria seen”
MC.5	Item (xsi:type=“G1_HIERARCHICAL_GROUP”)	-	-

MC.5.1.1	Name/value	“CULTURE”	“CULTURE”
MC.5.2 (MC5C)	Children	-	-
MC5C.1	Item (xsi:type=”G1_HIERARCHICAL_VALUE”)		-
MC5C.1.1.1	Name/value	“Findings”	“Findings”
MC5C.1.2.1	Value/value (xsi:type=”G1_PLAIN_TEXT”)	/record/comment/field where [@type] = “CULTURE”	“No growth after seven days”
MC5C.2	Item (xsi:type=”G1_HIERARCHICAL_GROUP”)	-	-
MC5C.2.1.1	Name/value	“Micro-organism identified”	“Micro-organism identified”
MC5C.2.2 (MO)	Children	-	-
MO.1	Item (xsi:type=”G1_HIERARCHICAL_VALUE”)		-
MO.1.1.1	Name/value	“Micro-organism”	“Micro-organism”
MO.1.2	Value (xsi:type=”G1_TERM_TEXT”)	-	-
MO.1.2.1	value	/record/data/organism	“Escherichia coli”
MO.1.2.2	primary	-	-
MO.1.2.2.1	Umls_entry/cui	??	??
MO.1.2.2.2	Code	??	??
MO.1.2.2.3	Expansions	??	??
MO.1.2.3	Term_set	“SNOMED”	“SNOMED”
MO.2	Item (xsi:type=”G1_HIERARCHICAL_GROUP”)		-
MO.2.1.1	Name/value	“Therapeutic susceptibilities”	“Therapeutic susceptibilities”
MO.2.2 (MT)	Children	-	-
MT.1	Item (xsi:type=”G1_HIERARCHICAL_GROUP”)		-
MT.1.1.1	Name/value	/record/data/antibiotic	“CEPHALOTHIN/CEPHALEXIN”
MT.1.2 (MTC)	Children	-	-
MTC.1	Item (xsi:type=”G1_HIERARCHICAL_VALUE”)		-
MTC.1.1.1	Name/value	“Susceptibility Description”	“Susceptibility Description”
MTC.1.2	Value (xsi:type=	IF /record/data/sensitivity	“Susceptible”

	"G1_PLAIN_TEXT")	= "R" THEN "Resistant" ELSIF /record/data/sensitivity = "S" THEN "Susceptible" ELSE Null	
MC.6	Item (xsi:type="G1_HIERARCHICAL_VALUE")	-	-
MC.6.1.1	Name/value	"General Comments"	"General Comments"
MC.6.2	Value (xsi:type="G1_PLAIN_TEXT")	/record/comment/field [@type] & "=" & /record/comment/field	"WHITE BLOOD CELLS (N<10) = >1000,ANTI- MICROBIAL AGENTS = Not detected"
M.3	Context	-	-
M.3.1	Recorder	"unknown"	"unknown"
M.3.2	Comment/value	Null	Null
M.3.3	Observation_time/value	/record/received-time	"8/05/1997 13:11:00"
M.3.4	Activity_time/value	/record/collection-time	"8/05/1997"
M.3.5	Location	Null	Null
M.4	Subject/value	"Self"	"Self"
M.5 (MP)	Protocol (xsi:type="G1_LIST_PROPOSITION")	-	-
MP.1	Name/value	"Microbiology Test Protocol"	"Microbiology Test Protocol"
MP.2	Item (xsi:type="G1_HIERARCHICAL_VALUE")	-	-
MP.2.1.1	Name/value	"Laboratory"	"Laboratory"
MP.2.1.2	Value/value (xsi:type="G1_PLAIN_TEXT")	/record/lab-name	"RAH-LAB"
MP.3	Item (xsi:type="G1_HIERARCHICAL_VALUE")	-	-
MP.3.1.1	Name/value	"Laboratory ID"	"Laboratory ID"
MP.3.2.1	Value/value (xsi:type="G1_PLAIN_TEXT")	/record/lab-number	"2193686CHEM"
MP.4	Item (xsi:type="G1_HIERARCHICAL_VALUE")	-	-
MP.4.1.1	Name/value	"Requested Time"	"Requested Time"
MP.4.2.1	Value/value (xsi:type="G1_DATE_TIME")	/record/requested-time	"10/06/1997"

The following table describes the generic XML to GEHR mapping for a *Radiology* transaction – included within the set of event clinical transactions of an extract.

Field ID	GEHR Transaction Field	Mapped Generic XML	Example Value
E.3 (R)	Item {Radiology versioned_transaction}		
R.1	Time_created	/record/comment/field where [@type] = "Verified"	15/03/2001 9:16:35
R.2	Versions/item	-	-
R.2.1	audit		
R.2.1.2	Hcp_authorising	/record/reported-by	"WALLACE,M"
R.2.1.3	Time	[\$Time] {Time of version creation}	15/03/2001 9:15:06
R.2.1.4	Reason	IF last version with same (lab-number, mrn) values THEN Reason = "Final" ELSE Reason = "Interim"	"Final Version"
R.2.1.5	Ehr_source_id	"OACIS"	"OACIS"
R.2.1.6	Language	"EN"	"EN"
R.2.1.7	Territory	"AUS"	"AUS"
R.2.1.8	Code_set	"ISO8859-1"	"ISO8859-1"
R.2.2	Content {EHR Content}	-	-
R.2.2.1	Archetype_id {For this transaction assume that /record/row-type = "R"}	"Gehr.trans.event.radiology"	"Gehr.trans.event.radiology"
R.2.2.2 (RO)	Content {Organiser Root}	-	-
RO.1	Name/value	"Report"	"Report"
RO.2.1	Organisers/Item	-	-
RO.2.1.1.1	Name/value (PLAIN_TEXT)	/record/master-panel	"Radiology Examination"
RO.2.1.2.1 (RC)	Content/Item	-	-
RC.1	Archetype_id {For this transaction assume that /record/row-type = "R"}	"gehr.cont.observe.radiology_details"	"gehr.cont.observe.radiology_details"
RC.2.1	Concept/value	"Radiological logical information apart from clinical information"	"Radiological logical information apart from clinical information"

RC.3	Proposition (xsi:type="G1_TREE_PROPOSITION")	-	-
R.3	Context	-	-
R.3.1	Recorder	"unknown"	"unknown"
R.3.2	Comment/value	Null	Null
R.3.3	Observation_time/value	/record/collection-time	"12/05/1997 14:37:00"
R.3.4	Activity_time/value	/record/collection-time	"12/05/1997 14:37:00"
R.3.5	Location	Null	Null
R.4 (RP)	Protocol (xsi:type="G1_LIST_PROPOSITION")	-	-
RP.1	Name/value	"Radiology Details Protocol"	"Radiology Details Protocol"
RP.2	Item (xsi:type="G1_HIERARCHICAL_VALUE")		-
RP.2.1.1	Name/value	"Service"	"Service"
RP.2.2.1	Value/value (xsi:type="G1_PLAIN_TEXT")	/record/lab-name	"RAH-RAD"
RP.3	Item (xsi:type="G1_HIERARCHICAL_VALUE")	-	-
RP.3.1.1	Name/value	"Laboratory ID"	"Laboratory ID"
RP.3.2.1	Value/value (xsi:type="G1_PLAIN_TEXT")	/record/lab-number	"97-027728"
RP.4	Item (xsi:type="G1_HIERARCHICAL_VALUE")	-	-
RP.4.1.1	Name/value	"Requested Time"	"Requested Time"
RP.4.2.1	Value/value (xsi:type="G1_DATE_TIME")	/record/requested-time	"12/05/1997"
RO.2.1.3.1 (ROS)	Organisers/Item	-	-
ROS.1.1	Name/value (PLAIN_TEXT)	"Findings"	"Findings"
ROS.2.1	Content/Item	-	-
ROS.2.1.1	Archetype_id {For this transaction assume that /record/row-type = "R"}	"gehr.cont.subj.clinical_information"	"gehr.cont.subj.clinical_information"
ROS.2.1.2.1	Concept/value	"Clinical information that are to be passed as a summary for the purposes of a test being carried out or a report being returned."	"Clinical information that are to be passed as a summary for the purposes of a test being carried out or a report being returned."
ROS.2.1.3 (RT)	Proposition (xsi:type="G1_TREE_PROPOSITION")	-	-

