ICT-2011-288048

EURECA

Enabling information re-Use by linking clinical Research and CAre

IP
Contract Nr: 288048

Deliverable: D2.3 Initial report on the interface layer towards external systems

Due date of deliverable: 07-31-2013
Actual submission date: (11-06-2013)

Start date of Project: 01 February 2012
Duration: 42 months

Responsible WP: Philips

Revision: <outline, draft, **proposed**, accepted>

<table>
<thead>
<tr>
<th>Dissemination level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PU</strong> Public</td>
<td>X</td>
</tr>
<tr>
<td><strong>PP</strong> Restricted to other programme participants (including the Commission Service)</td>
<td></td>
</tr>
<tr>
<td><strong>RE</strong> Restricted to a group specified by the consortium (including the Commission Services)</td>
<td></td>
</tr>
<tr>
<td><strong>CO</strong> Confidential, only for members of the consortium (excluding the Commission Services)</td>
<td></td>
</tr>
</tbody>
</table>
0 DOCUMENT INFO

0.1 Author

<table>
<thead>
<tr>
<th>Author</th>
<th>Company</th>
<th>E-mail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahmed Ibrahim</td>
<td>Philips</td>
<td><a href="mailto:Ahmed.a.ibrahim@philips.com">Ahmed.a.ibrahim@philips.com</a></td>
</tr>
<tr>
<td>Brecht Claerhout</td>
<td>Custodix</td>
<td><a href="mailto:brecht.claerhout@custodix.com">brecht.claerhout@custodix.com</a></td>
</tr>
<tr>
<td>Jasper van Leeuwen</td>
<td>Philips</td>
<td><a href="mailto:Jasper.van.leeuwen@philips.com">Jasper.van.leeuwen@philips.com</a></td>
</tr>
<tr>
<td>Kerstin Rohm</td>
<td>Fraunhofer</td>
<td><a href="mailto:kerstin.rohm@ibmt.fraunhofer.de">kerstin.rohm@ibmt.fraunhofer.de</a></td>
</tr>
<tr>
<td>Kristof De Schepper</td>
<td>Custodix</td>
<td><a href="mailto:kristof.deschepper@custodix.com">kristof.deschepper@custodix.com</a></td>
</tr>
<tr>
<td>Cédric Vansuyt</td>
<td>Custodix</td>
<td><a href="mailto:Cedric.vansuyt@custodix.com">Cedric.vansuyt@custodix.com</a></td>
</tr>
<tr>
<td>Raúl Alonso</td>
<td>UPM</td>
<td><a href="mailto:ralonso@infomed.dia.fi.upm.es">ralonso@infomed.dia.fi.upm.es</a></td>
</tr>
<tr>
<td>Santiago Aso</td>
<td>UPM</td>
<td><a href="mailto:saso@infomed.dia.fi.upm.es">saso@infomed.dia.fi.upm.es</a></td>
</tr>
<tr>
<td>Sergio Paraíso</td>
<td>UPM</td>
<td><a href="mailto:sparaiso@infomed.dia.fi.upm.es">sparaiso@infomed.dia.fi.upm.es</a></td>
</tr>
<tr>
<td>Haridimos Kondylakis</td>
<td>FORTH</td>
<td><a href="mailto:kondylak@ics.forth.gr">kondylak@ics.forth.gr</a></td>
</tr>
<tr>
<td>Yarema Mazuryk</td>
<td>Philips</td>
<td><a href="mailto:yarema.mazuryk@philips.com">yarema.mazuryk@philips.com</a></td>
</tr>
</tbody>
</table>

0.2 Documents history

<table>
<thead>
<tr>
<th>Document version #</th>
<th>Date</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>V0.1</td>
<td>19/06/2013</td>
<td>Starting version, template</td>
</tr>
<tr>
<td>V0.2</td>
<td>19/06/2013</td>
<td>Definition of ToC</td>
</tr>
<tr>
<td>V0.3</td>
<td></td>
<td>First complete draft</td>
</tr>
<tr>
<td>V0.4</td>
<td></td>
<td>Integrated version (send to WP members)</td>
</tr>
<tr>
<td>V0.5</td>
<td></td>
<td>Updated version (send PCP)</td>
</tr>
<tr>
<td>V0.6</td>
<td></td>
<td>Updated version (send to project internal reviewers)</td>
</tr>
<tr>
<td>Sign off</td>
<td></td>
<td>Signed off version (for approval to PMT members)</td>
</tr>
<tr>
<td>V1.0</td>
<td></td>
<td>Approved Version to be submitted to EU</td>
</tr>
</tbody>
</table>

0.3 Document data

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Interface layer, EHR, clinical data warehouse, trial management systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Editor Address data</td>
<td>Name: Ahmed Ibrahim</td>
</tr>
<tr>
<td></td>
<td>Partner: Philips Electronics Nederland BV</td>
</tr>
<tr>
<td></td>
<td>Address: High Tech Campus 34</td>
</tr>
<tr>
<td></td>
<td>Phone: +31-611 714 731</td>
</tr>
<tr>
<td></td>
<td>E-mail: <a href="mailto:ahmed.a.ibrahim@philips.com">ahmed.a.ibrahim@philips.com</a></td>
</tr>
<tr>
<td>Delivery date</td>
<td>11-06-2013</td>
</tr>
</tbody>
</table>
0.4 Distribution list

<table>
<thead>
<tr>
<th>Date</th>
<th>Issue</th>
<th>E-mailer</th>
</tr>
</thead>
<tbody>
<tr>
<td>06.11.2013</td>
<td>v1.0</td>
<td>'<a href="mailto:Eduardo.GONZALEZ-OTERO@ec.europa.eu">Eduardo.GONZALEZ-OTERO@ec.europa.eu</a>'</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:CNECT-ICT-288048@ec.europa.eu">CNECT-ICT-288048@ec.europa.eu</a></td>
</tr>
</tbody>
</table>

Table of Contents

0 DOCUMENT INFO ............................................................................................................. 2
  0.1 Author ..................................................................................................................... 2
  0.2 Documents history .................................................................................................. 2
  0.3 Document data ....................................................................................................... 2
  0.4 Distribution list .................................................................................................... 3

1 INTRODUCTION .............................................................................................................. 5

