ICT-2011-288048

EURECA

Enabling information re-Use by linking clinical Research and CAre

IP
Contract Nr: 288048

Deliverable 9.5: Report on preparation of the deployment environment for the clinical pilots

Due date of deliverable: (09-30-2014)
Actual submission date: (11-28-2014)

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

Responsible WP: FORTH

Revision: <Final>

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Service	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (excluding the Commission Services)	

0 DOCUMENT INFO

0.1 Author

Author	Company	E-mail
Lefteris Koumakis	FORTH	koumakis@ics.forth.gr
Haridimos Kondylakis	FORTH	kondylak@ics.forth.gr
Manolis Tsiknakis	FORTH	tsiknaki@ics.forth.gr
Kristof De Schepper	Custodix	kristof.deschepper@custodix.com
Jelle Van Den Driesche	Custodix	jelle.vandriesche@custodix.com
Kerstin Rohm	FhG IBMT	kerstin.rohm@ibmt.fraunhofer.de
Holger Stenzhorn	UdS	holger.stenzhorn@uniklinikum-saarland.de
Cyril Krykwinski	IJB	cyril.krykwinski@bordet.be
Monique Hendriks	Philips	monique.hendriks@philips.com
Sheng Yu	UOXF	sheng.yu@oncology.ox.ac.uk
Zhisheng Huang	VUA	huang@cs.vu.nl
Wytze Vlietstra	VUA	w.j.vlietstra@vu.nl
Jeroen Keijser	Philips	jeroen.keijser@philips.com

0.2 Documents history

Document version #	Date	Change
V0.1	08/01/2014	Starting version, template
V0.2	08/01/2014	Definition of ToC
V0.3	30/09/2014	First complete draft
V0.4	15/10/2014	Integrated version (send to WP members)
V0.5	12/11/2014	Updated version (send to project internal reviewers)
Sign off	21/11/2014	Signed off version (for approval to PMT members)
V1.0	28/11/2014	Approved Version to be submitted to EU

0.3 Document data

Keywords	Clinical Pilots, Preparation Environment
Editor Address data	Name: Manolis Tsiknakis Partner: FORTH Address: N. Plastira 100 Email: tsiknaki@ics.forth.gr
Delivery date	11-28-2014

0.4 Distribution list

Date	Issue	E-mailer

Table of Contents

1 Introduction

This deliverable reports on the preparation of the technical and procedural environment for the installation of EURECA technologies and tools. Specifically, we first identify and report the status of the deployment of the currently existing applications and services within the EURECA platform and also the status of the available data sets. Then, given the 15 clinical scenarios identified in the requirement elicitation phase of the project [1] [2], shown also in Table 1, this document describes the current status of the deployment environment for each one of these clinical scenarios.

Table 1: The list of the clinical scenarios

		Responsible	
	Scenarios	**Technical (Leading developer)**	**Clinical (Evaluation Leader)**
1	Personal medical information recommender	FORTH & FhG IAIS	UdS
2	Data mining of consultation	FhG IAIS	UdS
3	Contextualized overview	VUA	UOXF
4	Trial recruitment	Custodix	GBG
5	Protocol feasibility	Custodix	IJB, Maastro
6	Update of guidelines	VUA	Maastro
7	Hypothesis generation	UOXF	UOXF
8	Outcome prediction	UOXF	UOXF
9	Diagnostic classifier	UOXF	UOXF
10	Prediction of SAEs/SUSARs	Philips	UdS
11	Patient Diary & Long-term follow-up	FORTH	UOXF
12	Automatic detection and reporting of SAEs/SUSARs	FhG IBMT	UdS
13	Microbiology SAE	FhG IBMT	UdS
14	Reporting episodes of febrile neutropenia	IJB	IJB
15	Cancer registry and tumour bank reporting	IJB	IJB

For each clinical scenario the responsible partners for the development, the clinical deployment and the evaluation are identified. Moreover, we present the current status of each scenario. To be able to deploy each scenario all components within EURECA architecture should be available and integrated and the test datasets also. So the status on the aforementioned components is reported and also the plans for their integration. Finally the exact environment for the deployment and the corresponding setup are explained and commented.

1.1 Structure of the Deliverable

The remaining of this document is structured as follows:

Chapter 2 presents an overview of the applications and services and their deployment environments. Also the current status of the available data sets is described.

Then, Chapter 3 reports the current status on the deployment for each one of the aforementioned clinical scenarios. The responsible partners for evaluation and technical development are identified and the necessary components as well. Then the status of each one of these components is reported and details of all the components required by a clinical scenario and their integration are provided. A preliminary planning follows including a short description of the deployment environment, the setup and the data that will be used.
Finally, Chapter 4 concludes this deliverables and provides an outlook for the next steps.

2 Deployment Environment for Clinical Scenarios collection utilities

2.1 Component Overview

For each EURECA service and application there will be running three deployment environments. One dedicated to the further development of the service/application, one stage environment for the migration between development and pre-production environment and one pre-production environment where the stable applications and services will finally run. Table 2 summarizes the current status on each application and service in the development environment set up so far grouped by the architectural layer they belong to [4]. The other 2 environments will be set-up the following months. Moreover, we have to note that the services of the platform/infrastructure/security layers will only have one deployment environment environments grouped by to the architectural layer they belong to [4].

Table 2: The status of the EURECA applications and services in the three deployment environments

EURECA component	Development environment
Yakobo (local protocol feasibility)	Yaboko Trial feasibility v0.1.0
Yakobo (trial recruitment)	Yaboko Trial recruitment v0.1.0
Personal Health Record Extensions	IndivoXv0.1.0
Personal Medical Information Recommender (PMIR)	PMIR in IndivoX 0.1.0
Microbiology Module	Microbiology Module in ObTiMA v1.0.0
Outcome Prediction tool	N.A.
SAE Prediction tool	N.A.
Application Services	
TriaMeRe (trial management)	TriaMeRe_v1.0.0.338,
Query Engine	Query_Engine_v2.0.0
Single Patient Query Engine	N.A.
Semantic Integration Services	
Data Push Service (ETL)	data_push_v1.0.0
	Collaborative Tools DataPush
	Fraunhofer DataPush
	Trial Recruitment DataPush
Query Builder Service	query_builder_v1.0.0
Query Execution Service (CDM/CIM)	query_execution_v1.0.0
	Trial Recruitment Exe
	New Collaborative Tools Exe
	Collaborative Tools Exe
	St Gallen Exe
	Fraunhofer Exe

	Yakobo Exe
Terminology Linking Service (BIOportal)	terminology_linking_v1.0.0
Free Text Service	free_text_v1.0.0
Core Dataset Services	core_dataset_v1.0.0
Provenance	N.A.
Autocomplete Service	auto_complete_v1.0.0
Data Access Services	
Data Export Guidelines	data_export_guidelines_v1.0.0
EURECA Common Data Model	common_data_model_v2.6.2
Infrastructure Services	
Service Registry	Service_Registry_v1.0
Messaging Service	N.A.
Platform Management Services	
Platform Management Service	N.A.
Security Services	
Identity Manager	IdM_v1.2.3
Security Token Service	Security_Token_Service_v1.2.3
Identity Provider	Identity_Provider_v1.2.3
Authorisation Service	Authorisation_Service_v1.0.0
Auditing Service	N.A.
Audit Manager	N.A.
Policy Administration Point	Policy Administration Point v1.2.3
Patient Identity Management System	PIMS_v3.0.0

2.2 Available data sets

Table 3 presents the status of the available datasets within EURECA platform. As we can notice there are 10000 patients artificially created using the APDG engine, 4673 real patients from GBG, 3120 patients from SIOP, 5 fake patients from Fraunhofer, 219 real patients with breast cancer from Oxford, 80 fake patients from IJB, 409 real patients from MAASTRO and finally 29 real patients from IJB. All data are available, ready to be exploited within the EURECA clinical scenarios and are described in the respective deliverables of WP9 in more detail and more specifically in D9.1 [3]

Table 3: The list of available data sets

Dataset	# patients	Already accessible though EURECA CDM
APDG (artificially generated patients)	10000	yes
GBG	4673	yes
SIOP	3120	yes
Fraunhofer	5	yes
Oxford	219	yes
IJB (artificially generated patients)	80	yes
MAASTRO Lung Survival Dataset	409	yes
IJB	29	yes

3 Report on the Deployment of the Clinical Scenarios

This chapter describes the deployment environment and the current status of the 15 clinical scenarios.