	ON")		
RT.1.1	Name/value	"Clinical Information"	"Clinical Information"
RT.2	Root (xsi:type="G1_HIERARCHICAL_GROUP")	-	-
RT.2.1.1	Name/value	"Clinical Information"	"Clinical Information"
RT.2.2	Children	-	-
RT.2.2.1 (RTC)	Item (xsi:type="G1_HIERARCHICAL_VALUE")		-
RTC.1.1	Name/value	"Comment"	"Comment"
RTC.2.1	Value/value (xsi:type="G1_PARAGRAPH")	-	-
RTC.2.1.1.1 (RTCP)	Items/item (xsi:type="PARAGRAPH_ITEM")	-	-
RTCP.1.1	Items/item (xsi:type="PLAIN_TEXT_LIST")	-	-
RTCP.1.1.1	Value	/record/comment/field	"nodosa. Dry cough febrile."
ROS.2.1.4 (RS)	Context	-	-
RS.1	Provider	/record/reported-by	"WALLACE,M"
RS.2	Provider_Relationship	"HCP"	"HCP"
RS.3	Time_Provided	/record/comment/field where [@type] = "Verified"	15/03/2001 9:16:35
ROS.2.1.5	Subject	"Self"	"Self"
RO.3	Archetype_id	"gehr.org.radiology_report"	"gehr.org.radiology_report"
R.2.2.4	Is_persistent	-	-
R.2.2.5	Persistence_Status	-	-

The mapping shown above references the following coded term tables.

Common Biochemistry Elements Lookup Table

Common Biochemistry Elements (Serum)	LOINC	Notes on LOINC	UMLS CUIs
Sodium	2951-2		C0037473
Potassium	2823-3		C0032821
Chloride	2075-0		C0008203
Bicarbonate	1963-8		C0005367
Anion Gap	1863-0	Anion gap 4 (there is a three as well)	C0003074
Glucose	14749-6	Random	C0017725
Urea	22664-7		C0041942
Creatinine	2160-0		C0010294

Urea/Creatinine Ratio	3097-3	This is point, there is a 24H as well for urine	C0365243
Urate	14933-6		C0041980
Phosphate	2777-1		C0031603
Calcium	2000-8	Total	C0006675
Ionised Calcium	1995-0		C0373561
Albumin	1751-7		C0001924
Globulins	2336-6	Many others!	C0017649
Total Protein	2885-2		C0555903
Total Bilirubin	1975-2		C0368753
Gamma Glutamyl Transpeptidase	2324-2		C0017040
Alkaline phosphatase	6768-6		C0002059
Alanine Aminotransferase	1742-6		C0001899
Aspartate aminotransferase	1920-8		C0004002
Lactate dehydrogenase	2532-0	There are many fractions available	C0022917

Units-Property Lookup Table

Units	Property
mmol/L	Concentration
umol/L	Concentration
U/L	Concentration
G/L	Concentration
meq/L	Concentration

6.3 Problems and Issues

Assuming that the inconsistencies and errors in the data have been filtered out as much as is possible in the pre-processing stage, there were still a number of issues, which required handling with respect to the mapping from generic XML into the GEHR XML extract. These issues are described in this section.

Terminology

One problem faced in mapping the data into a standard format was that standard term sets and/or standard terminology was not used in the extracted OACIS data. This was a problem in first, standardising how the data should be mapped (particularly in the case of comment types), and secondly in categorising the data values into coded term sets.

Interim-versus-Final Test Results

Another issue that needed consideration in the mapping process, was how to handle repeated test results – that is, test results with the same laboratory identifier (lab-number), which may

be marked as either “I” (Interim) or “F” (Final). It was decided that the combination of ‘lab-number’ and ‘mrn’ were adequate to identify repeated sets of test results, and that test results from rows with the same values for these fields would be mapped into different versions of the appropriate transactions from the same extract. In the case of Biochemistry test results, the reason for each new Version was recorded as either “Interim Version” (in the case where /record/data/status = “I”) or “Final Version” (in the case where /record/data/status = “F”). Some confusion, however, is created by the fact that two rows with the same ‘lab-number’ and ‘mrn’ may both be ‘Final’. However, this was not considered to be a major issue for the mapping.

Grouping Items

It is interesting to notice that in the microbiology archetype antibiotics are grouped by organism name (in the ‘Culture’ group), but that in the original OACIS data each antibiotic used was contained in a separate test result (leading to repeating organism-names). It was therefore necessary to group the test results by organism, so that all antibiotics used on the same organism were grouped together. While this is easily done using XSLT, it is a case that needs to be taken into consideration, when building a generic mapping tool to automate the translation script generation.

Mapping the Microbiology Comment Fields

In the case of Microbiology rows, much of the information content 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 extract. The rest of the comments were mapped to the transactions’ 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. However as previously discussed, 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.

Comment types, which were mapped into specific GEHR fields, include “MICROSCOPY”, “MACROSCOPY”, “CULTURE”, “SPECIMEN_SITE”, “SPECIMEN”, “Reported By” and “Verified” (the last two being for Radiology). Because the interpretation of these terms was not standardised by the laboratories, however, there is no guarantee that our interpretation of these comment types is indeed correct. Also, it is quite possible that due to the limited size of the sample data set we reviewed, that there are key comment types, which could have been mapped, that were not. To address these issues an agreement would need to be reached by the laboratories on the standardised interpretation of a basic set of comment types used in their data.

7. Using Mappings to Translate XML Data

7.1 Overview of Process

Once a mapping has been defined between the OACIS data and the fields in the target GEHR EHR extract, this mapping can be used to generate an XSLT script to automatically transform the generic XML data into GEHR-compliant data conforming to the specified clinical archetypes.

For the purposes of this project, the XSLT script generation was done manually. However, to make the development process as automated as possible, it is recommended that a mapping tool be developed to automatically generate appropriate XSLT scripts for different categories of data that require transformation. Defining the mapping between the OACIS data fields and the resulting GEHR extract fields, however, is necessarily a manual process.

The XML data generated in this process is a set of GEHR EHR extracts, which are ready for exchange between EHR systems, and for direct import into a GEHR kernel.