2 EXTERNAL EHR, CDW AND CTMS DATA SOURCES .......................................................... 6
  2.1 Electronic Health Records .......................................................................................... 6
      2.1.1 CUSTOM EHRS .................................................................................................. 6
      2.1.1.1 HL7 messages generation from data source ................................................... 6
      2.1.1.2 Tabbed / CSV exportation ............................................................................. 8
      2.1.2 EHR – HL7 CLINICAL RESEARCH FILTERED QUERY ..................................... 9
      2.1.3 IHE INTEGRATION PROFILES ............................................................................ 10
      2.1.3.1 Retrieve Form for Data Capture .................................................................. 10
      2.1.3.2 Cross Enterprise Document Sharing (XDS) .................................................. 12
      2.1.3.3 Clinical Research Document (CRD) ............................................................ 13
      2.1.3.4 Cross-Community Access (XCA) .................................................................. 14
      2.1.3.5 Cross-Enterprise Document Reliable Interchange (XDR) ............................. 15
      2.1.3.6 Clinical Research Process Content (CRPC) .................................................. 15
      2.1.3.7 IHE Based Document Sharing Solution (Sample) .......................................... 16
      2.1.3.8 Conclusions .................................................................................................. 17
  2.2 Clinical Data Warehouses .......................................................................................... 18
      2.2.1 TRANSMART .................................................................................................... 18
      2.2.1.1 Data warehouse ........................................................................................... 18
      2.2.1.2 Data access .................................................................................................. 20
      2.2.1.3 Core API ...................................................................................................... 24
      2.2.1.4 Conclusion ................................................................................................... 25
  2.3 Clinical Trial Management Systems ............................................................................ 26
      2.3.1 OPENCLINICA .................................................................................................. 26
      2.3.1.1 Requirements of the EURECA scenarios related to OpenClinica: .................. 26
      2.3.1.2 The CDISC ODM standard .......................................................................... 26
      2.3.1.3 Import data into OpenClinica ...................................................................... 26
      2.3.1.4 Concept of interfacing .................................................................................. 28
2.3.1.5 Cross validation performed by OpenClinica for imported data........29
2.3.1.6 Export data from OpenClinica to external systems..................29
2.3.2 OBTIMA .............................................................................30
2.3.2.1 The clinical trial management system ObTiMA.........................30
2.3.2.2 Providing ObTiMA data for EURECA Data Push Service..........31
2.4 PHR .......................................................................................32
2.4.1 INDIVO-X ............................................................................32
2.4.1.1 Data export from IndivoX.........................................................32
3 CONCLUSIONS ...........................................................................36
3.1 EURECA Interface layer ..........................................................37
3.1.1 SEMANTIC INTEGRATION LAYER ........................................37
4 REFERENCES ................................................................................39
5 APPENDIX A CLINICAL RESEARCH DOCUMENT PRE-POPULATED
DATA CONTENT .............................................................................40
1 Introduction

This document reports on the analysis performed for EURECA task 2.3. In EURECA, task 2.3 includes the definition of a standards based interface layer towards external systems. The goal of the interface layer is to hide the heterogeneity of external EHR, CDW or CDMS data sources from the EURECA services. The interface layer will be defined as a set of abstracted interfaces (e.g. providing functional access through standard EURECA APIs, providing direct data access, etc.). This document focuses on assessing the external data sources of the EURECA clinical partners (which are described in the use cases), as these are very relevant for the clinical pilots. The task will include (but is not limited to) an analysis of the abstraction layer requirements and design with respect to:

- Providing data access towards the external data sources, taking into account the information and knowledge models used by the data sources and the EURECA services.
- Assessing the possibilities of exposing functionalities of the source systems, e.g. existing query functionality in an EHR or other relevant functionality.
- Analyzing the need for “intermediate” (shadow) extracts of source data as part of the interface layer.

The following types of external data sources are assessed:

- Electronic Health Record (EHR)
- Clinical data warehouse (CDW)
- Clinical Trial Management System (CTMS)
- Personal Health Records (PHR)

For EHRs, we describe the ETL process used to incorporate EHR systems in the EURECA environment that do not have a standardized query interface. All EHR systems of the EURECA clinical partners (see deliverable 1.3) fall under this category. Additionally, the HL7 Clinical Research Filtered Query is described and an investigation of relevant IHE profiles is performed. For CDWs we describe tranSMART\(^1\), a clinical data warehouse that is currently gaining a lot of traction in the clinical research world. For CTMSes, openClinica and Obtima are described. And finally, for PHRs we describe Indivo X\(^2\).

\(^1\) [http://www.transmartproject.org/]
\(^2\) [http://indivohealth.org/]
2 External EHR, CDW and CTMS data sources

2.1 Electronic Health Records

2.1.1 Custom EHRs

Electronic Health Records (EHRs) are the current basis for collecting electronic health information about individual patients in digital format. Despite the fact that the information is represented in a format that should facilitate its sharing across different health care infrastructures, a set of tools is required to perform this sharing correctly. The development effort required varies depending on the infrastructure where the clinical information is stored. This implies an individual effort for/with each data provider to allow sharing of information.

In order to allow data exportation from each clinical partner to the infrastructure developed by the EURECA project, it is necessary to provide a set of tools. These tools will set aid export of the data stored by each partner. Due to the diversity of technologies, methods and infrastructures used by each clinical entity for managing their data, the idea of having an implementation that provides a fully automatic and customized tool for each partner has been discarded. Following this approach, the set of tools won't implement a customized complete solution, but it will ease the required effort of the Information Technology department from each clinical partner.

The set of tools that have been developed consist of a set of coding rules, templates, mappings and guidelines that will apply for each data source. There are three existing use cases for these tools:

- The EHR is capable of exporting data to HL7 v3 or HL7 v2.x messages: coding rules will be applied and the data will be sent directly to the data push service.
- Data is stored in databases with legacy data model: guidelines and templates will be provided to transform data from databases to HL7 v3 messages using open source tools, such as Mirth Connect.
- Data is stored in tabular plain-text files or EHR database is not accessible so data has to be exported to a tabular / CSV format: a set of mappings to a HL7-based CSV will be used to adapt the original data.

2.1.1.1 HL7 messages generation from data source

Clinical databases usually interact by means of HL7 messages. Although the HL7 standard is widely used, each IT provider and system has its own internal data model and makes use of different technologies. In order to be able to export data from a database, we need a tool to access the data, and the rules needed to adapt the information to common scheme of HL7 v3 messages. In this context, we propose the use of Mirth Connect which is capable of accessing different kind of database management systems (such as MySQL, PostgreSQL, Oracle, Microsoft SQL Server and DBC). It also provides functionality for data transformation.

The process to transform data from a database to a HL7 v3 message using the Mirth Connect tool follows multiple steps (Figure 1). Firstly the connection to the database is
established which makes it possible to query for data and access it during the transformation process. For this initial step, a general tutorial describing how to perform this stage of the process will be provided. It is deemed sufficient and feasible to provide the data provider with a generic tutorial only, as the data provider will be very familiar with its own database.

For the transformation process, which has a higher complexity, more extensive documentation is required. This documentation includes a tutorial focusing on the management of the Mirth Connect tool and a set of templates to adapt the data format into a HL7 v3 message (Figure 2). The aim of this stage is to link elements extracted from the database to fields of a HL7 v3 message. For this reason, guidelines that explain what kind of information must be stored at a field of a HL7 v3 message, and templates that define the structure that the message should have, are mandatory.
This process will lead to the formation of a HL7 v3 message that wraps the information from the selected database and conforms to the requirements of the data push service. Therefore, the message extracted can be validated at the Data Push Service and then can be finally integrated at the Common data model.

### 2.1.1.2 Tabbed / CSV exportation

One of the formats supported when introducing data into the CDM is the Comma-Separated Values (CSV) format. This format will be mainly used when there is no possibility of accessing the original data in the database directly (in which case we should use the Mirth Connect tool) or the data cannot be exported (directly) through an HL7 message.