3.1 Personal medical information recommender

The Personal Medical Information Recommender (PMIR) is a tool that allows clinicians and patients to obtain objective information (e.g. about treatments) about the patient/disease.

The interesting web documents are identified by doctors who user PMIR Management interface to insert them to the internal database of the system. Then those documents are semantically annotated using the XEROX NLP tool and the annotations are stored in the database as well. Then those annotations are exploited by the PMIR Search engine to allow searching in the information indexed in those web documents. Besides searching, PMIR also automatically recommends interesting web documents to the patients using data mining techniques.

Responsible Partner for Evaluation	UdS
Target Group	Patients
Leading Developer Partner	FORTH/FhG IAIS
Technical Collaborators	XEROX
Short Description of the scenario	Doctors insert useful documents. Those documents are annotated using XEROX tool and stored in the database. Then patients are allowed to search for useful information using the PMIR Semantic Search engine through the PHR system. User interaction is logged and then useful recommendations appear to them using data mining techniques.
Components Involved	**Component Name**: PMIR Semantic annotator **Short Description**: The system used for insert/delete/update and annotate the interesting web documents. **Status**. Already Available **Component Name**: XEROX NLP tool **Short Description**: The tool used to annotate the web documents with ontology terms **Status**. Already Available **Component Name**: PHR client **Short Description**: The system that the patients will be using in order to be able to search in those web documents. **Status**. Already Available **Component Name**: PMIR Client **Short Description**: The specific app inside the PHR

	system that allows searching for useful information **Status**. Already Available **Component Name**: PMIR Service **Short Description**: The service that used data mining to recommend automatically useful documents to the patient **Status**. Already Available.
Components Integration	**Short description:** The components are already described within the PHR. **Date**: Components are already integrated **Effort**: No more effort is required for the integration
Deployment environment	**Location:** Clinical Setting **Restrictions**: Local Virtual Machine, deployed in the UdS hospital. The virtual machine is already available waiting to be deployed to the hospital.
Environment Setup	Virtual machine of the tools with no external internet connection. 10 different people will use the system. They will be clinicians, molecular biologists and data managers.
Data Needed	Data needed (useful web documents) already available. The data needed are not patient data but external web documents available freely at the web.
Preliminary Planning	• PMIR Data Mining algorithms finalized by December 2014 • Technical Evaluation (precision, recall) finalized by March 2015 • Usability Evaluation finalized by March 2015 • Clinical testing on April for 1 month

3.2 Data mining of consultation

In prospective clinical trials many consultations are performed. A part of the questions of such consultations are repeatedly asked. It would be helpful to generate an automatic answer to questions asked during consultations.
For this scenario the following steps are relevant:

- Select the trial and the documentation of the consultations
- This documentation can be available in a structured or unstructured way
- In case of an unstructured way data mining of the text is needed to extract relevant information that will be stored in a structured way
- The structured data of the trial will be used as a source for data mining for a specific question of a consultation
- All answers to the same consultation question will be selected and analysed to create an answer to the consultation question
- This answer will be validated by literature mining
- As a results of this validation a final answer will be created

Responsible Partner for Evaluation	UdS
Target Group	Patients

Leading Developer Partner	FhG IAIS
Technical Collaborators	-
Short Description of the scenario	In prospective clinical trials many consultations are performed. A part of the questions of such consultations are repeatedly asked. It would be helpful to generate an automatic answer to questions asked during consultations.
Components Involved	**Component Name**: Model Generation Client **Short Description**: The system that the patients will be using in order to be able to search in those web documents. **Status**. Already Available **Component Name**: Consultation Client **Short Description**: The system that the patients will be using in order to be able to search in those web documents. **Status**. It will available at the end of December 2014
Components Integration	**Short description:** The integration of the components is a trivial task **Date**: It will available at the end of December 2014 **Effort**: No more effort is required for the integration
Deployment environment	**Location:** Clinical Setting **Restrictions**: Local Virtual Machine, deployed in the UdS hospital. The virtual machine is already available waiting to be deployed to the hospital.
Environment Setup	Virtual machine of the tools with no external internet connection. For 30 upcoming consultations within the SIOP trial the tool will run and the recommendations given by the tool will be examined
Data Needed	Personal Consultations from Norbert Graf of the SIOP trial. The data are already released.
Preliminary Planning	• Tolls and algorithms finalized by December 20154. • Technical Evaluation finalized by December 2015 • Clinical testing finalized by March 2015

3.3 Contextualized overview

A clinician, while looking at a patient record (EHR), is presented with an overview of the availability of relevant documents from a wide variety of sources.
The overview consists of summaries of categories of relevant documents.
For each category we see in what way it is related to the current patient case, and how many documents there are. Examples could be:

- "10 records of patients with same disease and age"
- "2 trials with matching eligibility criteria"
- "50 trials with matching criteria except age"

This gives the clinician an overview of what's available. He/she can quickly assess the value of each result category, and select them to see more results.

The following table summarizes the status of the deployment for the contextualized overview scenario.

Table 4: Contextualized overview scenario

Responsible Partner for Evaluation	UOXF
Target Group	Clinicians
Leading Developer Partner	VUA
Technical Collaborators	None
Short Description of the scenario	The physician facing contextualization task aims to aid physicians in their information retrieval tasks during a patient – physician consultation. By integrating the patient data contained in the EHR, results of queries can be specialized to the patient's specific context, which is thought to enhance efficiency and quality of care.
Components Involved	**Component Name**: Query Execution Service **Short Description:** A Service allowing to query the patient data required for this clinical scenario **Status:** Already available. **Component Name**: Contextualization Application **Short Description:** The application allowing the contextualization of useful information for a clinician as already described. **Status:** December 2014.
Components Integration	**Short description:** The contextualization application should be integrated with the query execution service. **Date**: As soon as possible **Effort**: Getting through the security layer, obtaining relevant SPARQL query.
Deployment environment	**Location:** In case of extension: clinical setting, if no extension: research setting **Restrictions**: Local Virtual Machine
Environment Setup	Virtual machine of the tools with no external internet connection.
Data Needed	MAASTRO, Oxford, or prospective in Oxford. Fields to be extracted provided by VUA and feasibility to be confirmed by MAASTRO and Oxford.
Evaluation Factors	**Proposed type of evaluation** : Clinical evaluation

	Preconditions: if extension approved **Evaluation factors/parameters:** To be discussed, See planning below
Preliminary Planning	• Tool completed by December 2014 • Mapping completed start of 2015 • Rules for first round and interface evaluation by Feb 2015 • First round of evaluation on fake data Feb 2015 • Interface evaluation Feb 2015 • Design of clinical questionnaire Feb-June 2015 • Feedback to developer and update of interface by June 2015 • Clinical evaluation June 2015-Oct 2015

3.4 Trial recruitment

The recruitment of a clinical trial is currently a very time-consuming task. By using NLP techniques to understand inclusion and exclusion criteria and matching those criteria to the Electronic Health Record system, the selection of eligible patients can be greatly improved.
The following table summarizes the status of the deployment for the trial recruitment scenario.

