Electronic Health Records for Clinical Research

Executive Summary for deliverable D7.1: Establish specification for data acquisition and standards used including a concept for local interfaces

Project acronym: EHR4CR
Project full title: Electronic Health Records for Clinical Research
Grant agreement no.: 115189
Budget: 16 million EURO
Start: 01.03.2011 - End: 28.02.2015
Website: www.ehr4cr.eu

Coordinator: AstraZeneca

Managing Entity: EuroRec

The EHR4CR project is partially funded by the IMI JU programme

Document description

Deliverable no:	D7.1 (executive summary)		
Deliverable title:	Establish specification for data acquisition and standards used including a concept for local interfaces		
Description:	First deliverable of Work Package 7		
Status:	Final		
Version:	0.2	Date:	14/08/2012
Deadline:			
Editors:	Fleur Fritz, Martin Dugas		

Document history

Date	Revision	Author(s)	Changes
08/08/2012	0.1	Fleur Fritz	Initial draft
10/08/2012	0.2	Martin Dugas	Review

Table of Content

1 Background

This document describes the Work Package 7 (WP7) deliverable of the first year within the EHR4CR project. The EHR4CR Project aims to create a platform to reuse patient level data from electronic health records for clinical research purposes. More information can be found at: http://www.ehr4cr.eu/.

The overall objective of WP7 is to demonstrate the functionality of the tools and services provided by the platform (Work Packages 3-6) and to evaluate the EHR4CR platform in the areas of clinical study design and execution with a specific focus towards a set of mutually acceptable medical domains agreed on by the pilot sites and EFPIA partners in alignment with Work Package 1.

The main interdependencies between the different work packages and respective deliverables for year 1 are depicted in Figure 1.

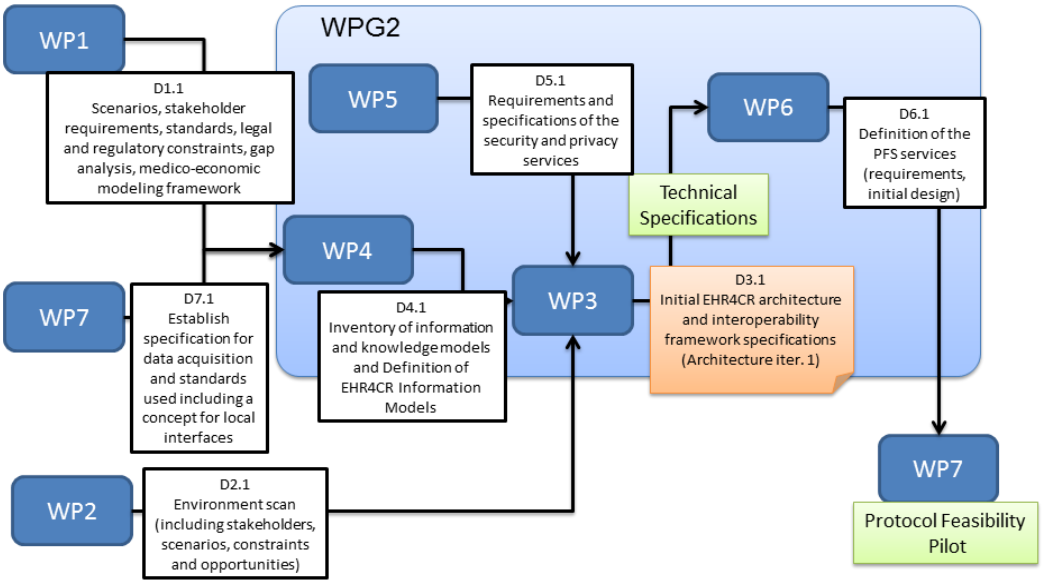

Figure 1 Work Package Relationships and Deliverables

2 WP7 Objectives of Year 1

WP7 will pilot the platform tools and services at 11 different hospital sites which in the text below are referred to as data provider sites. The piloting is divided into four clinical scenarios: protocol feasibility, site and patient identification and recruitment, clinical trial execution and adverse event monitoring. The first year of the project focuses on the first scenario of protocol feasibility, in particular to establish a specification for data acquisition and standards used including a concept for local interfaces. This deliverable focuses on an analysis of the local infrastructure, especially the data available at the pilot sites and the respective source systems. The WP7 output has been divided into tasks with specified activities for each year. The tasks for the first year are described below.

3 Task 7.1 Local Interfaces with EHR4CR Platform

Within the first year an inventory of local data items was built, including data elements from electronic health records (EHR), clinical data warehouses (CDW) and clinical data management systems (CDMS). As the piloting activities will be performed with actual data from real hospital IT environments, it was decided to agree upon a fixed set of data items currently available at the data provider sites and, for the year 1 version, needed to assess the feasibility of clinical trials. In accordance with WP 4 (Semantic interoperability) a limited set of data elements were defined in a multistage process which now represents the inventory of local data items of both EHR/CDW and CDMS.

4 Task 7.2 Protocol Feasibility

The year one activity for this task is to establish the inventory of data sources and to describe data schemes and data quality. This information is matched with the EFPIA clinical studies analysis, in particular with respect to inclusion and exclusion criteria, in order to select suitable pilot studies.