In the CSV file each of the attributes in the HL7 RIM model is stored in a field of a comma-delimited CSV. HL7 RIM contains a large quantity of attributes which results in CSV files containing many fields. In many situations only a small number of these fields are required to introduce the relevant data. Therefore, a list of the allowed fields in the CSV will be provided and each clinical entity will be responsible for selecting the required fields. To make sure that the format of the received CSV is the same in all data providers, a script will be provided that will reformat the CSV, including the empty fields that are not required by each clinical entity and reordering those fields if necessary.

Due to the structure of the HL7 RIM where acts and entities are related in a many-to-many cardinality, a division of the information in two CSV files was performed. One of the CSVs contains all the information related to the acts from the clinical trials and the other one contains the information of the entities related to those acts. These entities include the patients, devices, etc. The two CSV files will be related through the field EntityID. A small example of part of such CSV files is included in Table 1 and Table 2.

**Act CSV**
In this example we can observe that the relationship between the two CSV examples is established between the Patient_ID in the Act CSV and the Entity_ID in the Entity CSV. Another important point to notice in these CSV examples is that there may be multivalued fields such as targetSiteCode or methodCode (multiple target sites or methods related to the act itself). To represent these multivalued fields a special separator is used.

To allow the data providers to understand where to place each piece of data into the CSV, some guidelines will be provided explaining each field and the process that must be followed when performing the mapping from the original data to the CSV fields.

### 2.1.2 EHR – HL7 Clinical Research Filtered Query

The Clinical Research Filtered Query Service Functional Model (CFQ SFM) defines a single service with two distinct profiles:

- The first is an EHR service provider that queries all EHR records to provide data on (counts of) patients who satisfy the inclusion or exclusion criteria
- The second service is a protocol repository provider that receives a list of protocols that the clinical data (and thus the patient) satisfies

The CRFQ SFM is specifically restricted to applications in the clinical trials context (due to the explicitly defined business need) and thus should not be seen as a generalized query service. However, its clinical trials context is shared with several EURECA use cases and therefore it might be interesting. (It should be noted that CRFQ SFM can be seen as foundational model for other (query) services).

The first profile specifies a service on an EHR repository to execute (eligibility) criteria against the hosted patient data and returns a list of qualified patients. The service (List Qualified Patients) takes a set of the metadata from a single protocol and returns an array of patient IDs (which may be pseudo-IDs depending on security considerations). Of course, the usefulness of the proposed service depends on the expressiveness of the formalism used to express the eligibility criteria. Unfortunately, the specification leaves the formalism unspecified. It only refers to the ASPIRE project and its dataset as an example. The overall objective of ASPIRE is to create a structured representation of a core set of encoded protocol eligibility criteria, using accepted medical terminology and vocabulary standards.

---

when available. Unfortunately, the ASPIRE dataset has a too coarse granularity to be of use.

The second service (List Qualified Protocols) takes a set of patient data from a single patient and returns an array of protocols (IDs, names, etc). Also this service specification lacks a sufficient elaborated (technical) specification. The specification does not cover how the required patient data can be obtained and in what formats it can be submitted.

Although the Clinical Research Filtered Query Service Functional Model specification cannot be directly of use in the EURECA platform and project, it is a standard for which development should be tracked, as the concepts specified in the functional model are highly relevant.

### 2.1.3 IHE Integration Profiles

Integrating the Health Enterprise (IHE) (see also deliverable 2.1) developed a set of standards, called integration profiles, which are detailed specifications for communication among systems to address key clinical use cases, all based on established standards. IHE Profiles address critical interoperability issues related to information access for care providers and patients, clinical workflow, security, administration and information infrastructure. Each profile defines the actors, transactions and information content required to address clinical use cases by referencing appropriate standards. Some of the IHE integration profiles might be interesting in the context of interfacing external (EHR) systems.

<table>
<thead>
<tr>
<th>IHE Profile</th>
<th>Brief description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audit Trail &amp; Node Identification (ATNA)</td>
<td>Basic security through (a) functional access controls (b) defined security audit logging and (c) secure network communications.</td>
</tr>
<tr>
<td>Consistent Time (CT)</td>
<td>Enables system clocks and time stamps of computers in a network to be synchronized (median error less than 1 second).</td>
</tr>
<tr>
<td>Patient Demographics Query (PDQ)</td>
<td>Lets applications query by patient demographics for patient identity from a central patient information server.</td>
</tr>
<tr>
<td>Patient Identifier Cross Referencing (PIX)</td>
<td>Lets applications query for patient identity cross-references between hospitals, sites, health information exchange networks, etc.</td>
</tr>
<tr>
<td>Patient Demographics Query HL7 v3 (PDQv3)</td>
<td>Extends the Patient Demographics Query profile leveraging HL7 version 3.</td>
</tr>
<tr>
<td>Patient Identifier Cross Referencing</td>
<td>Extends the Patient Identifier Cross-Reference profile leveraging HL7 version 3.</td>
</tr>
</tbody>
</table>
**HL7 v3 (PIXv3)**

| Retrieve Form for Data Capture (RFD) | Enables EHR applications to directly request forms from clinical trial sponsors and public health reporting. |
| Clinical Research Document (CRD) | Clinical Research Document describes the content pertinent to the clinical research use case required within the Retrieve Form for Data-Capture (RFD) pre-population parameter. |
| Cross-Community Access (XCA) | Allows to query and retrieve patient electronic health records held by other communities. |
| Cross-Enterprise Document Reliable Interchange (XDR) | Exchanges health documents between health enterprises using a web-service based point-to-point push network communication. |
| Clinical Research Process Content | This profile is based on Retrieve Process for Execution profile, which provides general framework for messaging interaction. It further details the content which is specific to research domain. |

### Table 3 IHE IT Integration Profiles

Table 3 lists a number of IHE integration profiles, which can be relevant in EURECA context. Some of the profiles (ATNA, CT) are being relied on by others; therefore have to be implemented as well.

Open Health Tools project\(^5\) provides client-side implementation (in Java) of several key IHE profiles:

- ATNA: Audit Trail and Node Authentication
- MPQ: Multi Patient Query
- PAM: Patient Demographics Source
- PIX: Patient Identifier Cross-Referencing
- PDQ: Patient Demographics Query
- SVS: Shared Value Sets
- XCA: Cross Community Access
- XDR: Cross-Enterprise Document Reliable Interchange
- XDS: Cross-Enterprise Document Sharing
- XUA: Cross-Enterprise User Assertion

### 2.1.3.1 Retrieve Form for Data Capture

The Retrieve Form for Data Capture Profile (RFD, Figure 3) provides a method for gathering data within a user’s current application to meet the requirements of an external system. RFD supports the retrieval of forms from a form source, display and completion of a form, and return of instance data from the display application to the source application [1].