Table 5: Trial recruitment scenario

Responsible Partner for Evaluation	Custodix
Target Group	Clinical Trial recruiters e.g. research nurses
Leading Developer Partner	Custodix
Technical Collaborators	Philips, UPM
Short Description of the scenario	The evaluation of the trial recruitment tool will be done by clinical trial recruiters. Virtual machines will be configured at the sites of the service owners (UPM, Custodix and Philips). On these machines the trial recruitment services will be deployed and integrated. Patient data (provided by the evaluation site) will be ETL'ed to the CDM. Trial metadata and formalized eligibility criteria will be stored in the trial management service. The end-user tool itself will be installed at the validation site. The trial recruiters will use the tool to evaluate the "recruit patient for a given trial" scenario.
Components Involved	**Component Name:** Trial Recruitment Client **Short Description:** The application front-end of the trial recruitment is provided by this client. It offers a graphical user interface (GUI) that offers the functionality of the trial recruitment service in an intuitive

	and user-friendly way to the end-user (the investigator). **Status:** Available (Yakobo standalone tool) **Component Name:** Trial Recruitment Service **Short Description:** The main driver component of the recruitment is the trial recruitment service. It offers the functionality needed to complete the different steps in the recruitment flow. For this, it integrates and connects the other services that are needed in the trial recruitment. **Status:** Available (Yakobo standalone tool) **Component Name:** Locker Service **Short Description:** The Locker service provides private document and blob storage for services (on a per-user bases). Other services can use this service to store for instance user specific settings, documents, etc. **Status:** Available (Philips Service) **Component Name:** Trial Management Service **Short Description:** This service is responsible for providing trial registering, querying and editing functionality. It offers registering services that enable a trial administrator to generate general trial information, add eligibility criteria to a trial, define different trial arms in a trial, etc. All this trial information is stored in a trial meta-data repository of the site. Another important service is that the information stored in the trial repository, for example the list of trials, can be easily accessed by other services of the site. **Status:** Available (Philips Service) **Component Name:** Query Engine **Short Description:** An eligibility criterion is matched with the information of a selected patient in the criteria matching service. It provides an interface that enables other EURECA services to send matching requests. The query engine will query the requested information of the patient by sending the query that is included in the eligibility criterion to the CIM based query service of the different available data warehouses. The eligibility criterion itself is retrieved from the trial management service. The outcome of the matching is sent back to the requesting service. **Status:** Available (Custodix Service) **Component Name:** Free Text Service **Short Description:** This service is responsible for free text querying of the different available data sources (e.g. the EHR data warehouse) in the EURECA platform. It offers free text searching functionality in

	order to query structured and unstructured data. **Status:** Under construction (UPM Service) **Component Name:** Query Execution Service **Short Description:** This service provides functionality to query the datasets of the EHR and other data warehouses available on the site through the semantic layer. It abstracts the underlying data sources for the upper EURECA services and presents data to applications according to a single integrated data model. More information about this service can be found in the semantic layer view section. **Status:** Available (UPM Service) **Component Name:** Auto Complete Service **Short Description:** This service is responsible for extract concepts present on the Core Dataset and the metadata of this concept on the CDM. **Status:** Available (UPM Service) **Component Name:** Query Builder Service **Short Description:** This service is responsible of generate query template in XML format for querying on the Query Engine Service. **Status:** Available (UPM Service)
Components Integration	**Short description:** The integration of the different components was done in an iterative way. The first (main) integration round was done the months before the first review. After this the components where refined and gradually further integrated. Final integration is targeted in Q1 2015. **Date**: 01/04/2013 (last main integration) **Effort**: integration effort was done in iterative steps

Deployment environment	**Location:**
	• Custodix VMs
	  ○ Query Builder
	• UPM VMs
	  ○ Query Execution Service
	  ○ Query Builder Service
	  ○ Auto Complete Service
	  ○ Free Text Service
	• Philips VMs
	  ○ Locker Service
	  ○ Trial Management Service
	• Philips Tools:
	  ○ Yakobo Recruitment Tool

	• Custodix Security VMs
	  ○ Authorisation Service
	  ○ Authentication Service
	  ○ Auditing Service
	Restrictions:
	The services working with patient data will be installed at the hospital site (Query builder, UPM services, Yakobo tool). The clinical sites that are going to test the service are IJB, Maastro and GBG. From each one at least 3 nurses will be used to test the system.

Environment Setup	Locker Service, Trial Management Service • Java environment, JBoss application server, MySQL Yakobo Recruitment Tool • Windows 7 or higher, .net framework 4.0 or higher, high resolution monitor Query Builder, Authorisation Service, Authentication Service • Java environment, Tomcat application server, MySQL Query Execution Service, Auto Complete Service, Query Builder Service, Free Text Service • Java environment, Tomcat application server, MySQL
Data Needed	• GBG datasets • IJB datasets • Maastro datasets Initial list, depends on the evaluation sites (restriction on datasets)
Evaluation Factors	**Proposed type of evaluation** : The end-user tool (Yakobo) will be evaluated by a selected set of research nurses and others with experience with trial recruitment. The evaluation will be done in sessions, structured which are structured like this: • Introduction of tool (presentation by developer) • Let the user work with the tool • Formal feedback from end-user (questionnaires) • Informal feedback from end-user **Preconditions**: • Test users should have experience in the trial recruitment domain **Evaluation factors/parameters**: Questionnaire results
Preliminary	Current Preliminary planning:

Planning	• Finishing development of the different components (bug fixing) – end of December • Discussing with evaluation sites the required setup, the data used, restrictions, etc. – Q1 2015 • Final integration of the components – Q1 2015 • Setting up the environments for the different evaluations (preparing data, creating test scenarios and setting up configurations) – Q2 2015 • Testing the different setups – Q2 2015 • Evaluation at sites – Q2 2015

3.5 Protocol feasibility

This scenario describes if a new clinical trial is feasible to start according to the estimation of recruitment potential.
Two versions of this scenario are possible:

- Based on EHR/PHR/HIS data
- Based on other data sources

For legal reasons the first version can only be done at the hospital site.
This scenario should always be used before the scenario for the section and inclusion of patients into trials.
The following steps are needed:

- Patient data from HIS/EHR/PHR are exported and anonymized (data warehouse)
- Hospital can do data mining to select the cohort of patients that fits recruitment criteria best
- The number of aggregated data gives the answer if the protocol or the trial is feasible to develop

The following table summarizes the status of the deployment for the protocol feasibility scenario.

Table 6: Protocol feasibility scenario

Responsible Partner for Evaluation	Custodix
Target Group	Study managers
Leading Developer Partner	Custodix
Technical Collaborators	UPM
Short Description of the scenario	The evaluation of the trial feasibility tool will be done by study managers. Virtual machines will be configured at the sites of the service owners (UPM, Custodix and Philips). On these machines the trial feasibility services will be deployed and integrated. Patient data (provided by the evaluation site) will be ETL'ed to the CDM.

	Trial metadata and formalized eligibility criteria will be stored in the trial management service. The end-user tool itself will be installed at the validation site. The trial recruiters will use the tool to evaluate the "assesses the feasibility of a clinical protocol under development" scenario.
Components Involved	**Component Name:** Distributed Protocol Feasibility Client **Short Description:** The application front-end of the distributed protocol feasiblity is provided by this client. It offers a graphical user interface (GUI) that offers the functionality of the distributed protocol feasiblity service in an intuitive and user-friendly way to the end-user (the investigator). **Status:** End of December (Custodix web client) **Component Name:** Distributed Protocol Feasibility Service **Short Description:** The main driver component of the feasiblity is the distributed protocol feasiblity service. It offers the functionality needed to complete the different steps in the feasiblity flow. For this, it integrates and connects the other services that are needed in the distributed protocol feasiblity. **Status:** End of December (Custodix Service) **Component Name:** Trial Management Service **Short Description:** This service is responsible for providing trial registering, querying and editing functionality. It offers registering services that enable a trial administrator to generate general trial information, add eligibility criteria to a trial, define different trial arms in a trial, etc. All this trial information is stored in a trial meta-data repository of the site. Another important service is that the information stored in the trial repository, for example the list of trials, can be easily accessed by other services of the site. **Status:** Available (Philips Service) **Component Name:** Service Registry **Short Description:** This service is responsible for providing service registering, querying and editing functionality. It allows businesses to efficiently discover and communicate with each other using web services. **Status:** End of December (Custodix Service) **Component Name:** Query Engine **Short Description:** An eligibility criterion is matched with the information of a selected patient in the criteria matching service. It provides an interface that enables other EURECA services to send matching requests. The query engine will query the requested information of the patient by sending the query that is included in the eligibility criterion to the CIM based query service of the different available data warehouses. The eligibility criterion itself is retrieved from the trial management

	service. The outcome of the matching is sent back to the requesting service. **Status:** Available (Custodix Service) **Component Name:** Query Execution Service **Short Description:** This service provides functionality to query the datasets of the EHR and other data warehouses available on the site through the semantic layer. It abstracts the underlying data sources for the upper EURECA services and presents data to applications according to a single integrated data model. More information about this service can be found in the semantic layer view section. **Status:** Available (UPM Service) **Component Name:** Auto Complete Service **Short Description:** This service is responsible for extract concepts present on the Core Dataset and the metadata of this concept on the CDM. **Status:** Available (UPM Service) **Component Name:** Query Builder Service **Short Description:** This service is responsible of generate query template in XML format for querying on the Query Engine Service. **Status:** Available (UPM Service)
Components Integration	**Short description:** The integration of the different components will be done in an iterative way. Components will be refined and gradually further integrated. Final integration is targeted in Q1 2015. **Date**: - (last main integration) **Effort**: integration effort will be done in iterative steps