7.2 XSLT Script

Below is a small segment of the XSLT script used to transform generic XML data into GEHR EHR extracts:

```
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
  xmlns="http://www.gehr.org/namespace/GOM/" version="1.0">
  <xsl:output method="xml" encoding="UTF-8" indent="yes" />
  <!-- . . . . . -->
  <!-- Extract -->
  <xsl:template name="Extract">
    <xsl:element name="ehr_source">
      <xsl:value-of select="$ehr_source" />
    </xsl:element>
    <xsl:element name="creation_time">
      <xsl:call-template name="G1_GEHR_DATE_TIME">
        <xsl:with-param name="value" select="$extract_creation_time" />
      </xsl:call-template>
    </xsl:element>
    <xsl:element name="hcp_created_by">
      <xsl:value-of select="$extract_hca_created_by" />
    </xsl:element>
    <xsl:element name="destination_ehr_source">
      <xsl:value-of select="$extract_destination_ehr_source" />
    </xsl:element>
    <xsl:element name="hcp_authorising_acquisition">
```

```

        <xsl:value-of select="$extract_hca_authorising_acquisition" />
    </xsl:element>
    <xsl:element name="subject">
        <xsl:call-template name="SubjectVersionedTransaction" />
    </xsl:element>
    <!-- Create event_clinical_transactions if row type is known -->
    <xsl:if test="contains($VALID_ROW_TYPES, row_type)">
        <xsl:element name="event_clinical_transactions">
            <xsl:element name="item">
                <xsl:call-template name="LabTestVersionedTransaction" />
            </xsl:element>
        </xsl:element>
    </xsl:if>
</xsl:template>
<! . . . . . >

```

where \$ehr_source = "", \$extract_creation_time = "", \$extract_hca_created_by = "Zar Zar Tun", \$extract_destination_ehr_source = "DSTC_1" and \$extract_hca_authorising_acquisition = Scott Germann".

This code segment introduces the XSLT style sheet, defines the values of the first five GEHR extract fields (ehr_source to hcp_authorising_acquisition) and creates a new event clinical transaction if the row-type of the current row is valid (ie. "B", "M" or "R").

7.3 GEHR-compliant XML

In this section we present three examples of the XML that is generated by the XSLT script, which transforms the generic XML data from OACIS into a GEHR EHR extract.

7.3.1 Biochemistry

An example of an XML GEHR extract representing a row of Biochemistry OACIS data is as follows:

```

<?xml version="1.0" encoding="UTF-8"?>
<ehr_extract xmlns="http://www.gehr.org/namespace/GOM/">
    <ehr_source>OACIS</ehr_source>
    <creation_time><value>15/03/2001 9:15:00</value> </creation_time>
    <hcp_created_by>Dr Sam Heard</hcp_created_by>
    <destination_ehr_source>DSTC_1</destination_ehr_source>
    <hcp_authorising_acquisition>Scott Germann</hcp_authorising_acquisition>
    <subject>
        <time_created><value>15/03/2001 9:15:01</value></time_created>
        <versions><item>
            <audit>
                <hca_authorising>Scott Germann</hca_authorising>
            </audit>
        </item>
    </subject>

```

```

<time><value>15/03/2001 9:15:02</value></time>
<reason>Initial Version</reason>
<ehr_source_id>OACIS</ehr_source_id> <language>EN</language>
<territory>AUS</territory> <code_set>ISO8859-1</code_set>
</audit>
<content>
  <archetype_id>gehr.trans-persist.patient-demographics</archetype_id>
  <concept><value>Patient Demographics</value></concept>
  <content>
    <archetype_id>gehr.org.patient-identity</archetype_id>
    <concept><value>Patient Identity</value></concept>
    <name><value>Patient Identity</value></name>
    <content><item>
      <archetype_id>gehr.cont-define.patient-identity</archetype_id>
      <concept><value>Patient Identity</value></concept>
      <proposition xsi:type="G1_TREE_PROPOSITION" xmlns:xsi="
        http://www.w3.org/2000/10/XMLSchema-instance">
        <name><value>Patient Identity</value></name>
        <item xsi:type="G1_HIERARCHICAL_GROUP">
          <name><value>Patient Identity</value></name>
          <children>
            <item xsi:type="G1_HIERARCHICAL_VALUE">
              <name><value>EHR Source Patient ID</value></name>
              <value xsi:type="G1_PLAIN_TEXT">
                <value>001054</value> </value></item>
            <item xsi:type="G1_HIERARCHICAL_VALUE">
              <name><value>First Name</value></name>
              <value xsi:type="G1_PLAIN_TEXT">
                <value>Paul</value></value></item>
            <item xsi:type="G1_HIERARCHICAL_VALUE">
              <name><value>Last Name</value></name>
              <value xsi:type="G1_PLAIN_TEXT">
                <value>Duncan</value></value></item>
            <item xsi:type="G1_HIERARCHICAL_VALUE">
              <name><value>Date of Birth</value></name>
              <value xsi:type="G1_GEHR_DATE">
                <value>3/08/1923</value></value></item>
            <item xsi:type="G1_HIERARCHICAL_VALUE">
              <name><value>Sex</value></name>
              <value xsi:type="G1_PLAIN_TEXT">
                <value>male</value></value></item>
          </children></item></proposition></item> </content>
    </content>
  </content>
  <is_persistent>TRUE</is_persistent>
  <persistence_status>Pers_active</persistence_status>
</content>

```



```

    <reference><normal_range><lower>137</lower><upper>145
      </upper></normal_range></reference>
  </item>
  <item xsi:type="G1_HIERARCHICAL_VALUE">
    <name xsi:type="G1_TERM_TEXT"><value>Potassium
      </value><primary><umls_entry><cul>C0032821</cul>
      </umls_entry></primary></name>
    <value xsi:type="G1_QUANTITY"><value>4.3</value>
      <property>Concentration</property><units>mmol/L
      </units></value>
    <reference><normal_range><lower>3.1</lower><upper>4.2
      </upper></normal_range></reference>
  </item>
  <item xsi:type="G1_HIERARCHICAL_VALUE">
    <name xsi:type="G1_TERM_TEXT"><value>Glucose
      </value><primary><umls_entry><cul>C0017725</cul>
      </umls_entry></primary></name>
    <value xsi:type="G1_QUANTITY"><value>4.4</value>
      <property>Concentration</property><units>mmol/L
      </units></value>
    <reference><normal_range><lower>3.8</lower><upper>5.5</
      upper></normal_range></reference>
  </item></children></item>
  <item xsi:type="G1_HIERARCHICAL_VALUE">
    <name><value>Comment</value></name>
    <value xsi:type="G1_PLAIN_TEXT"><value>Plasma</value>
    </value></item>
  <item xsi:type="G1_HIERARCHICAL_GROUP">
    <name><value>Context</value></name>
    <children>
      <item xsi:type="G1_HIERARCHICAL_GROUP">
        <name><value>Challenge</value></name>
        <children>
          <item xsi:type="G1_HIERARCHICAL_VALUE">
            <name><value>Action</value></name>
            <value xsi:type="G1_TERM_TEXT"><primary>
              <umls_entry><cul>C0015633</cul></umls_entry>
              </primary></value></item></children>
          </item></children>
        </item></children></item></proposition>
    <context>
      <recorder>unknown</recorder>
      <observation_time><value>10/06/1997 9:15:00</value>
      </observation_time>
      <activity_time><value>10/06/1997</value></activity_time>
    </context>