---

\(^5\) [https://www.projects.openhealthtools.org/sf/sfmain/do/home](https://www.projects.openhealthtools.org/sf/sfmain/do/home)
This profile describes manual filling out of the form, while allowing some pre-population of data from the local EHR system. The form itself is XHTML based, which supports submit transaction. Optionally Form Manager can support XForms [2]. Upon user’s request from form filler application (local EHR for example) the form is retrieved from the external form manager and displayed within form filler application. Some of the data might be prepopulated, format of prepopulated data is defined by domain-specific IHE content profiles (see [3]), is application specific and out of the scope of RFD integration profile. Semantics of the form itself is out of the scope of the profile as well. As of now IHE specifies following domains:
- Cardiology
- Dental
- Endoscopy
- Eye Care
- Laboratory
- Anatomic Pathology
- Patient Care Coordination
- Pharmacy
- Quality, Research and Public Health
- Radiation Oncology
- Radiology

This profile covers one patient at a time only (as the rest of the IHE profiles actually), so to get data for multiple patients this profile has to be combined with patient demographics query (PDQ) and patient cross-referencing (PIX) profiles.

2.1.3.2 Cross Enterprise Document Sharing (XDS)
Cross-Enterprise Document Sharing (Figure 4) enables a number of healthcare delivery organizations belonging to an XDS Affinity Domain (e.g., a community of care) to cooperate in the care of a patient by sharing clinical records in the form of documents as they proceed with their patients’ care delivery activities. Federated document repositories and a document registry create a longitudinal record of information about a patient within a given XDS Affinity Domain. XDS as well as the rest of IHE integration profiles is document format neutral, any type of clinical information is
supported. This profile is based on ebXML Registry standards⁶. Document consumer can query the document registry, however the query is defined in the registry. The implementation of queries is up to the document registry, no specific querying mechanisms are defined by IHE, and consumer has to transmit only the query id and query parameters to invoke it. This makes querying on registry flexible and easily customizable, as well as protecting the registry from malicious queries.

Figure 4 Cross Enterprise Sharing Actors and Transactions

This profile heavily relies on ATNA and CT profiles, as well as on existing IT infrastructure within healthcare enterprise (document repository, document registry). Therefore implementation effort is high.

2.1.3.3 Clinical Research Document (CRD)
The Clinical Research Document Profile (CRD, Figure 5) specifies a standard way to generate a so-called clinical research document from EHR data provided in the CDA [4] standard. In summary, the CRD profile is just like the RFD profile except it is more specific about the pre-population XML (see Appendix A) requirements used when retrieving a form, some optional functionality is more tightly specified as required, a new transaction is created and is used to facilitate the archiving of the pre-population data, and other actors groupings are added to enhance the security of CRD actors.

⁶ http://www.ebxml.org/specs/
As well as in RFD case to get data for multiple patients this profile has to be combined with PDQ & PIX. Unlike RFD, the process can be fully automated, which is more desirable in EURECA context.

2.1.3.4 Cross-Community Access (XCA)

The Cross-Community Access profile supports the means to query and retrieve (see XDS, Figure 6) patient relevant medical data held by other communities. A community is defined as a coupling of facilities/enterprises that have agreed to work together using a common set of policies for the purpose of sharing clinical information via an established mechanism. Facilities/enterprises may host any type of healthcare application such as EHR, PHR, etc. A community is identifiable by a globally unique id called the homeCommunityId. Membership of a facility/enterprise in one community does not preclude it from being a member in another community. Such communities may be XDS Affinity Domains, which define document sharing using the XDS profile or any other communities, no matter what their internal sharing structure.
2.1.3.5 Cross-Enterprise Document Reliable Interchange (XDR)

Cross-Enterprise Document Reliable Interchange (XDR, Figure 7) provides document interchange using a reliable messaging system. This permits direct document interchange between EHRs, PHRs, and other healthcare IT systems in the absence of a document-sharing infrastructure such as XDS Registry and Repositories.

XDR requires less infrastructure comparing to XDS, however combination with PDQ/PIX is required to make sharing solution more usable.

2.1.3.6 Clinical Research Process Content (CRPC)

CRPC profile (see Figure 8) aims on resolving complexity of research protocols, which guide the conduct of trials. Existing Retrieve Process for Execution IHE profile is used as a general framework for interaction between the actors. Further CRPC details the content, which is specific to the research domain. The profile is focusing on the following use cases:
- Study definition sharing;
- Recruitment at the point of care;
- Enrolment through the EMR;
- Generation of patient-specific timeline of study events;
- Adjustment of patient-specific timeline of study events.

The main focus of all the above mentioned use-cases is avoiding manual double administration of the clinical trial in a clinical trial management system and other systems involved (such as EMR, etc.)

2.1.3.7 IHE Based Document Sharing Solution (Sample)
In [5] IHE presented a sample document sharing solution based on IHE profiles (Figure 9).
IHE uses a so-called "meet-in-the-middle" approach for service identification. This approach in an essence is about balancing between the needs (services identified by the top-down approach, driven by the business) and functionality already provided by existing IT assets (bottom-up approach). Services depicted on the left side of Figure 9 are IHE profile “wrapper” services – implementing actors & transactions exactly as specified in IHE integration profiles. Services on the left side are implementing consumer behaviour described in correspondent IHE integration profiles. In addition to that they provide higher abstraction-level interfaces for the next layer of service consumers, making document sharing easier for those. The shaded area is actually the scope that is determined by IHE Integration profiles.

2.1.3.8 Conclusions

An obvious benefit of IHE-compliant data exchange with external systems in EURECA would be ability to unify communication with all the data sources. There are however few serious disadvantages of building such a solution in EURECA context:

- IHE profiles are document format agnostic, therefore data push service would still have to know semantic/syntactic specifics of data coming in from each data source;
IHE “wrapper” services would have to be built per participating healthcare institution. Due to the way how the data is obtained & transformed in EURECA, these wrapper services will not be used too frequently, while implementation effort can be rather high. However, in case if applications would work directly on the data from external data sources IHE compliant document sharing would be a good solution, unifying document exchange and hiding specifics of each data source (including querying capabilities) in IHE wrapper services. In deliverable D3.3 we will describe which of the IHE profiles discussed here can be matched with the EURECA use cases.

2.2 Clinical Data Warehouses

2.2.1 TranSMART

TranSMART is a knowledge management platform that enables scientists to develop and refine research hypotheses by investigating correlations between genetic and phenotypic data, and assessing their analytical results in the context of published literature and other work. Two major features of tranSMART are:

1. Data Explorer. Data Explorer is an i2b2-based tool that represents the concepts in tree form and lets you compare two sets of study groups based on one or more points of comparison. One can define both the criteria that populate the study groups and the points of comparison between the study groups.

2. Search Tool. TranSMART provides a Google-like search tool that lets you search across multiple data sources for information related to items of interest, such as biomarkers, diseases, genes, and gene signatures.

Both features use different data sources which reside in the same data warehouse.

The objective is to Extract, Transform and Load (ETL) the data, which resides in the CDM, into tranSMART. Then, tranSMART users can use the Data Explorer to query data that was extracted from the CDM, in aggregate form.

2.2.1.1 Data warehouse

The data warehouse (PostgreSQL based) is divided into distinct zones, spread out over three servers (Figure 10). The purpose of these separate areas is to ensure that raw data can be continually loaded into the data warehouse and transformed into correct, complete, and schema-compliant data, without affecting users of the production system [6].