Deployment environment	**Location:**
	• Custodix VMs
	  ○ Distributed protocol feasibility client
	  ○ Distributed protocol feasibility service
	  ○ Service registry
	  ○ Query engine
	• UPM VMs
	  ○ Query Execution Service
	  ○ Query Builder Service
	  ○ Auto Complete Service
	• Philips VMs
	  ○ Trial Management Service
	• Custodix Security VMs
	  ○ Authorisation Service
	  ○ Authentication Service
	  ○ Auditing Service
	
	Restrictions:
	• The services working with patient data should be installed at the hospital site. More specifically will be installed in the IJB,

	MAASTRO and GBG and tested by at least 3 study managers.
Environment Setup	Trial Management Service • Java environment, JBoss application server, MySQL Query Builder, Authorisation Service, Authentication Service, Service Registry, Distributed Protocol Feasibility Client, Distributed Protocol Feasibility Service • Java environment, Tomcat application server, MySQL Query Execution Service, Auto Complete Service, Query Builder Service • Java environment, Tomcat application server, MySQL
Data Needed	• GBG datasets • IJB datasets Initial list, depends on the evaluation sites (restriction on datasets)
Evaluation Factors	**Proposed type of evaluation** : The web front-end will be evaluated by a selected set of study managers and others with experience with protocol feasiblity. The evaluation will be done in sessions, structured which are structured like this: • Introduction of tool (presentation by developer) • Let the user work with the tool • Formal feedback from end-user (questionnaires) • Informal feedback from end-user **Preconditions**: • Test users should have experience in the feasibility domain **Evaluation factors/parameters**: • Questionnaire results • Timing tables
Preliminary Planning	Current Preliminary planning: • Further develop the service registry, distributed protocol feasibility client, distributed protocol feasibility service – end of December • Discussing with evaluation sites the required setup, the data used, restrictions, etc. – Q1 2015 • Final integration of the components – Q1 2015 • Setting up the environments for the different evaluations (preparing data, creating test scenarios and setting up configurations) – Q2 2015 • Testing the different setups – Q2 2015 • Evaluation at sites – Q2 2015

3.6 Update of guidelines

This scenario describes how guidelines can be developed and regularly updated from data mining of clinical trials and literature.
The following steps are needed:

- Select a guideline and items in the guideline that should be updated
- Use these items for data mining in CT/HIS, Literature and trial databases
- Search only for data beyond the date of the guideline
- After data collection do an automatic listing of the updated items
- The end-user will select the relevant items from these listings
- These updated items will replace the old items in the guideline
- The guideline is updated and a new version with the date of update is stored

The following table summarizes the status of the deployment for the update of guidelines scenario.

Table 7: Update of guidelines scenario

Responsible Partner for Evaluation	Maastro
Target Group	Guideline developers/ clinical professionals
Leading Developer Partner	VUA
Technical Collaborators	
Short Description of the scenario	A guideline developer or clinical professional check the existing guideline recommendation, and select one to see whether or not there exists any new finding for it.
Components Involved	**Component Name**: Guideline Update Client **Short Description:** The client with the graphical user interface of the tool. **Status:** Will be available on January 2015 **Component Name**: Guideline Update Service **Short Description:** The guideline update service implements the functional part of the step by step process for checking the new guideline evidences **Status:** Will be available on January 2015 **Component Name**: Literature Mining Service **Short Description:** Literature Mining Service provides analysis of relevant literatures, based on selected relevance measure. **Status:** Will be available on May 2015 **Component Name**: Guideline Update Management **Short Description:** Guideline Update Management provides the interface for the guideline administrator to show the evaluation and analysis results of the guideline update processing. The results include new

	research findings and their evidence levels which are evaluated by using the methods from evidence-based medicine. **Status: :** Will be available on May 2015
Components Integration	**Short description:** A component of guideline update will be developed and integrated in SemanticCT system, and the EURECA platform. **Date**: End of 2014 **Effort**:
Deployment environment	**Location:** A local machine (**wasp.cs.vu.nl)** at VUA implementing the Java Services of the aforementioned components. **Restrictions**: none
Environment Setup	JAVA Runtime environment 1.6 or higher
Data Needed	None
Evaluation Factors	**Proposed type of evaluation** : Taking two versions of a guideline and check whether or not the new evidences in the second version have been founded, from the request on the update of recommendation in the first version of guidelines. **Preconditions**: **Evaluation factors/parameters**: Not defined yet.
Preliminary Planning	Components finalized till May 2015

3.7 Hypothesis generation

New hypothesis are needed to start new clinical trials. Promising results from the literature can be automatically identified via data-mining and serve as base for new hypothesis.
Before starting a new clinical trial a new research question is needed. Such a question is of utmost importance and is part of hypothesis generation. Analysing all available data from previous trials, guidelines, literature and others, can support this process. It can also help to find biomarkers that are relevant for the disease suggesting their use in the trial for evaluation or validation purposes.
A trial chair, while designing a new trial, is supported in the design of the eligibility conditions (inclusion and exclusion) based on background information:

- design of previous trials
- results of previous trials
- SUSAR reports from previous trials
- Previous epidemiological (retrospective) studies.
- published literature

Responsible Partner for Evaluation	UOXF
Target Group	Bioinformaticians or clinical researchers
Leading Developer Partner	UOXF

Technical Collaborators	UOXF
Short Description of the scenario	A tool that allows clinicians to generate and evaluate hypotheses in the context of designing a clinical trial. Data are uploaded via KDF functionality. Script are uploaded via KDF. Jobs are managed via KDF. The evaluation will consist of interviews with the users, who will be presented with the tool and a questionnaire to assess the usefulness of this tool.
Components Involved	**Component Name**: **KDF web client** Short Description: • The KDF web client is the GUI of the Knowledge Discovery Framework (KDF) Status: • KDF web client ready from FhG. Algorithm available from UOXF. NCBI data retrieved via R script. **Component Name**: **KDF manager** Short Description: • The central module which controls the workflow of the KDF. The web based GUI calls the functions of KDF manager to perform operations in the analysis process. Status: • Java component for data mapping/retrieving from the semantic interoperability layer available but not yet integrated to KDF, work in progress. • Available / development goes on **Component Name**: **KDF script processing module** Short Description: • The component to process scripts that are loaded into the KDF Status: • The component ready from FhG **Component Name**: **KDF local storage module** Short Description: • The component to process scripts that are loaded into the KDF Status: • The component ready to process and store data locally but development goes on to handle other data types
Components Integration	**Short description:** Java component for data mapping/retrieving available (UPM) but not yet integrated to KDF, work in progress. **Date**: December **Effort**: ready but integration with KDF is work in progress by FhG.
Deployment environment	**Location:** • Instance of KDF running at UOXF. Script currently running on UOXF instance. **Restrictions**:

	None known
Environment Setup	Used via the KDF. Local configuration.
Data Needed	Can be applied to different data types depending on application. At the moment it is used on Oxford, GBG and fake datasets. Data available for testing. NCBI data needed but retrieved via R script.
Evaluation Factors	**Proposed type of evaluation** : • proof of concept **Preconditions**: • None
Preliminary Planning	December 2014 Interviews will be set up to complete the evaluation procedure. A virtual machine that has the tool running and any clinical data loaded will be presented to the interviewee. A questionnaire will be carefully designed to assess the usefulness of the tool. During the interview the user (bioinformatician or clinician) will be introduced to the functionalities of the tool and as the user is using the tool, the answers to the questions will be recorded. After the interview the virtual machine will be destroyed along with any sensitive data.

3.8 Outcome prediction

In rapid learning research we want to learn and validate outcome prediction models from routine patient care data.

The following table summarizes the status of the deployment for the outcome prediction scenario.