```

```

<subject>
  <value>Self</value>
</subject>
<protocol xsi:type="G1_LIST_PROPOSITION">
  <name><value>Biochemistry Test Protocol</value></name>
  <item xsi:type="G1_HIERARCHICAL_GROUP">
    <name><value>Biochemistry Test Protocol</value></name>
    <children>
      <item xsi:type="G1_HIERARCHICAL_VALUE">
        <name><value>Laboratory</value></name>
        <value xsi:type="G1_PLAIN_TEXT"><value>RAH-
          LAB</value></value></item>
      <item xsi:type="G1_HIERARCHICAL_VALUE">
        <name><value>Laboratory ID</value></name>
        <value xsi:type="G1_PLAIN_TEXT"><value>2193686CHEM
          </value></value></item>
      <item xsi:type="G1_HIERARCHICAL_VALUE">
        <name><value>Requested Time</value></name>
        <value xsi:type="G1_PLAIN_TEXT"><value>10/06/1997
          </value></value></item></children></item>
    </protocol></item></content></content>
    <is_persistent>False</is_persistent></content></item></versions></item>
  </event_clinical_transactions></ehr_extract>

```

7.3.2 Microbiology

An example of an XML GEHR extract representing a row of Microbiology OACIS data is as follows:

```

<?xml version="1.0" encoding="UTF-8"?>
<ehr_extract xmlns="http://www.gehr.org/namespace/GOM/">
  <ehr_source>OACIS</ehr_source>
  <creation_time><value>15/03/2001 9:15:00</value> </creation_time>
  <hcp_created_by>Dr Sam Heard</hcp_created_by>
  <destination_ehr_source>DSTC_1</destination_ehr_source>
  <hcp_authorising_acquisition>Scott Germann</hcp_authorising_acquisition>
  <subject>
    <time_created><value>15/03/2001 9:15:01</value></time_created>
    <versions><item>
      <audit>
        <hca_authorising>Scott Germann</hca_authorising>
        <time><value>15/03/2001 9:15:02</value></time>
        <reason>Initial Version</reason>
        <ehr_source_id>OACIS</ehr_source_id> <language>EN</language>
        <territory>AUS</territory> <code_set>ISO8859-1</code_set>
      </audit>
    </content>
  </subject>

```

```

<archetype_id>gehr.trans-persist.patient-demographics</archetype_id>
<concept><value>Patient Demographics</value></concept>
<content>
  <archetype_id>gehr.org.patient-identity</archetype_id>
  <concept><value>Patient Identity</value></concept>
  <name><value>Patient Identity</value></name>
  <content><item>
    <archetype_id>gehr.cont-define.patient-identity</archetype_id>
    <concept><value>Patient Identity</value></concept>
    <proposition xsi:type="G1_TREE_PROPOSITION" xmlns:xsi="
      http://www.w3.org/2000/10/XMLSchema-instance">
      <name><value>Patient Identity</value></name>
      <item xsi:type="G1_HIERARCHICAL_GROUP">
        <name><value>Patient Identity</value></name>
        <children>
          <item xsi:type="G1_HIERARCHICAL_VALUE">
            <name><value>EHR Source Patient ID</value></name>
            <value xsi:type="G1_PLAIN_TEXT">
              <value>001076</value> </value></item>
          <item xsi:type="G1_HIERARCHICAL_VALUE">
            <name><value>First Name</value></name>
            <value xsi:type="G1_PLAIN_TEXT">
              <value>Tom</value></value></item>
          <item xsi:type="G1_HIERARCHICAL_VALUE">
            <name><value>Last Name</value></name>
            <value xsi:type="G1_PLAIN_TEXT">
              <value>Davenport</value></value></item>
          <item xsi:type="G1_HIERARCHICAL_VALUE">
            <name><value>Date of Birth</value></name>
            <value xsi:type="G1_GEHR_DATE">
              <value>18/08/1953</value></value></item>
          <item xsi:type="G1_HIERARCHICAL_VALUE">
            <name><value>Sex</value></name>
            <value xsi:type="G1_PLAIN_TEXT">
              <value>male</value></value></item>
        </children></item></proposition></item> </content>
    </content>
    <is_persistent>TRUE</is_persistent>
    <persistence_status>Pers_active</persistence_status>
  </content>
</item></versions>
</subject>
<event_clinical_transactions>
  <item>
    <time_created><value>15/03/2001 9:15:05</value></time_created>
    <versions><item>

```

```

<audit>
  <hca_authorising>Scott Germann</hca_authorising>
  <time><value>15/03/2001 9:15:06</value></time>
  <reason>Final Version</reason><ehr_source_id>OACIS</ehr_source_id>
  <language>EN</language>
</audit>
<content>
  <archetype_id>gehr.trans-persist.patient-contact</archetype_id>
  <concept><value>Patient Contact</value></concept>
  <content>
    <archetype_id>gehr.org.test-result</archetype_id>
    <concept><value>Test Result</value></concept>
    <name><value>Test Result</value></name>
    <content>
      <item xsi:type="G1_OBSERVATION_CONTENT"
        xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance">
        <archetype_id>gehr.cont-observe.microbiology</archetype_id>
        <concept><value>Laboratory findings and comments for microbiology
          tests</value></concept>
        <proposition xsi:type="G1_TREE_PROPOSITION">
          <name><value>Microbiology Test Result</value></name>
          <item xsi:type="G1_HIERARCHICAL_GROUP">
            <name><value>Microbiology Test Result</value></name>
            <children>
              <item xsi:type="G1_HIERARCHICAL_VALUE">
                <name><value>Battery Name</value></name>
                <value xsi:type="G1_PLAIN_TEXT"><value>URINE -
                  MICROBIOLOGY</value></value></item>
              <item xsi:type="G1_HIERARCHICAL_GROUP">
                <name><value>Specimen collection</value></name>
                <children>
                  <item xsi:type="G1_HIERARCHICAL_VALUE">
                    <name><value>Site</value></name>
                    <value xsi:type="G1_PLAIN_TEXT"> <value>Midstream
                      </value></value></item></children></item>
                <item xsi:type="G1_HIERARCHICAL_GROUP">
                  <name><value>Culture</value></name>
                  <children>
                    <item xsi:type="G1_HIERARCHICAL_GROUP">
                      <name><value>Micro-organism identified</value></name>
                      <children>
                        <item xsi:type="G1_HIERARCHICAL_VALUE">
                          <name><value>Micro-organism</value></name>
                          <value xsi:type="G1_PLAIN_TEXT">
                            <value>PSA</value> </value></item>
                        <item xsi:type="G1_HIERARCHICAL_GROUP">

```

```

<name><value>Therapeutic susceptibilities</value>
  </name>
<children>
  <item xsi:type="G1_HIERARCHICAL_GROUP">
    <name><value>GENTAMICIN</value></name>
    <children>
      <item xsi:type="G1_HIERARCHICAL_VALUE">
        <name><value>Susceptibility description</value>
          </name>
          <value xsi:type="G1_PLAIN_TEXT">
            <value>Susceptible</value> </value></item>
          </children></item>
        <item xsi:type="G1_HIERARCHICAL_GROUP">
          <name><value>NORFLOXACIN</value></name>
          <children>
            <item xsi:type="G1_HIERARCHICAL_VALUE">
              <name><value>Susceptibility description</value>
                </name>
                <value xsi:type="G1_PLAIN_TEXT"> <value>
                  Resistant</value></value></item></children>
              </item></children></item></children></item>
            <item xsi:type="G1_HIERARCHICAL_VALUE">
              <name><value>Overall comment</value></name>
              <value xsi:type="G1_PLAIN_TEXT"><value>SPECIMEN
                SITE = Midstream</value></value></item>
            <item xsi:type="G1_HIERARCHICAL_VALUE">
              <name><value>Overall comment</value></name>
              <value xsi:type="G1_PLAIN_TEXT"><value>WHITE
                BLOOD CELLS (N&lt;10) = &gt;1000</value></value>
            </item></children></item></children></item>
          </proposition>
        <context>
          <recorder>unknown</recorder>
          <observation_time><value>2/11/1998 09:18:00</value>
            </observation_time>
          <activity_time><value>2/11/1998 08:35:00</value></activity_time>
        </context>
        <subject>
          <value>Self</value>
        </subject>
        <protocol xsi:type="G1_LIST_PROPOSITION">
          <name><value>Microbiology Test Protocol</value></name>
          <item xsi:type="G1_HIERARCHICAL_GROUP">
            <name><value>Microbiology Test Protocol</value></name>
            <children>
              <item xsi:type="G1_HIERARCHICAL_VALUE">