---

7 http://www.transmartproject.org/
8 https://www.i2b2.org/
Environment
The three servers where the zones reside are called environments. There are three environments:

1. Development environment. This environment is used by application developers, for data loading and transformation. The landing and working zone contains (semi) raw data and temporary tables. These zones are used heavily by the stored procedures (a part of the ETL process) and are emptied every time new data is loaded.
2. Staging/QA environment. Data testing takes place in this environment and is used by QA staff and project leaders.
3. Production environment. This environment is used by the end-users and contains production data.

Core and Data Marts
Every environment contains a copy of the data warehouse, which consists of:

1. Core data warehouse, which contains biomarker data.
2. Data Mart 1. This is the Search Tool database and contains stored search filters, search key words for auto-completion functionalities and user defined search queries. Queries are forwarded to the core data warehouse, i.e. there is a mapping between Data Mart 1 and the core data warehouse.
3. Data Mart 2. This data warehouse consists:
   a. An i2b2 database, which contains ontology and fact data (clinical data).
   b. Gene expression database, which is an extension of part a (reference to clinical data).
Clinical and gene expression data

TranSMART uses Data Mart 2, i.e. the i2b2 data model for clinical data and an extended model with gene expression data. Data Mart 2 consists of the following schemas:

1. I2B2METADATA. This schema contains the ontology data.
2. I2B2DEMODATA. This is the i2b2 star schema\(^9\) that contains ontology and (clinical) fact data.
3. DEAPP schema. This schema is an extension of the i2b2 schemas and contains gene expression data.

Figure 11 shows a partial graphical user interface of the Data Explorer. The figure shows a comparison between two groups. Each group is called a subset. Both groups share the (first) concept ‘Hematology’. The second concept is the provider (e.g. Subset 1 has provider ‘Hansen, Sydney, MD’).

The information in region 1 is retrieved from the I2B2METADATA.I2B2 table. The table contains detailed information about the i2b2 navigation tree in the Data Explorer feature. Region 2 shows the number of subjects that were involved in an observation with the ‘Hematology’ concept. This information is retrieved from the I2B2DEMODATA.OBSERVATION_FACT table. Region 3 shows statistics regarding the age (e.g. mean and average) of a patient. This information is retrieved from the I2B2DEMODATA.PATIENT_DIMENSION table.

2.2.1.2 Data access

TranSMART passes requests to i2b2 web services to access data, but some requests are directly passed to the database into the i2b2 tables instead of the interfaces provided by the web services. The i2b2 web services run on a JBOSS\(^{10}\) server.

---

\(^9\) http://en.wikipedia.org/wiki/Star_schema
\(^{10}\) http://www.jboss.org/
Web services
The i2b2 web services are divided into many cells. A few important cells used by tranSMART are the Ontology Management and Data Repository (CRC) cells.

Ontology Management
The Ontology Management cell manages the i2b2 ontology. The ontology follows a tree structure. External vocabularies / terminologies can be imported into this tree structure. The tree structure implies a subsumption relationship. There is no support for other/additional types of relationships, which has its implications when importing an ontology which supports different/multiple relationship types.

Data Repository (CRC)
The data repository cell (also called the Clinical Research Chart (CRC) repository cell) is one of the core cells. Its main requirements are:

- It must be able to hold healthcare information from many different venues and allow it to be queried rapidly even if there are hundreds of millions of rows.
- It must be easily combined with other project repositories to form large unified repositories.
- It must allow objects to be stored that are present in the genomic data.

The CRC is designed as a data warehouse, focusing on clinical data (a patient’s phenotype and genotype information). Data in the CRC is de-identified, except for encrypted patient notes. The CRC relies on other cells for authentication and metadata management.

The CRC provides two services, a setFinder web service and a PDO (Patient Data Object) web service. setFinder manages a user's setFinder queries. The queries are used to create a set of patients that satisfy the specified criteria. The API closely mimics the way the graphical user interface is designed and setFinder queries are composed of query constraints, a list of panels and its items. The criteria put constraints on concepts from the i2b2 ontology in order to select instances from the data mart. The PDO web service exposes the clinical data, providing access to patient information such as clinical observations, demographics and provider data.

We refer the reader to [7] and [8] for more information about the Ontology Management and Data Repository cells.

Shortcomings
In the near future, the dependency between tranSMART and i2b2 will be removed (therefore also removing the JBOSS dependency). According to the developers of tranSMART the model has already been extended to the point where standard i2b2 tools cannot capture all the information in the database [9]. A new API11 has been proposed of the Ontology Management and Data Repository cells.

Direct database access

11 https://github.com/thehyve/transmart-core-api
TranSMART passes some requests directly to the databases. These requests are related to the JavaScript front-end used in tranSMART [9]. A simple way to access the data in the database is via the tool pgAdmin\(^{12}\).

**I2b2 data model**
Figure 12 shows a subset of the i2b2 star schema. The arrows indicate the primary and foreign keys.

To lookup the average age of Subset 1 (mean in Figure 11), we use the following query:

\(^{12}\) [http://www.pgadmin.org/](http://www.pgadmin.org/)

SELECT avg(age_in_years_num) as AVG_AGE_SUBSET1  
FROM i2b2demodata.patient_dimension  
WHERE patient_num in(  
  SELECT patient_num  
  FROM i2b2demodata.observation_fact O  
  WHERE O.concept_cd in(  
    SELECT concept_cd  
    FROM i2b2demodata.concept_dimension  
    WHERE concept_path  
    LIKE '%%\Labtests\LAB\(LLB53) Hematology\(%'  
  )  
  AND  
  O.provider_id in(  
    SELECT provider_id  
    FROM i2b2demodata.provider_dimension  
    WHERE provider_path  
    LIKE '%%Providers\Medicine\Primary Care\Hansen, Sydney, MD\%'  
  )  
)  
The following query "give all the patient ids and ages for patients having a Hematology lab" illustrates AGE as a concept instead of an attribute of the patient dimension table.

SELECT patient_num, concept_cd  
FROM i2b2demodata.observation_fact O  
WHERE concept_cd like '%DEM|AGE%' AND  
  O.patient_num in(  
    SELECT patient_num  
    FROM i2b2demodata.observation_fact  
    WHERE concept_cd in(  
      SELECT concept_cd  
      FROM i2b2demodata.concept_dimension  
      WHERE concept_path like '%%\Labtests\LAB\(LLB53) Hematology\%'  
    )  
    AND  
    O.provider_id in(  
      SELECT provider_id  
      FROM i2b2demodata.provider_dimension  
      WHERE provider_path like '%%Providers\Medicine\Primary Care\Hansen,Sydney, MD%'  
    )  
  )  
Table 4 shows a partial result of the query.

<table>
<thead>
<tr>
<th>patient_num</th>
<th>concept_cd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000000101</td>
<td>DEM</td>
</tr>
<tr>
<td>1000000105</td>
<td>DEM</td>
</tr>
</tbody>
</table>

Table 4 Partial result of a query with age as a concept

DEAPP Extension
Figure 13 shows two tables of the DEAPP model: DE_SUBJECT_SAMPLE_MAPPING and DE_SUBJECT_MICROARRAY_DATA. The DEAPP model extends the i2b2 model by including the PATIENT_NUM and CONCEPT_CD.
2.2.1.3 Core API

The developers of *thehyve* created different API's which follow the design philosophy behind i2b2 and resemble the i2b2 Ontology Management and Data Repository cells.