Table 8: Outcome prediction scenario

Responsible Partner for Evaluation	UOXF
Target Group	Bioinformaticians or clinical researchers
Leading Developer Partner	UOXF
Technical Collaborators	UOXF
Short Description of the scenario	A tool that allows clinicians or researcher to predict the outcome of treatment for a patient. Data are uploaded via KDF functionality. Scripts are uploaded via KDF. Jobs are managed via KDF. The evaluation will consist of interviews with the users, who will be presented with the tool and a questionnaire to assess the usefulness of this tool.
Components Involved	**Component Name**: KDF web client Short Description: • The KDF web client is the GUI of the Knowledge Discovery Framework (KDF)

	Status • KDF web client ready from FhG. Algorithm for prediction available from UOXF. Script containing algorithm uploaded and tested on KDF by UOXF (previous version of KDF, needs to be rechecked). **Component Name**: KDF manager Short Description: • The central module which controls the workflow of the KDF. The web based GUI calls the functions of KDF manager to perform operations in the analysis process. Status: • Java component for data mapping/retrieving from the semantic interoperability layer available but not yet integrated to KDF, work in progress. Available / development goes on **Component Name**: **KDF script processing module** Short Description: • The component to process scripts that are loaded into the KDF Status: • The component ready from FhG **Component Name**: **KDF local storage module** Short Description: • The component to process scripts that are loaded into the KDF Status: The component ready to process and store data locally but development goes on to handle other data types.
Components Integration	**Short description:** Java component for data mapping/retrieving available (UPM) but not yet integrated to KDF, work in progress. **Date**: (December?) to be confirmed by FhG **Effort**: ready but integration with KDF is work in progress by FhG.
Deployment environment	**Location:** (physical location) Instance of KDF running at UOXF. Script currently running on UOXF instance. **Restrictions**: None known
Environment Setup	Used via the KDF. Local configuration.
Data Needed	(Oxford, GBG and fake datasets. Data available.
Evaluation Factors	**Proposed type of evaluation** : Selected user evaluation **Preconditions**: None **Evaluation factors/parameters**: To be defined
Preliminary Planning	December 2014 Interviews will be set up to complete the evaluation procedure. A virtual machine that has the tool running and any clinical data loaded will be presented to the interviewee. A questionnaire will be carefully designed to assess the usefulness of the tool. During the

	interview the user (bioinformatician or clinician) will be introduced to the functionalities of the tool and as the user is using the tool, the answers to the questions will be recorded. After the interview the virtual machine will be destroyed along with any sensitive data.

3.9 Diagnostic classifier

This scenario describes the process of training a diagnostic classifier to classify sarcoma based on patient's clinical data and genomic data. It would be a useful tool to improve the diagnosis of sarcoma because there are many types of this bone and soft tissue cancer and currently not all patients are correctly diagnosed.
Relevant clinical data is collected from the clinical systems in the Oxford University Hospitals trust in this step. The method of collection includes a potential automatic process that pulls data from a cancer-oriented data repository called the Cancer Outcome Service Dataset and a few manual steps to collect other related data.
The local technical teams are contacted at each clinical site by the chair via emails or telephone to start the training; they receive the study design, data pre-processing guidelines and the QC guidelines for the initial training process and every time that there are modification to the dataset specification.
The technical team at each site starts the pre-process of the data as required by the data pre-processing guidelines (e.g. image segmentation and image processing, genomic data summarization and normalization, patient data anonymization etc).
Pre-processed data and QC statistics are generated at each site; the anonymous data are generated.

The following table summarizes the status of the deployment for the diagnostic classifier scenario.

Table 9: Diagnostic classifier scenario

Responsible Partner for Evaluation	UOXF
Target Group	Bioinformaticians or clinical researchers
Leading Developer Partner	UOXF
Technical Collaborators	UOXF
Short Description of the scenario	A tool that allows clinicians or researcher to classify patients into diagnostic groups. Data are uploaded via KDF functionality. Script are uploaded via KDF. Jobs are managed via KDF. The evaluation will consist of interviews with the users, who will be presented with the tool and a questionnaire to assess the usefulness of this tool.
Components Involved	**Component Name**: KDF web client Short Description: • The KDF web client is the GUI of the Knowledge Discovery Framework (KDF) Status:

	• KDF web client ready from FhG. Algorithm for prediction available from UOXF. Script containing algorithm uploaded and tested on KDF by UOXF (previous version of KDF, needs to be rechecked). **Component Name**: KDF manager Short Description: • The central module which controls the workflow of the KDF. The web based GUI calls the functions of KDF manager to perform operations in the analysis process. Status: • Java component for data mapping/retrieving from the semantic interoperability layer available but not yet integrated to KDF, work in progress. Available / development goes on **Component Name**: **KDF script processing module** Short Description: • The component to process scripts that are loaded into the KDF Status: • The component ready from FhG **Component Name**: **KDF local storage module** Short Description: • The component to process scripts that are loaded into the KDF Status: • The component ready to process and store data locally but development goes on to handle other data types
Components Integration	**Short description:** Java component for data mapping/retrieving available (UPM) but not yet integrated to KDF, work in progress. **Date**: December **Effort**: • ready but integration with KDF is work in progress by FhG.
Deployment environment	**Location:** (physical location) • Instance of KDF running at UOXF. Script currently running on UOXF instance. **Restrictions**: • None known
Environment Setup	Used via the KDF. Local configuration.
Data Needed	Can be applied to different data types depending on application. At the moment it is used on Oxford, GBG and fake datasets. Data available for testing. Other data maybe from clinical evaluators?
Evaluation Factors	**Proposed type of evaluation** : Selected user evaluation **Preconditions**: None **Evaluation factors/parameters**: To be defined

Preliminary Planning	December 2014 Interviews will be set up to complete the evaluation procedure. A virtual machine that has the tool running and any clinical data loaded will be presented to the interviewee. A questionnaire will be carefully designed to assess the usefulness of the tool. During the interview the user (bioinformatician or clinician) will be introduced to the functionalities of the tool and as the user is using the tool, the answers to the questions will be recorded. After the interview the virtual machine will be destroyed along with any sensitive data.

3.10 Prediction of SAEs/SUSARs

The objectives of the prediction of SAEs/SUSARs scenario are two:

1. A researcher, who wants to construct a prediction model, wants to do a literature search first to find out which features might be interesting to include in the prediction modelling. A researcher who has constructed a prediction model wants to validate the model using comparison to literature

2. A researcher enters data of a new patient to obtain the probability of developing an SAE for this particular patient.

This scenario needs data of a large number of patients, containing patient characteristics such as gender and age, clinical variables from the EHR and clinical trial management systems and pre-processed information from lab data such as imaging and molecular data. This data should contain an indication of whether or not the patient has experienced the SAE of interested, accompanied by a date relative to the start of the treatment. For an individual prediction, only the patient features which are used in the prediction model are required. These may consist of the same patient characteristics, clinical variables and lab data as indicated above.
The requirements on the EURECA architecture are:

- The tool should be able to obtain patient data from the EURECA data environment (e.g. via querying).
- Availability of an auto-completion tool for the selection of features is desirable.
- Availability of a literature mining service and access to literature.

The following table summarizes the status of the deployment for the prediction of SAEs/SUSARs scenario.