```

```

        <name><value>Laboratory</value></name>
        <value xsi:type="G1_PLAIN_TEXT"><value>QEH-
          LAB</value></item>
      <item xsi:type="G1_HIERARCHICAL_VALUE">
        <name><value>Laboratory ID</value></name>
        <value xsi:type="G1_PLAIN_TEXT"><value>MB187959UR
          </value></value></item>
      <item xsi:type="G1_HIERARCHICAL_VALUE">
        <name><value>Requested Time</value></name>
        <value xsi:type="G1_PLAIN_TEXT"><value>2/11/1998
          </value></value></item></children></item>
    </protocol></item></content></content>
    <is_persistent>False</is_persistent></content></item></versions></item>
  </event_clinical_transactions></ehr_extract>

```

7.3.1 Radiology

An example of an XML GEHR extract representing a row of Microbiology OACIS data is as follows:

```

<?xml version="1.0" encoding="UTF-8" ?>
<ehr_extract xmlns="http://www.gehr.org/namespace/GOM/">
  <ehr_source>OACIS</ehr_source>
  <creation_time><value>24/05/2001 16:54:37</value></creation_time>
  <hcp_created_by>Dr Sam Heard</hcp_created_by>
  <destination_ehr_source>DSTC_1</destination_ehr_source>
  <hcp_authorising_acquisition>Scott Germann</hcp_authorising_acquisition>
  <subject>
    <time_created><value>24/05/2001 16:54:37</value></time_created>
    <versions><item>
      <audit>
        <hca_authorising>Scott Germann</hca_authorising>
        <time><value>24/05/2001 16:54:37</value></time>
        <reason>Initial Version</reason><ehr_source_id>OACIS</ehr_source_id>
        <language>EN</language><territory>AUS</territory>
        <code_set>ISO8859-1</code_set> </audit>
      <content>
        <archetype_id>gehr.trans-persist.patient-demographics</archetype_id>
        <concept><value>Patient Demographics</value></concept>
        <content>
          <name><value>Patient Identity</value></name>
          <content><item>
            <archetype_id>gehr.cont-define.patient-identity</archetype_id>
            <concept><value>Patient Identity</value></concept>
            <proposition xsi:type="G1_LIST_PROPOSITION" xmlns:xsi="
              http://www.w3.org/2000/10/XMLSchema-instance">
              <name><value>Patient Identity</value></name>

```

```

<item xsi:type="G1_HIERARCHICAL_GROUP">
  <name><value>Patient Identity</value></name>
  <children>
    <item xsi:type="G1_HIERARCHICAL_VALUE">
      <name><value>EHR Source Patient ID</value></name>
      <value xsi:type="G1_PLAIN_TEXT"><value>001054
        </value></value></item>
    <item xsi:type="G1_HIERARCHICAL_VALUE">
      <name><value>First Name</value></name>
      <value xsi:type="G1_PLAIN_TEXT"><value>Paul
        </value></value></item>
    <item xsi:type="G1_HIERARCHICAL_VALUE">
      <name><value>Last Name</value></name>
      <value xsi:type="G1_PLAIN_TEXT"><value>Duncan
        </value></value></item>
    <item xsi:type="G1_HIERARCHICAL_VALUE">
      <name><value>Date of Birth</value></name>
      <value xsi:type="G1_GEHR_DATE"><value>3/08/1923
        </value></value></item>
    <item xsi:type="G1_HIERARCHICAL_VALUE">
      <name><value>Sex</value></name>
      <value xsi:type="G1_PLAIN_TEXT"><value>male
        </value></value></item>
  </children></item></proposition></item></content>
  <archetype_id>gehr.org.patient-identity</archetype_id>
  <concept><value>Patient Identity</value></concept></content>
  <is_persistent>TRUE</is_persistent>
  <persistence_status>Pers_active</persistence_status>
</content>
</item></versions></subject>
<event_clinical_transactions>
  <item>
    <time_created>
      <value>05/06/1997 16:14</value>
    </time_created>
    <versions><item>
      <audit>
        <hca_authorising>WALLACE,MA</hca_authorising>
        <time><value>24/05/2001 16:54:37</value></time>
        <reason>Final Version</reason> <ehr_source_id>OACIS</ehr_source_id>
        <language>EN</language><territory>AUS</territory>
        <code_set>ISO8859-1</code_set></audit>
      <content>
        <archetype_id>gehr.trans-event.radiology</archetype_id>
        <concept><value>Radiology Report</value></concept>
      </content>
    </item>
  </versions>
</item>

```



```

<name><value>Report</value></name>
<organisers>
  <item>
    <name><value>Radiology Examination</value></name>
    <content>
      <item xsi:type="G1_OBSERVATION_CONTENT" xmlns:xsi="
        http://www.w3.org/2000/10/XMLSchema-instance">
        <archetype_id>gehr.cont-observe.radiology_details </archetype_id>
        <concept><value>Radiological Information</value> </concept>
        <context>
          <recorder>Unknown</recorder>
          <observation_time><value>5/06/1997 14:11:00</value>
            </observation_time>
          <activity_time><value>5/06/1997 14:11:00</value>
            </activity_time></context>
        <subject><value>Self</value></subject>
        <protocol xsi:type="G1_LIST_PROPOSITION">
          <name><value>Radiology Details Protocol</value></name>
          <item xsi:type="G1_HIERARCHICAL_GROUP">
            <name><value>Radiology Details Protocol</value></name>
            <children>
              <item xsi:type="G1_HIERARCHICAL_VALUE">
                <name><value>Service</value></name>
                <value xsi:type="G1_PLAIN_TEXT"><value>RAH-
                  RAD</value></value></item>
              <item xsi:type="G1_HIERARCHICAL_VALUE">
                <name><value>Laboratory ID</value></name>
                <value xsi:type="G1_PLAIN_TEXT"><value>97-
                  033119</value></value></item>
              <item xsi:type="G1_HIERARCHICAL_VALUE">
                <name><value>Requested Time</value></name>
                <value xsi:type="G1_GEHR_DATE_TIME">
                  <value>5/06/1997</value></value></item>
            </children></item></protocol></item></content>
        </organisers>
        <item>
          <name><value>Findings</value></name>
          <content>
            <item xsi:type="G1_SUBJECTIVE_CONTENT" xmlns:xsi="
              http://www.w3.org/2000/10/XMLSchema-instance">
              <archetype_id>gehr.cont-subj.clinical_information
                </archetype_id>
              <concept><value>Clinical Information</value></concept>
              <proposition xsi:type="G1_TREE_PROPOSITION">
                <name><value>Clinical Information</value></name>
                <item xsi:type="G1_HIERARCHICAL_GROUP">