**Clinical Data Ontology API**

The Clinical Data Ontology API\(^{15}\) is based on the Ontology Management cell. A ConceptsResource, which represents i2b2 concepts, is defined which can be implemented by clinical data sources to expose ontology trees with concepts. Table 5 shows the interface of the ConceptsResource.

<table>
<thead>
<tr>
<th>method</th>
<th>Input (type: name)</th>
<th>return value</th>
</tr>
</thead>
<tbody>
<tr>
<td>getAllCategories()</td>
<td>-</td>
<td>List&lt;OntologyTerm&gt;</td>
</tr>
<tr>
<td>getByKey()</td>
<td>String: conceptKey</td>
<td>OntologyTerm</td>
</tr>
</tbody>
</table>

Table 5 The ConceptsResource interface

---


The OntologyTerm is an i2b2 ontology metadata entry. Table 6 shows a method of the interface.

<table>
<thead>
<tr>
<th>method</th>
<th>return value</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>getFullName()</td>
<td>String</td>
<td><code>&lt;code&gt;i2b2\Diagnoses\Musculoskeletal and connective tissue (710)\Arthropathies (710-19)(714) Rheumatoid arthritis and other arthropathies (714-0) Rheumatoid arthritis&lt;/code&gt;</code></td>
</tr>
</tbody>
</table>

**Clinical Data Query API**

The Data Repository cell is quite more complex than the Ontology Management cell. It allows creating queries involving any number of ontology terms (and even other objects, such as other queries) and restrict data by value, possibly even doing unit conversions [9]. A QueriesResource\(^{16}\) is defined which can be exposed by clinical data sources as a means of querying clinical data. Table 7 shows the interface of the QueriesResource.

<table>
<thead>
<tr>
<th>method</th>
<th>Input (type: name)</th>
<th>return value</th>
</tr>
</thead>
<tbody>
<tr>
<td>runQuery</td>
<td>QueryDefinition: definition</td>
<td>QueryResult</td>
</tr>
<tr>
<td>getQueryResultFromId</td>
<td>Long: id</td>
<td>QueryResult</td>
</tr>
<tr>
<td>getQueryDefinitionForResult</td>
<td>QueryResult: result</td>
<td>QueryDefinition</td>
</tr>
</tbody>
</table>

**High Dimensional Query API**

This API\(^{17}\) supports querying of high dimension data such as assays (samples, biological tissues).

**Core-DB**

The Core-DB project\(^{18}\) is a Grails plugin that implements the clinical and high dimension API’s described in the previous sections. This project is the actual re-implementation of all the i2b2 functionalities used tranSMART.

**2.2.1.4 Conclusion**

The i2b2 web services can be (re)used, as an implementation of an interface layer for EURECA services, to access the data in tranSMART. However, the i2b2 web services do not capture all the information that resides in tranSMART and the interface layer requires i2b2 running on a JBOSS server (introducing an undesired dependency).


\(^{18}\) [https://github.com/thehyve/transmart-core-db](https://github.com/thehyve/transmart-core-db)
2.3 Clinical Trial Management Systems

2.3.1 openClinica
As described in Deliverable 2.1 [19], OpenClinica is a software platform for Electronic Data Capture and for managing clinical research studies. It helps both creation and management of CRFs, by collecting the large amount of data needed for running a clinical trial in electronic format, and also provides validation components to check user data and reporting tools for analysis of the data that have been collected [11].

2.3.1.1 Requirements of the EURECA scenarios related to OpenClinica:
OpenClinica will be used within the development of the following EURECA scenarios and tools:
- The scenarios “Pre-filling of CRF and AE reports”, “Automatic detection and reporting of SAEs/SUSARs” and “Long-term follow-up” need to import data files into OpenClinica from clinical data that should have been extracted automatically from an EHR.
- The scenario “Economic analysis” needs to export, and then analyse, data files from OpenClinica.

As all of these scenarios will have to exchange information by importing – using the “Event CRF” data importing feature in OpenClinica (e.g. to automatically fill eCRFs) – or exporting CRF data from OpenClinica towards developed tools (e.g. to analyse economic aspects of different procedures by joining data from clinical trials), an interface layer between the EURECA services and OpenClinica is needed.

2.3.1.2 The CDISC ODM standard
OpenClinica is built on the CDISC[20] Operational Data Model[21] (ODM) standard as it imposes a data type on each value – it is a strongly-typed system – this simplifies the sharing of clinical and metadata from multiple sources. It is defined as a Document Type Definition (DTD) format, and enables to acquire data from a paper or electronic CRF coming from a Clinical Research Organisation, and the storage of clinical trial data.

CDISC ODM defines different types of information associated with a clinical trial:
- Metadata about a clinical trial;
- Administrative data from a clinical trial;
- Clinical data.

There are two ways of loading these data in OpenClinica, via the Web-based user interface or via Web Services by SOAP requests.

2.3.1.3 Import data into OpenClinica
As previously stated, the data to be imported should be defined according to CDISC ODM (Operational Data Model), structured in a properly-formatted XML file format.

[20] Clinical Data Interchange Standards Consortium
[21] http://www.cdisc.org/stuff/contentmgr/files/0/919cb4ef843829170d470b37eb662aeb/misc/odm1_1_0.html
This XML file is generated using Object Identifiers (OIDs) which are used by OpenClinica to uniquely reference and identify the entities in the study.

OIDs are necessary for importing data. They can be determined by looking into the XML Study metadata file. For Study (Figure 14), Study Event Definitions (Figure 15), CRFs (Figure 16), Item Groups (Figure 17) and Items, the OIDs are created as part of Study Setup. For Study Subjects the OIDs are created when the Subject has been added to the Study.

**Figure 14 Determination of a Study OID**

```
<?xml version ...?>
<OM ... >

<Study OID="YourStudyOID"> <!-- Get the Study OID here. -->
```

**Figure 15 Determination of a Study Event OID**

```
<Protocol>

<StudyEventRef StudyEventOID="YourFirstStudyEventOID" .../> <!-- Get the OID for your first Study Event here -->

<StudyEventRef StudyEventOID="YourSecondStudyEventOID" .../> <!-- Get the OID for your second Study Event here -->
...
</Protocol>
```

**Figure 16 Determination of a Form (CRF) OID**

```
<FormDef OID="YourFirstFormOID" ...> <!-- This section contains the OIDs for the Item Groups in the CRF whose OID is YourFirstFormOID -->

<ItemGroupRef ItemGroupOID="YourFirstItemGroupOID" .../> <!-- Get the OID for the first Item Group in this CRF here -->
<ItemGroupRef ItemGroupOID="YourSecondItemGroupOID" .../> <!-- Get the OID for the second Item Group in this CRF here -->
...
</OpenClinica:FormDetails ... >
...
</OpenClinica:FormDetails>
</FormDef>
```
In order to import data from another system, the Study, Event Definitions, Event CRFs and Study Subjects for the data being imported must exist and must be scheduled for the Study Events related to this data.