In the protocol feasibility phase of a clinical trial the proposed eligibility criteria need to be refined in order to ensure optimal patient recruitment. Those criteria which are computable are transformed into specific inclusion/exclusion criteria and entered using an EHR4CR query workbench application. The EHR4CR platform system will cascade the query across multiple hospitals and, at each site, will search for patients who match these inclusion/exclusion criteria. To pilot this scenario the following preparatory work is necessary. Before the platform can be deployed at the sites, an overview of the data sources at the data provider sites, their available data elements (based on the data inventory) and their local infrastructure needs to be compiled... In addition information about current and recent trials of the participating EFPIA companies running at the data provider sites are required to select suitable trials for the piloting scenarios.

5 Deliverable 7.1

Based on the tasks described above the output of the first year deliverable includes:

- First version of a data inventory consisting of a list of data elements derived from eligibility criteria of selected EFPIA studies suitable to perform feasibility analysis.
- Aggregated data export from the sites to evaluate the local data inventory and assess data availability.
- Matching of EFPIA studies with pilot sites for evaluation and proof of concept scenario.
- Local infrastructure survey and experiences from data exports to get an overview for local interfaces.

5.1 Data Inventory

The current version of the data inventory was defined in a multi-stage process. In a first step all EFPIA partners were asked to provide the most common used patient eligibility criteria. This list was then analyzed by all data providers to provide estimations about the availability of the items in the EHRs. During a consensus meeting additional data items were identified through an analysis of five different clinical trials. The resulting list formed the first version of the data inventory named “Top 82 Data Elements”.

To assess the data access and availability of those 82 data items each of the 11 data providers were asked to query them within their data sources. All pilot sites were asked to provide a set of metadata related to each data item as well as the frequency expressed in the percentage of how many patients of the cohort the respective data item was compiled. The template addressed only de-identified patient data.

On average 50 of the 82 data items were available in all the data provider sources (coded as “percentage not 0”), 32 data items were not available. The “Top Ten” of the data items which were most frequently available (on average 25% to 75% of these “Top Ten” data items were completed) are the following:

1. Gender (data group “Demographics“)
2. Date of Birth (data group “Demographics“)
3. Diagnosis Text (data group “Diagnosis“)
4. Diagnosis Date (data group “Diagnosis“)
5. Hospitalization admission date (data group “Demographics“)
6. Case Status (data group “Demographics“)
7. Procedure Text (data group “Procedure“)
8. Creatinine in serum (data group “Laboratory Findings“)
9. SGPT (ALT) in serum (data group “Laboratory Findings“)
10. Diagnosis Code (data group “Diagnosis“)

The complete analysis is available upon request.

5.2 Study Matching

All EFPIA companies were asked to deliver a list of recently completed (within the last two years) or ongoing studies (with completed feasibility phase) being conducted at the eleven data provider sites. This request resulted in 203 trials from Amgen, AstraZeneca, Bayer, GSK, Janssen, Merck, Novartis, Sanofi and Roche. With respect to time and effort a subset of 41 trials was selected from the list, for which the companies were asked to provide eligibility criteria.

The first requirement was that each of the EFPIA companies and all the pilot sites should be represented at least once.

The second requirement was that the number of studies per site should be up to to five, if possible.

From these 41 studies, a new subset was created with the goal, to have each EFPIA company included at least once and have trials for which every site gets approvals and has the data to run feasibility queries. The subset contained twelve trials as shown in the matrix below:

Internal Study Nr/Code	EFPIA Partner	Disease Area	AP-HP	FAU	HUG	KCL	MUW	U936	UCL	Univdun	UoG	uom	WWU	Total
11899	Bayer	Cardiovascular	2		1								1	4
20050182	Amgen	Oncology		1	1									2
27919	Merck	Nervous system disorders			1						1			2
BIO111482	GSK	Oncology								1			1	2
CENA713B2315	Novartis	Neurology						1						1
COU-AA-301	Janssen	Oncology	1											1
D4320C00015	AstraZeneca	Oncology	1								1		1	3
EFC11785	Sanofi	Oncology		1							1			2
NC25113	Roche	Cardiovascular and Metabolic								1		1		2
OMB112517	GSK	Oncology					1							1
CSPP100A2368	Novartis	Cardiovascular					1							1
EGF106708	GSK	Oncology				1*			1					2
Total			4	2	3	1	2	1	1	2	3	1	3	23
* = Trial not run at site														

5.3 Local Interfaces

To understand the different system architectures, communication standards and application programming interfaces (APIs) that are available at the different pilot sites it was decided to use the HL7 Electronic Health Record System Functional Model (EHR-S-FM) Information Infrastructure (IN) section which was sent to all data provider sites as a questionnaire.

The results provided a high level functional profile of the local hospital systems in the sections Security, Functionality, Health Record Information and Management, Standards-based Interoperability and Work-flow and Business Rules. This profile provides a framework to elicit requirements that will inform the analysis and design of the EHR4CR pivot representation with respect to the set of functions that could be expected in each local EHR, such as terminology, information models and semantic interoperability, allowing it to anticipate the tools and services needed for each of the scenarios. At the same time the profile will inform each pilot site of the set of functions that might be expected to be developed to interoperate with the pivot representation.

For example, it seems that no single system is currently able to deliver across all requirements to an EHR4CR platform without some additional modifications or the creation of a gateway application, such as a data mart.

6 Conclusion

WP 7 aims to provide a specification for data acquisition and standards used including a concept for local interfaces within the EHR4CR project. A first version of a data inventory was defined to enable piloting of the use case protocol feasibility. This inventory is based upon data elements from clinical trials and local hospital EHRs. Exports with de-identified patient level data were conducted successfully at 11 sites to determine availability of EHR data. In addition, the local EHR infrastructure at these sites in five European countries has been analysed and described.