```

```

<name><value>Clinical Information</value></name>
<children>
  <item xsi:type="G1_HIERARCHICAL_VALUE">
    <name><value>Comment</value></name>
    <value xsi:type="G1_PLAIN_TEXT"><value>Reported
      by WALLACE,MA Verified 05/06/1997 16:14
      CLINICAL DETAILS: C/ diffuse enteritis, recent RIH
      repair. Tender abdomen. Continuing diarrhoea/
      anorexia. exclude 1. bowel obstruction. 2. Exclude
      megacolon. _~ABDOMEN_~ (33119) (23)
      Calcification is noted in the abdominal aorta which is
      tortuous and aneurysmal particularly in the L1 to L4
      region. There is a slight scoliosis concave to the left
      centered on T11 with secondary scoliosis concave to
      the right centered on L3. Degenerative change is noted
      in the lower lumbar spine associated with the
      scoliosis.</value></value></item></children>
  </item></proposition>
  <context>
    <provider>WALLACE,MA</provider>
    <provider_relationship><value>HCP</value>
    </provider_relationship>
    <time_provided><value>05/06/1997 16:14</value>
    </time_provided></context>
    <subject><value>Self</value></subject>
  </item></content></item></organisers></item></organisers>
  <archetype_id>gehr.org.radiology_report</archetype_id>
  <concept><value>Radiology Report</value></concept></content>
  <is_persistent>FALSE</is_persistent></content></item></versions>
</item></event_clinical_transactions></ehr_extract>

```

7.4 Problems and Issues

There were very few issues in performing this step of the transformation process. However, as previously mentioned, it would be highly desirable to automate the process of writing the XSLT script, based on the mapping from source to target data formats.

Another area of future improvement is in the area of coded term set mappings. For the purposes of this 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 XSLT script. However, it may be better for a number of reasons to write the XSLT script such that it directly references a live term set server. This would provide more up-to-date term translations, and be more scalable for very large term sets.

8. Importing and Viewing GEHR-XML

8.1 Overview of Process

Importing the GEHR-compliant EHR data into the GEHR kernel requires two steps to be performed. Firstly, all the appropriate GEHR archetype definitions (at the transaction, organiser and content levels) must be imported into the GEHR kernel to initialise the internal archetype models. Secondly, the GEHR XML extract itself must be imported, and in doing so validated against the archetype definitions that have previously been initialised. At this stage the data is ready for use within any clinical applications, which can communicate with the GEHR kernel.

To view the data imported into the GEHR kernel, a generic GEHR-XML to HTML transformation is performed (again using XSLT), allowing the formatted data to be viewed through a standard web browser. An 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.

8.2 Viewing the Data

In this section, we show how the GEHR XML extracts from Chapter 7 are displayed for viewing by the generic GEHR-extract viewer used for this project.

8.2.1 Biochemistry

The following screen shots illustrate how a row of Biochemistry data from OACIS is viewed through the generic instance viewer, after being transformed into a GEHR XML extract. This viewer has no transformation for clinical presentation at present but allows all the information stored in the transaction to be viewed. The patient information, for example, relates to the EHR as a whole and not to the individual transaction. Further, the UMLS concept unique identifiers are present to enable automatic processing, and would not normally be visible. The LOINC code is also displayed in some instances. These surrogates allow the mapping of HL7, translation to other languages and classification in a variety of systems such as ICD-10 and SNOMED.

The screenshot shows the 'EHR Instance Viewer' interface. The left sidebar contains a 'Contents' section with links for '[View EHR details...]', 'Subject' (Version dated 12/03/2011 9:00:00 [obj...]), 'Event Transactions' (Test Result [obj...]), and 'Test Result' (Version dated 12/03/2011 9:00:00 [obj...]). The main content area is titled 'Content "Patient Identity"' and shows a 'Concept "Patient Identity"' with an 'EHR Archetype in use' box containing 'EHR archetype ID: gdr:cont-idm:patient-identity' and 'Concept: Patient Identity'. Below this, a 'Group "Patient Identity"' section lists patient details: 'EHR Source Patient ID: 001024', 'First Name: Paul', 'Last Name: Duncan', 'Date of Birth: 1980/10/15', and 'Sex: male'.

Figure 5: Biochemistry Patient Data

The screenshot shows the 'EHR Instance Viewer' interface. The left sidebar contains a 'Contents' section with links for '[View EHR details...]', 'Subject' (Version dated 12/03/2011 9:00:00 [obj...]), 'Event Transactions' (Test Result [obj...]), and 'Test Result' (Version dated 12/03/2011 9:00:00 [obj...]). The main content area is titled 'Content "Biochemistry Test"' and shows a 'Concept "Biochemistry Test Result"' with an 'EHR Archetype in use' box containing 'EHR archetype ID: gdr:cont-obstrv-biochemistry' and 'Concept: Biochemistry Test Result'. Below this, a 'Group "Biochemistry Test"' section lists test details: 'Sample: Blood - unspecified - ELDP Blood - unspecified - LOINC', a 'Group "DIAGNOSTIC ANALYSIS"' section, and two test results: 'Urea' (145 - Concentration - mmol/L, UMLS Entry: C0037473, Normal range: 117 to 145) and 'Potassium' (4.3 - Concentration - mmol/L, UMLS Entry: C0032821, Normal range: 3.1 to 4.2).

Figure 6: Biochemistry Test Results

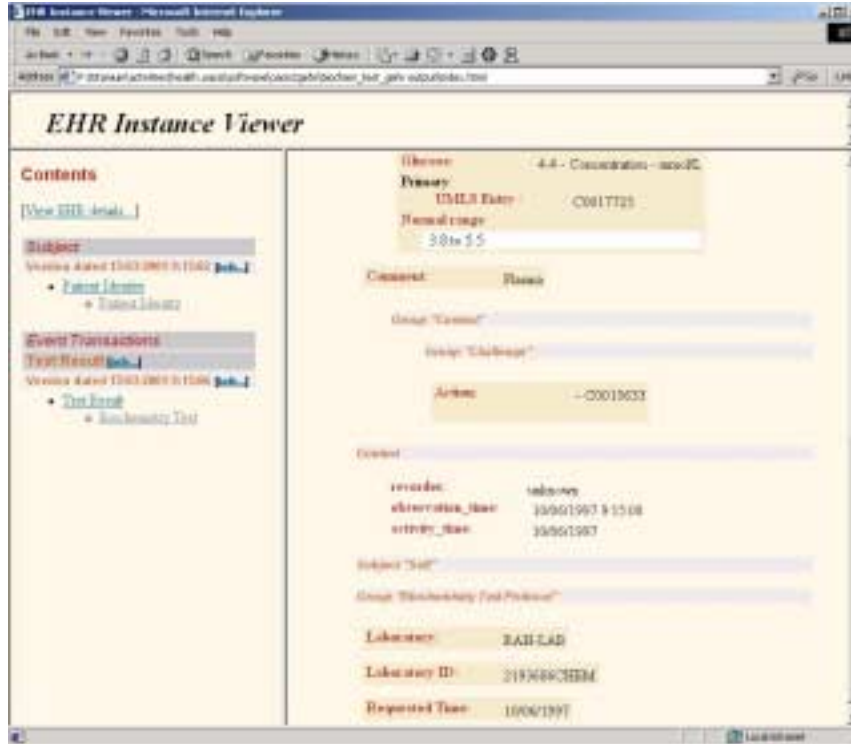


Figure 7: Biochemistry Test Results (More)

8.2.2 Microbiology

The following screen shots illustrate how a row of Microbiology data from OACIS would be viewed through the generic instance viewer, after being transformed into a GEHR XML extract.