Scheduled import jobs can be used to automate data import – the most fine-grained is a daily run – allowing the import of properly-formatted XML data file on a regular basis, a complete study although cannot be fully imported in this case. No solution exists at this stage to automatically import concepts like management of users, study and site metadata together with other applications.

Several data import tools [12] can be used to map the data file for import in another format than ODM XML. Among those, Mirth[22] is a cross-platform HL7 interface engine. It is installed on the OpenClinica server and enables HL7 messages to be sent between different systems and applications over a variety of protocols: databases, local file system, network shares, etc. Mirth can map tab-delimited based text files to the CDISC ODM output and allows automating the process of scheduling the data to be imported into OpenClinica.

A channel map needs to be created in the system as it is responsible for the creation of the CDISC XML file with the path OID and data from the subject identifier file, the database, and the tab-delimited text file, by mapping the file column names to the CRF item names. It is done querying the subject id for the subject OID, and adding the study related OID path from the study identifier file. Then the study identifiers file indicates the path of the CRF attributes to import to the Mirth Connect channel map (In OpenClinica it is possible to download all OIDs for each study). Finally it is possible to configure the global scripts in Mirth Connect to define both input and output directory, the study identifier to be used, and the database to be connected.

2.3.1.4 Concept of interfacing

Meta-data on clinical trials can be automatically transferred from the Study Management Tool (SMT) into OpenClinica [13] using OpenClinica Web services (OC-WS) and Simple Object Access Protocol (SOAP) support. The OpenClinica core system is used for interaction, data entry and user interfacing, the OpenClinica Web Services are allow automatic data transfer into the system and a database management system allows to store the data in relational tables.

2.3.1.5 Cross validation performed by OpenClinica for imported data

OpenClinica validates the XML files that have been imported by checking the file format and validity. The Alerts and Messages sidebar panel reports the problems when errors occur at some point in the validation process (e.g. invalid date format). Then it corrects any errors in the file and imports it over again.

Here are the validations and tests performed by OpenClinica on the data import file [11]:

- XML is well-formed against the ODM.
- The Study OID in the import file matches the current Study in the OpenClinica session you are performing the import from.
- Study Subject OIDs exist for the current Study.
- Study Event Definition OIDs exist for the current Study.
- OIDs for the CRF versions exist for the Study Event Definition OID.
- Study Subject OIDs are scheduled for the Study Events.
- Data values for the Items comply with the metadata. For example, if the metadata for an Item calls for a value of 1 or 2, the data value for that Item in your import file must be 1 or 2.
- Data values are in the correct format. For example, when an integer is required, a decimal value generates an error.

After this, OpenClinica runs an audit log of all actions performed on a specific record, including a timestamp of the modifications together with the user and the reason for the modification, with both old and new values of the record. Validation scripts can be built to routinely ensure reliability of data imported [14].

Rules can be defined as well in OpenClinica in order to validate data entry for CRFs during data entry process (e.g. initial data entry, double data entry, administrative data entry). Rules are generated as elements of an XML file, and are then uploaded into OpenClinica.

2.3.1.6 Export data from OpenClinica to external systems

Practically everything can be exported in CDISC ODM from OpenClinica (e.g. study metadata, structure and study data, discrepancy rules). With ODM standard, it is possible to import the study in another standard compliant application. Scheduled import jobs are used to extract data automatically for an existing dataset definition. In practice, only some vendor extensions cover more application-specific points, which prevent total compatibility.

As pointed out hereinbefore, the EURECA scenario on “economic analysis of difference clinical procedures” has specific needs for exploiting – and thus exporting – data files from OpenClinica.

The Data Mart features (only available in the OpenClinica Enterprise Edition) allow extracting data from OpenClinica for use with standard relational database for
reporting and analysis. It also allows creating reports for data using SQL reporting tools.

2.3.2 Obtima

2.3.2.1 The clinical trial management system ObTiMA

ObTiMA is an ontology-based clinical trial management system for designing and conducting clinical trials. The core functionality of ObTiMA has been developed within the integrated European project ACGT23 (Advancing Clinico Genomic Trials on Cancer), which provided a European biomedical grid infrastructure for cancer research. Since then, the application was continuously improved and extended. The department for Pediatric Oncology at Hematology at the UdS uses ObTiMA as their Clinical Trial Management System. The main components of ObTiMA are the Trial Builder and the Patient Data Management System.

**Trial Builder**

In the Trial Builder all aspects of a clinical trial can be specified. A trial chairman can define the outline and metadata of a trial in a master protocol. He can further setup treatment plans for guiding clinicians through individual patient treatment. Furthermore, the trial builder enables an ontology-based creation of CRFs. A graphical user interface allows defining content, navigation, and layout of CRFs to capture all patient data during a trial, e.g., medical findings or diagnostic data. The resulting descriptions are based on concepts from the Health Data Ontology Trunk (HDOT) for each CRF item along with metadata, e.g., data type and measurement unit, and used to setup the trial database.

The user interface makes the underlying aspects of the ontological metadata transparent and tries to overcome the gap between clinical practice and the actual logical representation of ontological concepts. When an item has been created based on a concept, its attributes are determined automatically, e.g., label, data type or answer possibilities but can be manually adopted.

Since many trials collect similar or equal data, it is possible to store components of or complete CRFs in a repository as templates. When setting-up a clinical trial, fitting CRF templates can either be directly reused or can be quickly created by plugging together existing CRF components. This in turn fosters the CRF standardization since CRFs can then readily be compared on the level single items (through ontological concepts) and also on the component level or in their entirety24.

The collected trial data is stored into ObTiMA’s trial database whose comprehensive metadata has been rendered in terms of the Health Data Ontology Trunk (HDOT). HDOT is a middle layer ontology developed by the Institute for Formal Ontology and Medical Information Science (IFOMIS) of the University of Saarland that provides the base to harmonize and link existing semantic resources under one umbrella [15].

**Patient Data Management System**

The second major functionality of ObTiMA is the Patient Data Management System. It guides the clinicians through the treatment of the individual patients according to the given treatment plans and provides a user interface to fill in the CRFs for a patient.

---

24 [http://www.optima.org](http://www.optima.org)
When the Patient Data Management is set up, the trial database is automatically derived from the ontology-based CRF definitions.

**Security**

Data safety and security as pseudonymization of private data is done according to roles and rights assigned to users of ObTiMA. The pseudonymization is realized through Custodix’s Privacy Enhanced Storage Framework (PESF).

2.3.2.2 Providing ObTiMA data for EURECA Data Push Service

ObTiMA will provide a mapping mechanism, which exports ObTiMA trial data to HL7 v3 messages. The main focus will be set on the export of the nephroblastoma trial data of SIOP/GPOH.

**Integration of the EURECA Core Data Set**

ObTiMA enables an ontology-based creation of CRFs. The EURECA Core Data Set will be integrated into ObTiMA, so the resulting descriptions can be stored based on these concepts.

**Export Interface**

An export interface (Figure 18) will allow users to initiate an export of clinical trial data from ObTiMA into a EURECA Data Push Service capable format.