Table 10: Prediction of SAEs/SUSARs scenario

Responsible Partner for Evaluation	UdS
Target Group	Physicians
Leading Developer Partner	Philips
Technical Collaborators	-

Short Description of the scenario	Enabling clinicians to be more involved in the process of building prediction models for Serious Adverse Events by providing a domain specific tool for generating these type of models within the EURECA framework.
Components Involved	**Component Name**: Query execution service **Short Description:** Allows data connection via CDM **Status:** already available **Component Name**: SAE prediction client **Short Description: T**he GUI offering the functionality of the SAE prediction service. **Status:** already available **Component Name**: SAE prediction service **Short Description:** The feature selection & preparation component offers functionality to select features to include in the analysis and to define how the data should be pre-processed. **Status:** already available
Components Integration	**Short description:** The remaining • Integration with security/authentication finished. Integration with Semantic Interoperability Layer ongoing. **Date**: asap **Effort**:
Deployment environment	**Location:** • Using anonymized data, so N.A. **Restrictions**: -
Environment Setup	Client can be installed on user's pc. Client deals with secure connection to CDM
Data Needed	SIOP, MAASTRO
Evaluation Factors	**Proposed type of evaluation**: • proof of concept, subjective evaluation with selected users from UdS and possibly MAASTRO **Preconditions**: demonstrator of UI available, data available **Evaluation factors/parameters**: • functionality (The set of functions covers all the specified tasks and user objectives) • usability • portability
Preliminary Planning	User evaluation December/January Validation of results: 2015

3.11 Patient Diary & Long-term follow-up

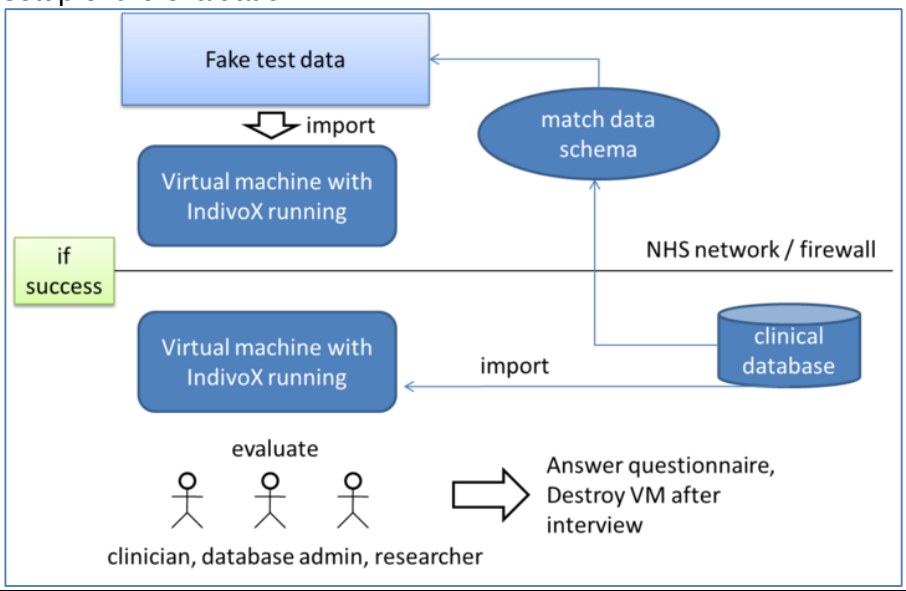
This scenario deals with the possibilities of a patient diary. Such a diary can be used in clinical trials, where there are specific eCRFs for patients. In this eCRFs the trial chairman can define what the patient can be asked. This can include the following items:

- Follow-up of late effects

- Quality of life data
- New surgical data
- Relapse data
- Second malignancy data
- Laboratory values
- Imaging data

Evaluation will be held inside the NHS UK. FORTH will provide a fully functional Virtual Machine with the PHR and the EURECA extensions. Real patient data will be imported to the system (bulk import). Clinicians from NHS will use & evaluate the tool.
Actions needed for the realization of the evaluation.

- Data mapping: FORTH will to provide the schema with the PHR clinical features needed. Then an import mechanism will be created. Oxford will create queries for the NHS EHR system to retrieve the data.
- Installing VM: Minor changes are needed. Import mechanism to be added.

Responsible Partner for Evaluation	UOXF
Target Group	Patients, Clinicians
Leading Developer Partner	FORTH
Technical Collaborators	UPM, Custodix
Short Description of the scenario	The evaluation of the Patient Diary and long term follow up will be conducted by clinicians. A dedicated virtual machine will be created and installed in the NHS UK. Then using a bulk mechanism (using the EURECA CDM) will import patient data. The bulk import functionality will be tested in advanced with fake test data. Clinicians will use the web application and evaluate the system. The following figure highlights the flow of operations for the preparation and the setup of the evaluation. 
Components Involved	**Component Name**: Personal Health Record Short Description:

	• The PHR web application and the EURECA extensions (smart alerts, auto-complete service) Status: • already available **Component Name**: Security/Authentication Short Description: • required to connect to the PHR web application Status: • already available **Component Name**: EURECA CDM Short Description: • Automatic import of data into the PHR environment using the EURECA data models. Status: • already available
Components Integration	**Short description:** • Integration with security/authentication finished. • Minor changes are needed. Import mechanism to be added. **Date**: asap **Effort**:
Deployment environment	**Location:** • NHS UK **Restrictions**: • Evaluation will be held inside the NHS UK.
Environment Setup	FORTH will provide a fully functional Virtual Machine with the PHR and the EURECA extensions
Data Needed	Real patient data will be imported to the system (bulk import). Data and the virtual machine will be destroyed after the evaluation.
Evaluation Factors	**Proposed type of evaluation**: • Clinical evaluation **Preconditions**: • The clinical evaluation will be done in the NHS firewall. All the components must be installed in a single virtual machine and run inside the firewall. **Evaluation factors/parameters**: • functionality • usability • portability • Survey PHR system evaluation criteria/questionnaire • Design interview with clinicians who evaluate the PHR system • Answer the question: ○ Would you recommend this tool to your patients?
Preliminary Planning	User evaluation early 2015

3.12 Automatic detection and reporting of SAEs/SUSARs

This scenario describes how SAEs and SUSARs can be detected and predicted before a treatment is given to a patient.

The following steps are needed for the automatic detection of SAEs/SUSARs:

- Select a patient and relevant clinical data from the HIS/EHR/PHR
- Select a drug or a treatment that will be given to a patient
- Do data mining in databases of EMA for SAEs and literature mining
- Show possible SAEs and list the molecular pharmacogenomics
- Perform molecular analysis of the pharmacogenomics in the blood of the patient
- Specify the individual risk of an SAE or SUSAR for the drug / treatment tested in this scenario

The following steps are needed for the automatic reporting of SAEs/SUSARs:

- SAE and SUSAR Definitions are described and used
- At regular time points that can be fixed (e.g. daily) the HIS/EHR/PHR databases are queried for SAEs and SUSARs.
- The SAEs and SUSARs are send to a physician
- He needs to validate the SAEs and SUSARs
- After validation an automatic report is created according to GCP criteria
- The SAEs and SUSARS are uploaded to the SAE database at the European Medical Agency (EMA)

Automatic detection and reporting of SAEs/SUSARs scenario will use the same environment with the Microbiology SAE scenario (the following scenario) and so one can simply reuse the mentioned description.

3.13 Microbiology SAE

The Microbiology SAE scenario helps a clinician to get fast knowledge and analyses about antibiotic treatments, specific infectious agents, their resistance profile, and possible serious side effects.
The scenario aims to identify infectious complications and potential life threatening SAEs. Therefore we need an early knowledge about infectious agents and their resistance profile, which will help the wards to choose pre-emptively the correct antibiotic treatment for a patient. Therefore we will implement a service, which will integrate data of different sources inside a hospital. Sufficient statistical analyses will be possible through a query interface based on the complete data set.
The following data sources from the UdS will be summarized in the EURECA data warehouse:

- EHR data: Communication server(synchronised from HIS)
- Laboratory data: M\Lab

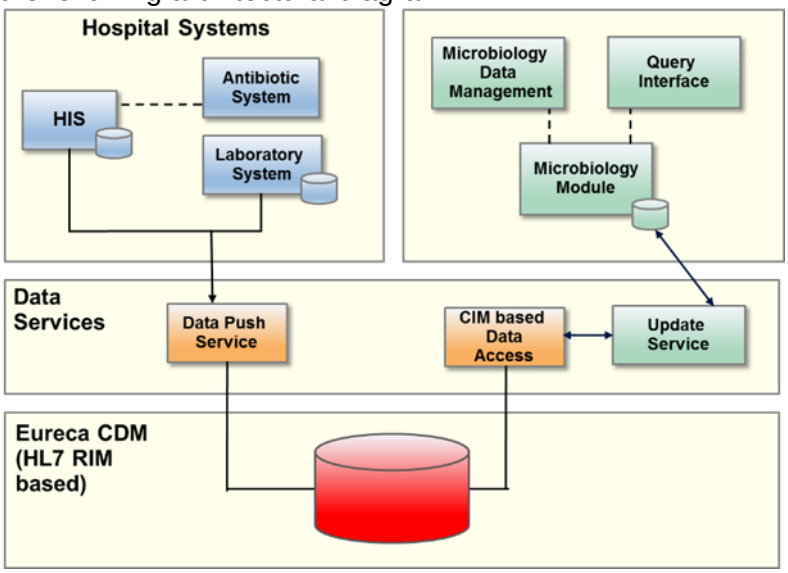
The corresponding query interface will be carried out in the Clinical Trial management system: ObTiMA.