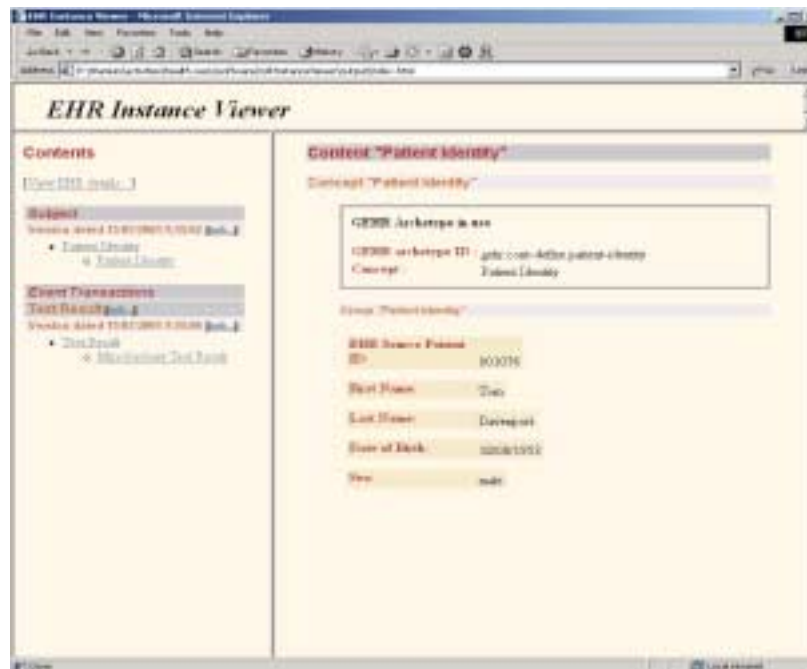


Figure 8: Microbiology Patient Data

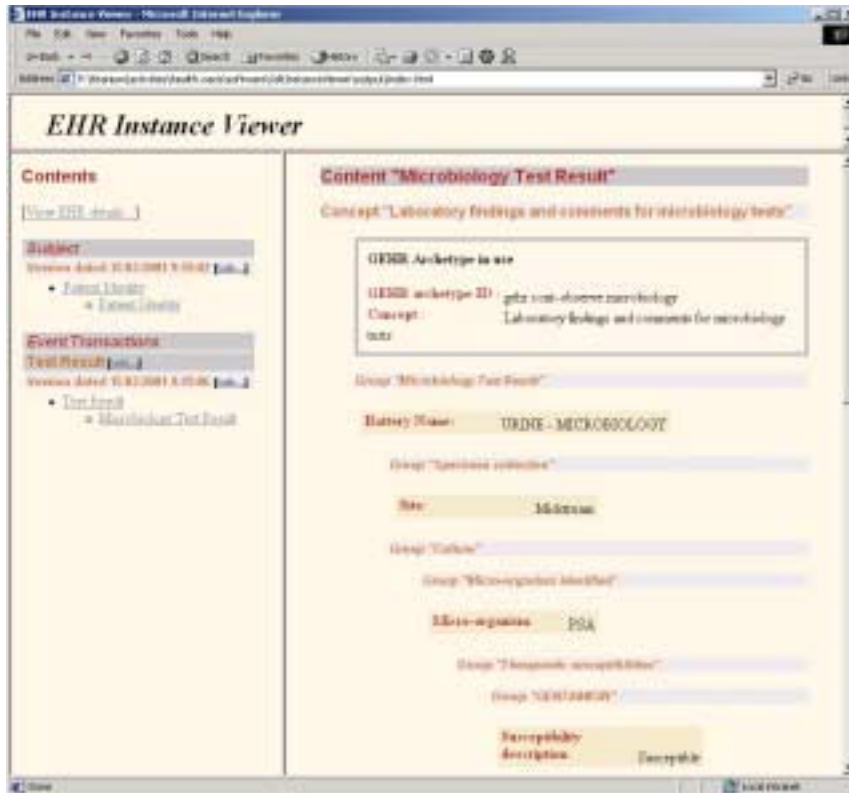


Figure 9: Microbiology Test Results

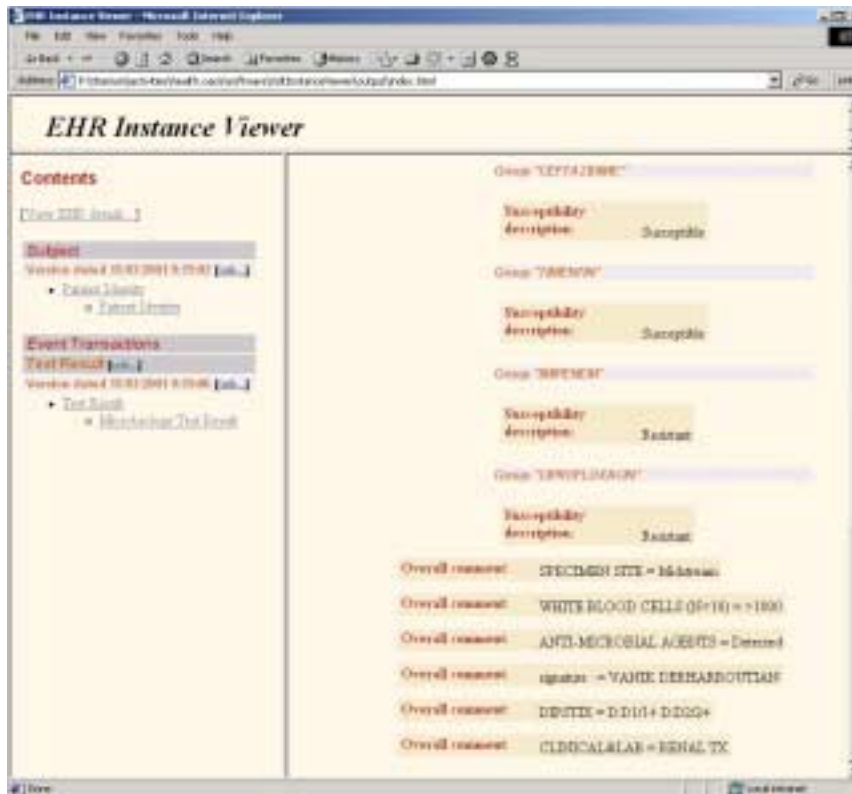


Figure 10: Microbiology Test Results (More)

8.2.3 Radiology

The following screen shots illustrate how a row of Radiology data from OACIS would be viewed through the generic instance viewer, after being transformed into a GEHR XML extract.