![Figure 18 Screenshot export interface](image)

**HL7v3 message generation and pushing**
The HL7v3 messages generation will be realized through an XML generator that is based on a generic HL7 CDA XML schema. Therefore, the respective trial data of ObTiMA have to be mapped to the EURECA Common Data Model schema. The Data Push Service will receive the generated message by a WebService function. The message can then be validated and finally saved into the EURECA Data Warehouse. As long as CRF entries have are coded by the EURECA Core Data Set, the terminology linking service enables that these data will be stored semantically annotated.

2.4 PHR

2.4.1 Indivo-X

Indivo-X\(^ {25}\) is a personal health platform that tries to enable individuals to own and manage a complete, secure, digital copy of their health and wellness information. Indivo-X integrates health information across sites of care and over time. It is free, open-source and web-based and it uses open standards including those from the SMART\(^ {26}\) platforms project. A screenshot of the platform can be seen in Figure 19.

![Indivo-X Screenshot](image)

Figure 19 Example showing the available apps in Indivo-X

2.4.1.1 Data export from IndivoX

IndivoX already implements one mechanism for exporting Data, via the pull mechanism and within EURECA another mechanism has been implemented via pull mechanism. Those two methods are explained below:

\(^ {25}\) [http://indivohealth.org/](http://indivohealth.org/)

\(^ {26}\) [http://smartplatforms.org/](http://smartplatforms.org/)
**Push**

The push mechanism allows a user to push his own data to the Eureca DWH. To do that he has to select the appropriate app, and then to select the data that he/she would like to push to the DWH. The app that allows this functionality is shown in Figure 20.

![Indivohealth](image)

*Figure 20 Pushing data to the Eureca DWH*

To achieve that, initially the IndivoX API is used. However, the API allows an authenticated user to export data as JSON, RDF and XML data and the EURECA DWH imports messages as HL7 messages. So data exported from Indivo-X should be transformed in HL7-messages. This transformation is being done on-the-fly each time using a web service that has been created. A screenshot of some of the correspondences between the exported data and the HL7 messages are shown in Figure 21.
After the transformation, another web service pushes the HL7 messages to the DWH.

**Pull**

Besides pushing data to the DWH, Indivo-X allows a patient to share data with a specific carenet. The patient authenticates a Carenet, and then the specific Carenet can use the API calls to retrieve the XML, JSON or RDF data. A Carenet can be a doctor, a group of doctors, a research project or a team of people. A screenshot of this functionality is shown in Figure 22. A patient can drag-and-drop the different apps to the individual carenet to allow data sharing with that carenet. Then the people in that carenet are authorized to query the patient data when using Indivo-X API calls.
Figure 22 Authorising EURECA carenet to access patient
3 Conclusions

This document reports on the analysis performed for EURECA task 2.3. In EURECA task 2.3 a standards based interface layer towards external systems will be defined. The goal of the interface layer is to hide the heterogeneity of the external EHR, CDW or CDMS data sources for the EURECA services. The interface layer will be defined as a set of abstracted interfaces (e.g. providing functional access through standard EURECA APIs, providing direct data access, etc.).

As can be seen in Figure 23 the interface layer borders with the external systems. It provides interfaces to the external systems that can be used to push (ETL) data into the EURECA platform and it provides interfaces to the components in the EURECA platform to access data residing in external systems.

As a first step towards the interface layer, the relevant external data sources of the EURECA clinical partners have been assessed in this deliverable as these are very relevant for the clinical pilots. For the various data sources, this document describes how data can be exported or accessed.

For EHRs, we described the ETL process used to incorporate EHR systems in the EURECA environment that do not have a standardized query interface. All EHR systems of the clinical partners fall under this category. Additionally, the HL7 Clinical Research Filtered Query has been described. For CDWs we described tranSMART\(^{27}\), a clinical data warehouse that is currently gaining a lot of traction in the clinical research world. For CTMSes, we described openClinica and Obtaina. And finally, for PHRs we described Indivo X\(^{28}\).

\(^{27}\) http://www.transmartproject.org/
\(^{28}\) http://indivohealth.org/
3.1 EURECA Interface layer

The EURECA platform is developed as a multilayer architecture (see Figure 24) as is detailed deliverable 2.2 “Initial EURECA architecture”. The interface specifications for the different services in this multi-layered architecture will be discussed in deliverable 2.4, “Interoperability Framework and definition of Integration Profile” (month 24).

3.1.1 Semantic Integration Layer

The Semantic Integration Layer (Figure 25) is comprised by two main components; Common Data Model (CDM) and Core Dataset. CDM refers to all the information contained in the data warehouse and the schema of that data. Core Dataset is the ontology based information model of the platform.

This layer provides homogeneous access to the information contained in the CDM. For that purpose, a set of services has been developed that implements these components and the endpoints to retrieve and store information and this will be discussed in more detail, as stated before, in deliverable 2.4.
Figure 25 Semantic layer interfaces
4 REFERENCES


## 5 Appendix A Clinical Research Document Pre-Populated Data Content

<table>
<thead>
<tr>
<th>Optionality</th>
<th>Data Element or Section Name</th>
<th>Template ID</th>
<th>Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Header elements</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Required</td>
<td>Date of Birth</td>
<td>patientRole/patient/birthTime</td>
<td></td>
</tr>
<tr>
<td>Required</td>
<td>Gender</td>
<td>patientRole/patient/administrativeGenderCode</td>
<td></td>
</tr>
<tr>
<td>Optional</td>
<td>Ethnicity</td>
<td>patientRole/patient/ethnicGroupCode</td>
<td></td>
</tr>
<tr>
<td>Required if available</td>
<td>Race</td>
<td>patientRole/patient/raceCode</td>
<td></td>
</tr>
<tr>
<td><strong>Sections</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Required</td>
<td>Active problems</td>
<td>1.3.6.1.4.1.19376.1.5.3.1.3.6</td>
<td></td>
</tr>
<tr>
<td>Required if available</td>
<td>History of past illness</td>
<td>1.3.6.1.4.1.19376.1.5.3.1.3.8</td>
<td></td>
</tr>
<tr>
<td>Required if available</td>
<td>Procedures</td>
<td>2.16.840.1.113883.10.20.1.12</td>
<td></td>
</tr>
<tr>
<td>Required if available</td>
<td>Social History</td>
<td>1.3.6.1.4.1.19376.1.5.3.1.3.16</td>
<td></td>
</tr>
<tr>
<td>Required</td>
<td>Medications</td>
<td>1.3.6.1.4.1.19376.1.5.3.1.3.19</td>
<td></td>
</tr>
<tr>
<td>Required if available</td>
<td>Coded Vital Signs</td>
<td>1.3.6.1.4.1.19376.1.5.3.1.5.3.2</td>
<td></td>
</tr>
<tr>
<td>Required if available</td>
<td>Detailed Physical examination</td>
<td>1.3.6.1.4.1.19376.1.5.3.1.1.9.15</td>
<td></td>
</tr>
<tr>
<td>Required if available</td>
<td>Allergies and other adverse reactions</td>
<td>1.3.6.1.4.1.19376.1.5.3.1.3.13</td>
<td></td>
</tr>
<tr>
<td>Required if available</td>
<td>Coded Results</td>
<td>1.3.6.1.4.1.19376.1.5.3.1.3.28</td>
<td></td>
</tr>
</tbody>
</table>