The following table summarizes the status of the deployment for the Microbiology SAE scenario.

Table 11: Microbiology SAE scenario

Responsible	UdS

Partner for Evaluation	
Target Group	Patients with nosocomial resistant or multi resistant infections
Leading Developer Partner	Fraunhofer IBMT
Technical Collaborators	UdS (including internal collaboration with the data (management) centre of the medical centre)
Short Description of the scenario	The Microbiology SAE scenario helps a clinician to get fast knowledge and analyses about antibiotic treatments, specific infectious agents, their resistance profile, and possible serious side effects.
Components Involved	The requirements of the Microbiology SAE scenario are realized through the Microbiology Safety Service (MSS). The following systems are required in order to execute this clinical scenario. -Hospital information System: SAP -Laboratory System: M/Lab -Antibiotic Systems: SAP Module However, a direct usage of the data on these Hospital Information Systems is not possible because of several hospital internal restrictions. For that reason, predefined data will be daily exported from the SAP system and the M/Lab system to a specific **Communication Server**. **Component Name:** Communication Server **Short Description:** The Communication Server forwards the data to the EURECA Data Push Service. **Status:** Not available, will be available on the end of May 2015 **Component Name:** EURECA CDM **Short Description:** The EURECA CDM that will store the required information **Status:** Already available **Component Name:** Microbiology Module **Short Description:** The Microbiology Module enables user interaction via a Graphical User Interface (GUI). The Microbiology Module is integrated in the Web Application ObTiMA. **Status:** Already available https://obtima.org/demo/review/ **Component Name:** Data Push Service **Short Description:** The Data Push-Service pushes the clinical data from the Communication Server to the EURECA CDW **Status:** Already available

	Component Name: Update Services **Short Description:** The Update services facilitate upload of the microbiology data from the EURECA CDW to the Microbiology Module over CIM-based Data Access. **Status:** No real connection will be implemented with the EURECA CDW, but a dummy update service will be implemented
Components Integration	**Short description:** The integration of the components will lead to the following architectural diagram.  **Date**: May 2015 **Effort**: If all components are available the integration effort is minimal.
Deployment environment	The department for Pediatric Oncology and Hematology at the Saarland University (UdS) will evaluate the MSS. The tools will run for a time frame of 4 weeks. **Location:** Data (management) centre of the Saarland University Medical Center **Restrictions:** Fulfilment of all of the legal and technical requirements towards data privacy and protection regarding patient data
Environment Setup	Local installation of all the workflow-related infrastructure and tools within the data centre's premises: For this a virtual machine running SLES 11 SP3 on 8 GB RAM and 80GB disk space was installed. The virtual machine has direct, secure access to the HIS and attached systems (i.e. SAP and M/Lab). Connection to the server is firewalled and only possible from within the medical centre - it is also restricted to browser-access for common users. Also, on the virtual machine, the necessary infrastructure applications to run UdS' own tools, i.e. ObTiMA and the connectors to SAP and M/LAB as well as the EURECA DWH are setup, i.e. Java 8u20, PostgreSQL 9.2.9, Tomcat 8.0.12.
Data Needed	The tool will run using data from the SIOP trial within the Obtima.

Evaluation Factors	**Proposed type of evaluation:** Clinical evaluation by physicians treating patients fulfilling the given treatment profile **Preconditions**: The workflow needs to be fully established which means that all necessary data must be retrievable from SAP and M/LAB. This data must also be processable in the sense that the data has to be fully mapped to the respective concepts from the CD. **Evaluation factors/parameters:** Since this is an end-user scenario, the factors are focused on the correctness and completeness of the retrieved results as well as on the usability and ease-of-use.
Preliminary Planning	Implementation start January 2015 Components ready May 2015 Technical Evaluation June 2015 No clinical evaluation will be performed

3.14 Reporting episodes of febrile neutropenia

The scenario aims to detect an episode of febrile neutropenia in patients and to determine whether or not this episode is chemotherapy induced by automatically analyzing patients' EHR data using adapted NLP tools. The goal is to detect an episode of febrile neutropenia (chemotherapy treatment side effect) by extracting the following relevant information. Indicative problems to be solved include:

- Fever
- Abnormally low number of neutrophils
- Determine whether the patient received prophylactic antibiotics and/or antifungals
- Determine the cytotoxic drug's name and administration schema (i.e. how long ago, dosage ...)
- Determine whether the neutropenia is chemotherapy-induced
- Determine whether the patient received antibiotics and/or antifungals to control the infection
- Date of admission / Date of fever onset
- Prior episodes of febrile neutropenia
- Outcome of episode of ferile neutropenia
- Clinical and biological documentation of the infection

Table 12: Reporting episodes of febrile neutropenia scenario

Responsible Partner for Evaluation	Institut Jules Bordet (IJB)
Target Group	Epidemiologists, statisticians and data managers (e.g. the Statistics & Epidemiology Unit of IJB).
Leading Developer Partner	Institut Jules Bordet (IJB)
Technical Collaborators	Xerox
Short Description of the scenario	The scenario aims to detect an episode of febrile neutropenia in patients and to determine whether or not this episode is chemotherapy induced by automatically

	analysing patients' EHR data using adapted NLP tools.
Components Involved	**Component name:** Febrile Neutropenia Detection Service **Short Description:** The main driver component of the febrile neutropenia episodes detection tools is the febrile neutropenia detection service. It offers the functionality needed to complete the different steps in detection flow. For this, it integrates and connects the other services that are needed in the detection. This service adapts the NLP medical concept identifier which identifies medical concepts out of freetext data, and the NLP relation extractor which extracts relation between concepts (RDF triples) out of freetext data. **Status:** End of June 2014 (First prototype) **Component name:** Patient Identity Management Service **Short Description:** Patients are managed in the EURECA platform in the patient identity management service. It is responsible for the registration, consultation and editing of patient meta-data relevant for the EURECA services (real patient data is part of the common data warehouse). A patient registered in this service has a link with the common data warehouse by using a common patient ID. **Status:** Available (Custodix Service) **Component name:** Free Text Service **Short Description:** This service is responsible for free text querying on the indexing database linked to EURECA CDM. It offers free text searching functionality in order to query unstructured data that have been processed using NLP to be loaded in the CDM. **Status:** Under construction (UPM Service) **Component name:** Query Engine **Short Description:** Criteria for episodes of febrile neutropenia are matched with the information of a selected patient in the febrile neutropenia detection service. It provides an interface that enables other EURECA services to send matching requests. The criteria matcher will query the requested information of the patient by sending the query that is included in the criterion to the query execution service of the different available data warehouses. **Status:** Available (Custodix Service) **Component name:** Query Execution Service **Short Description:** This service provides functionality to query the datasets of the CDW and other data warehouses available on the site

	through the semantic layer. It abstracts the underlying data sources for the upper EURECA services and presents data to applications according to a single integrated data model. More information about this service can be found in the semantic layer view section. **Status:** Available (UPM Service) **Component name:** Query Builder Service **Short Description:** This service is responsible of generate query template in XML format for querying on the Query Engine Service. **Status:** Available (UPM Service)
Components Integration	**Short description:** The integration of the different components is done in an iterative way. The main integration will be done for the release of the first prototype, and will be improved until July 2014 for the last version of the tool. **Date**: End of June 2014 (First prototype) until July 2014 (Last version of the tool). **Effort**: Integration effort in iterative steps.

Deployment environment	**Location:**
	• IJB Tool:
	  ○ Febrile Neutropenia Detection Service
	• Custodix VMs
	  ○ Patient Identity Management Service
	  ○ Query Builder
	• UPM VMs
	  ○ Query Execution Service
	  ○ Query Builder Service
	  ○ Free Text Service
	• Custodix Security VMs
	  ○ Patient Identity Management Service
	
	Restrictions:
	The services working with patient data will be installed at the hospital site (IJB). It will be tested by one epidemiologist, statisticians and data managers.