The screenshot displays the 'EHR Instance Viewer' interface. On the left, the 'Contents' pane shows a tree view with 'Subject' (Version: 2006-03-20 09:14:50Z) and 'Event Transactions' (Version: 2006-03-20 09:14:50Z). The 'Event Transactions' section is expanded to show 'Radiology Report'. The main content area, titled 'Content "Patient Identity"', shows the following data:

Concept: "Patient Identity"	
GEHR archetype name	
GEHR archetype ID:	gohr-coat-dfian-patient-identity
Concept:	Patient Identity
Group: "Patient Identity"	
EHR Source Patient	
ID:	001004
First Name:	Paul
Last Name:	Davies
Date of Birth:	19661923
Sex:	male

Figure 11: Radiology Patient Data

The screenshot displays the 'EHR Instance Viewer' interface. On the left, the 'Contents' pane shows a tree view with 'Subject' (Version: 2006-03-20 09:14:50Z) and 'Event Transactions' (Version: 2006-03-20 09:14:50Z). The 'Event Transactions' section is expanded to show 'Radiology Report'. The main content area, titled 'Content "Radiological Information"', shows the following data:

Concept: "Radiological Information"	
GEHR archetype name	
GEHR archetype ID:	gohr-coat-ovcrvs-radiology_details
Concept:	Radiological Information
Context:	
occurs:	Unknown
observation_time:	5061997 14:11:00
activity_time:	5061997 14:11:00
Subject "Paul"	
Group: "Radiology Details Protocol"	
Laboratory ID:	
	97-453119
Requested Time:	
	5061997

Figure 12: Radiology Information and Protocol

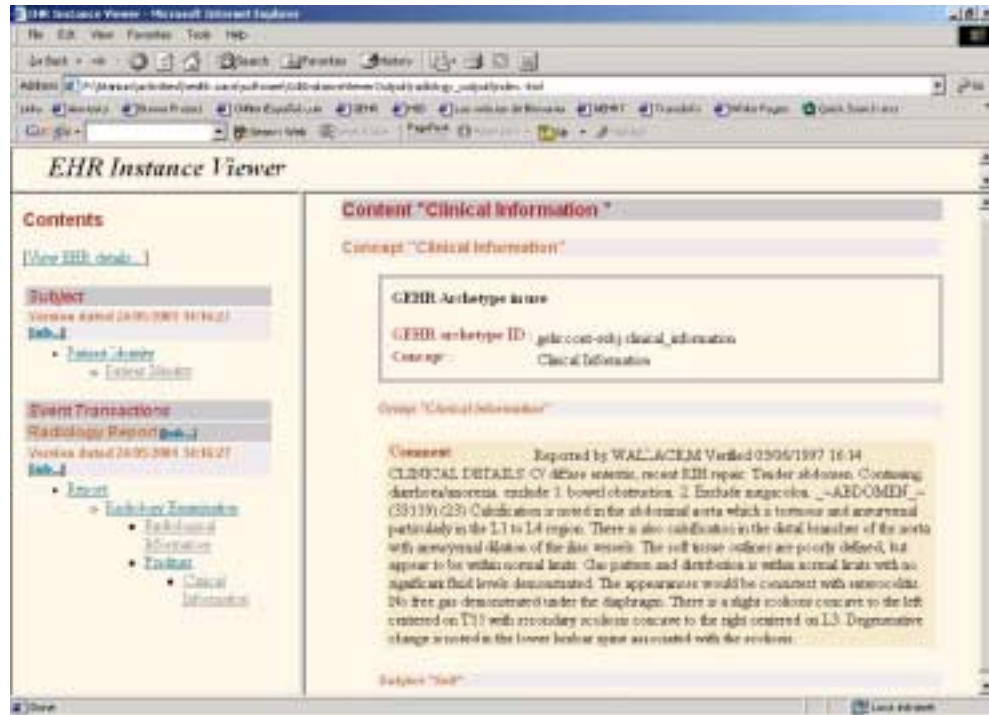


Figure 13: Radiology Clinical Information

8.3 Problems and Issues

The only real issue faced in importing the data into the GEHR kernel was that the format used for dates in the OACIS system needed to be defined in the GEHR kernel. Alternatively, the date could have been transformed into the default date format used in the kernel. However, this was seen as unnecessary.

With respect to viewing the data, there was a choice as to whether to view the distinct XML extracts generated from each row of OACIS data individually, or to display all the transactions related to a single patient together. It was decided that each extract exported from the OACIS system would represent individual sets of test results, but that once the data was imported into the GEHR kernel, that a single extract could be exported to describe all transactions relevant to each person.

9. Conclusions

The Hospital to GP communication between non-GEHR and GEHR-compliant systems has been completed in a short time frame. This process will allow all GEHR compliant systems to have access to the data. The transformation of data from OACIS to GEHR has been achieved with the maximum preservation of data and designed to work in a real time environment. Both ‘push’ and ‘pull’ scenarios should be catered for with the methodology provided. Importantly, if this information is added at the outset via the GEHR kernel then considerably more information will be available for automatic processing. Perhaps the most important information that has been hidden in text fields in an inconsistent manner in the OACIS data is the urine microbiology. It is clear that radiology reports vary very widely and are more like clinical reports in other settings than microbiology or biochemistry reports. No images are available for the radiology reports in OACIS.

Extracting the data from a large clinical database with heterogeneous data is challenging for all systems. The GEHR system has been designed to allow incorporation of health information without changing the internal data model and enabling generic processing of this information through the use of clinical models called GEHR archetypes. This project allowed the testing of the capability of GEHR to incorporate information from a very large clinical database, while assessing the ‘quality’ of that data for such a transformation.

The extraction of the raw data used the query capability of the OACIS system and generated records as delimited strings. The patient details were altered in this process. This text file consisted of one record per ‘report’ with the patient identification held within the record. 15000 such records were processed.

A Perl script (available from the GEHR website) was used to transform the raw data to a generic XML data format. XML was chosen as there is a growing toolset to work with data in this format and future applications are likely to be able to export in this format. This step has been labelled as pre-processing and is specific to the OACIS setting. Some important ‘interpretation’ is required within this process to preserve information. There is no limit to the extent of pre-processing that may be required to preserve as much information as possible – and while such pre-processing can be very complex and freeform in a single download setting which can be run and re-run to perfect the process– it must be limited to ‘safe’ transformations in a real time setting.

The GEHR archetypes developed for incorporation of the OACIS data were:

- Biochemistry
- Microbiology
- Radiology

These generic archetypes included requirements from the LOINC user manual for HL7 messages and the DICOM radiology standard. The project enabled the testing of the archetypes with real clinical data.

Having transformed the data to XML and developed the archetypes, a mapping process was developed to place each data element, its units and normal values, into the correct point in the GEHR record. The mapping tables express these links and are used by a standard tool to transform the generic XML into XML that is conformant with the GEHR Object Model (GOM). This process was carried out using the XML transformation engine (XSLT). The XSLT scripts are also available from the GEHR website.

This data was then imported into the GEHR kernel using a standard import tool and displayed in the generic GEHR data viewer developed by the DSTC.

The steps taken can be run in real time running the PERL script and XSLT script for each record or set of records received.

This project has enabled the development of a process to import data from a large clinical database into GEHR and potentially run this process in real time in a clinical setting. It has been broken down into a specific (to OACIS) pre-processing stage and a generic transformation and import phase which can be used by the GEHR kernel in many settings in the future. The later phase uses mapping tables and XSLT to achieve produce the GEHR format. This process is particularly suitable for incorporation of HL7 messages in the future.

The archetypes produced have endeavoured to include the requirements for incorporation of HL7 messages in the future and have to this point only been tested with the OACIS data. The complexity of the archetypes is more than was required for this project. No development of the archetypes was required when processing the large dataset.

This project has required skills in data extraction, text processing, XML and XML transformation, HL7 and LOINC and clinical practice as well as GEHR development. The team at Flinders University, DSTC, SA Health Commission and Ocean Informatics have brought these skills to the project and undertaken a complex task in a relatively short period of time.

Further development work will be required to utilise the GEHR kernel in a live situation accepting HL7 messages and text from hospital systems. Having incorporated the data into GEHR it should be available to all systems in a form suitable for safe processing. This work may be of interest to those developing an approach to Health Connect as well as communication between hospital and primary care.