Environment Setup	Febrile Neutropenia Detection Tool: Java environment Query Builder, Patient Identity Management Service: Java environment, Tomcat application server, MySQL Query Execution Service, Query Builder Service: Java environment, Tomcat application server, MySQL
Data Needed	IJB Breast Structured dataset IJB Breast free text dataset
Evaluation Factors	**Proposed type of evaluation**: Evaluation will be made on comparing results from NLP tools with manual reporting cases of episodes of febrile neutropenia. **Preconditions**: Manually report episodes of febrile neutropenia among patients in the dataset. **Evaluation factors/parameters**: Comparison manual/automatic extraction Questionnaire results Timing tables
Preliminary Planning	**First prototype:** End of June 2014 **Final version:** End of July 2014 Each version will be iteratively tested onsite (IJB) during the development process.

3.15 Cancer registry reporting

The scenario aims to fill part of the internal cancer registry by automatically extracting information related to the categorization of any new tumor case (e.g. morphology, topography, tumor staging) out of patients' EHR data using adapted NLP tools.

Table 13: Cancer registry reporting scenario

Responsible Partner for Evaluation	Institut Jules Bordet (IJB)
Target Group	Help data managers to fill in the cancer registry ((e.g. the Statistics & Epidemiology Unit of IJB).
Leading Developer Partner	Institut Jules Bordet (IJB)
Technical	Xerox

Collaborators	
Short Description of the scenario	The scenario aims to fill part of the internal cancer registry by automatically extracting information related to the categorisation of any new tumour case (e.g. morphology, topography, tumour staging) out of patients' EHR data using adapted NLP tools.
Components Involved	**Component name:** Cancer Registry Reporting Service **Short Description:** The main driver component of the cancer registry reporting tools is the cancer registry reporting service. It offers the functionality needed to complete the different steps in detection flow. For this, it integrates and connects the other services that are needed in the screening. This service adapts the NLP medical concept identifier which identifies medical concepts out of freetext data, and the NLP relation extractor which extracts relation between concepts (RDF triples) out of freetext data. **Status:** End of December 2014 (First prototype) **Component name:** Cancer Registry Management Service **Short Description:** Cancer registry is locally managed on site. It interacts with the cancer registry reporting service and populates the cancer registry DB. **Status:** Available (Local IJB Service) **Component name:** Patient Identity Management Service **Short Description:** Patients selected for screening are managed in the EURECA platform in the patient identity management service. It is responsible for the registration, consultation and editing of patient meta-data relevant for the EURECA patient screening (real patient data is part of the CDW). A patient registered in this service has a link with the Common data warehouse by using a common patient ID. **Status:** Available (Custodix Service) **Component name:** Free Text Service **Short Description:** This service provides functionality to query the semantic core datasets of the CDW and other data warehouses available on the site through the semantic layer. **Status:** Under construction (UPM Service) **Component name:** Query Engine **Short Description:** The criteria matcher exposes functionality to create patient cohorts. Items of the cancer registry to extract are matched with the information of patients in the query service. It provides an interface that enables other EURECA services

	to send matching requests. The patient cohort that matches the criteria is sent back to the requesting service. **Status:** Available (Custodix Service) **Component name:** Query Execution Service **Short Description:** This service provides functionality to query the datasets of the CDW and other data warehouses available on the site through the semantic layer. It abstracts the underlying data sources for the upper EURECA services and presents data to applications according to a single integrated data model. More information about this service can be found in the semantic layer view section. **Status:** Available (UPM Service) **Component name:** Query Builder Service **Short Description:** This service is responsible of generate query template in XML format for querying on the Query Engine Service. **Status:** Available (UPM Service)
Components Integration	**Short description:** The integration of the different components is done in an iterative way. The main integration will be done for the release of the first prototype, and will be improved until March 2014 for the last version of the tool. **Date**: End of December 2014 (First prototype) until March 2014 (Last version of the tool). **Effort**: Integration effort in iterative steps.

Deployment environment	**Location:**
	• IJB Tool:
	  ○ Cancer Registry Reporting Service
	  ○ Cancer Registry Management Service
	• Custodix VMs
	  ○ Patient Identity Management Service
	  ○ Query Builder
	• UPM VMs
	  ○ Query Execution Service
	  ○ Query Builder Service
	  ○ Free Text Service

	• Custodix Security VMs ○ Patient Identity Management Service **Restrictions**: The services working with patient data will be installed at the hospital site (IJB). It will be tested mainly by data managers who are used to perform this task manually. And it will also be tested by one epidemiologist and one statistician.
Environment Setup	Cancer Registry Reporting Tool: Java environment Cancer Registry Management Service: Windev environment Query Builder, Patient Identity Management Service: Java environment, Tomcat application server, MySQL Query Execution Service, Query Builder Service: Java environment, Tomcat application server, MySQL
Data Needed	- IJB Breast Structured dataset - IJB Breast free text dataset - IJB Cancer Registry dataset
Evaluation Factors	**Proposed type of evaluation**: Evaluation will be made on comparing results from NLP tools with manual entries in the cancer registry. **Preconditions**: Technical partners to have access to the IJB Cancer Registry dataset, and statistics. **Evaluation factors/parameters**: Retrospective comparison with entries in the local cancer registry Questionnaire results Timing tables
Preliminary Planning	**First prototype:** End of December 2014 **Final version:** End of March 2014 Each version is iteratively tested onsite (IJB) during the development process.

4 CONCLUSIONS

One of the objectives in EURECA work package 9 is to prepare the technical and procedural infrastructure – in compliance with the defined legal and security framework of the project – for the installation of EURECA technologies and tools for their extensive evaluation and validation. This deliverable reports the current status and the preparation of the deployment environment for the clinical pilots. EURECA supports 15 clinical scenarios.

For each EURECA clinical scenario the technical leader for the development environment reported the current status. To be able to deploy each scenario all the needed EURECA components should be implemented and integrated as well as the needed datasets must be available and ready to use. So the status on the aforementioned components is reported and also the plans for their integration. Finally the exact environment for the deployment and the corresponding setup has been defined and commented.

For the most of the clinical scenarios the preparation of the deployment environment has already started and follows a solid plan in terms of time and integration. In some cases however, some deficiencies where revealed which however have low risk for the project. These deficiencies exist due to the involved EURECA components which are currently under development and will be delivered the latest in month 36 of the project (January 2015).

5 REFERENCES

[1] EURECA consortium, D1.2 Definition of relevant user scenarios based on input from users, 2013
[2] EURECA consortium, D1.1 User needs and specifications for the EURECA environment and software services, 2012
[3] EURECA consortium, D9.1 Report on the development environment and on the available test data, 2012
[4] EURECA consortium, D2.5 EURECA architecture and interface layer update, 2014

APPENDIX – Deployment Reporting Form

Responsible Partner for Evaluation	
Target Group	
Leading Developer Partner	
Technical Collaborators	
Short Description of the scenario	
Components Involved	**Component Name**: Short Description: (functionality, data needed) Status: (already available/will be available on) **Component Name**: Short Description: (functionality, data needed) Status: (already available/will be available on)
Components Integration	**Short description:** **Date**: **Effort**:
Deployment environment	**Location:** (physical location) **Restrictions**: if any (legal ethical)
Environment Setup	Detailed technical description of the setup needed for the deployment environment of the specific clinical scenario.
Data Needed	(Please select data among APDG, GBG, SIOP, FRA, Oxford, IJB (Fake). You can find information about available data you can find here: http://atlas.ics.forth.gr/EURECA/wiki/index.php/WP4#Datasets_.26_Query_execution_services_deployed And about the release status: http://atlas.ics.forth.gr/EURECA/wiki/index.php/Data_Status#Datasets_deployed_and_available_through_the_Semantic_Interoperability_Layer In case you need more/other data please be specific (where, when, how)
Evaluation Factors	**Proposed type of evaluation** : (clinical evaluation, proof of concept, (e.g. workshop, subjective evaluation with selected users etc.), parallel, sequential or retrospective) **Preconditions**: **Evaluation factors/parameters**: (see http://atlas.ics.forth.gr/EURECA/wiki/index.php/Evaluation_procedures)
Preliminary Planning	Dates etc.