# ICT-2011-288048

# EURECA

# Enabling information re-Use by linking clinical Research and CAre

IP
Contract Nr: 288048

# Deliverable:
# D11.1 Public summary of the project

Due date of deliverable: (29-02-2012)

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

Responsible WP11: Philips

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

| **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 |
|---|---|---|
| Anca Bucur | Philips | anca.bucur@philips.com |
| | | |
| | | |
| | | |

## 0.2 Documents history

| Document version # | Date | Change |
|---|---|---|
| V0.1 | 31.03.2012 | Starting version, template |
| V0.2 | | Definition of ToC |
| V0.3 | | First complete draft |
| V0.4 | | Integrated version (send to WP members) |
| V0.5 | | Updated version (send PCP) |
| V0.6 | 25.05.2012 | Updated version (send to project internal reviewers) |
| Sign off | | Signed off version (for approval to PMT members) |
| V1.0 | 03.07.2012 | Approved Version to be submitted to EU |
| | | |

## 0.3 Document data

| | |
|---|---|
| **Keywords** | Semantic interoperability, clinical research, secondary use, EHR, oncology |
| **Editor Address data** | Name: Anca Bucur<br>Partner: Philips<br>Address: HTC34 5-03<br>Phone: +31 63 1637570<br>Fax:<br>E-mail: anca.bucur@philips.com |
| **Delivery date** | 09.07.2012 |

## 0.4 Distribution list

| Date | Issue | E-mailer |
|---|---|---|
| April 2012 | | al_eureca_all@natlab.research.philips.com |
| July 2012 | | Benoit.ABELOOS@ec.europa.eu |
| | | |

# Table of Contents

# 1 Project INFO

| General Info | |
|---|---|
| Acronym | EURECA |
| Name | Enabling information re-Use by linking clinical Research and Care |
| Web page | www.eurecaproject.eu/ |
| Reference | FP7-ICT-2011-288048 |
| Contract type | IP |
| Start date | February 01, 2012 |
| End date | July 31, 2015 |
| Duration | 42 months |
| Coordinator | Philips |

| Participants | | | |
|---|---|---|---|
| No. | Name | Short name | Country |
| 1 | PHILIPS ELECTRONICS NEDERLAND B.V. | Philips | NL |
| 2 | FOUNDATION FOR RESEARCH AND TECHNOLOGY HELLAS | FORTH | GR |
| 3 | INSTITUT JULES BORDET | IJB | BE |
| 4 | CUSTODIX N.V. | CUSTODIX | BE |
| 5 | UNIVERSITAET DES SAARLANDES | UdS | DE |
| 6 | THE CHANCELLOR, MASTERS AND SCHOLARS OF THE UNIVERSITY OF OXFORD | UOXF | UK |
| 7 | FRAUNHOFER-GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V | FhG | DE |
| 8 | VERENIGING VOOR CHRISTELIJK HOGER ONDERWIJS WETENSCHAPPELIJK ONDERZOEK EN PATIENTENZORG | VUA | NL |
| 9 | BREAST INTERNATIONAL GROUP - AISBL | BIG | BE |
| 10 | GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER | LUH | DE |
| 11 | XEROX SAS | XEROX | FR |
| 12 | UNIVERSIDAD POLITECNICA DE MADRID | UPM | SP |
| 13 | STICHTING MAASTRICHT RADIATION ONCOLOGY MAASTRO CLINIC | MAASTRO | NL |
| 14 | ecancermedicalscience AG | eCancer | CH |
| 15 | EUROPEAN INSTITUTE FOR HEALTH RECORDS | EUROREC | FR |
| 16 | STONEROOS B.V. | SIT | NL |
| 17 | GBG FORSCHUNGS GMBH | GBG | DE |
| 18 | NATIONAL RESEARCH COUNCIL CANADA | NRC | CA |

# 2 RATIONALE

From academic medical research centres to community hospitals and other stakeholders, the healthcare industry continues to improve its capabilities for electronic data capture. Despite these advances, a gap remains in the ability of these IT systems to deliver knowledge and insight back to the very researchers and clinicians they are intended to support.

Additionally, there is a widening knowledge gap between the care provided in top research clinical sites and standard care sites, resulting in large differences in treatments and outcomes. In this context, the need to bring the latest therapy options validated in clinical research to each and every hospital must be addressed before being able to significantly reduce the numbers of patients that receive suboptimal treatment (e.g. overtreatment, wrong dose, etc.), or the wrong treatment. There are currently very few mechanisms and formally established channels for transferring the best practices to clinicians and the current dissemination means are insufficient.

**Challenges for Patient Safety**

The fact is that we will never know for sure whether a drug is entirely safe when it gets to the market. A main problem persists: access to the samples and to data that could help us to study rare adverse events:

- Usually, the size of safety databases that are generated during drug development are too small to do a good job in picking up rare safety events
- The largest safety information is produced when a drug is on the market

**Our current tools do not allow us to effectively capture this information and capitalize on its potential.**

"*The Transition from Pre-Clinical to Clinical Application of Safety Related Genomics"*
*Felix W. Frueh, CDER/FDA*
*IOM/FDA Emerging Safety Science Workshop, 2007.*

The current separation between clinical research and clinical care makes the detection of many serious patient safety issues difficult. Serious side effects[1],[2] of therapy and drugs that appear outside a clinical trial either due to a low incidence or a late onset are very difficult to detect and to explain in the absence of a feedback loop from standard care to research.

Despite identified benefits[3], the secondary use of care data for research, quality assurance and patient safety is rarely supported. Main barriers to enabling secondary use of data are the lack of interoperability, common standards and terminologies[4].

The recruitment rate of patients into clinical trials is currently low despite recognized better outcomes and the enrolment process is slow and inefficient, involving redundant data entry, inconsistencies and several manual verification steps. Additionally, the progress in treating rare diseases is currently hampered by the fact that there is not enough data to support research. Even when sufficient data was available, the acceleration of discoveries in basic sciences would not necessarily change at the same speed the way patients are treated. As discoveries can only be implemented after proper prospective validation, supporting a faster enrolment of patients in clinical trials would also have a direct impact on the speed of improvement of the standard of care.

[1] Breast Cancer Surveillance Practices Among Women Previously Treated With Chest Radiation, Oeffinger et al. *JAMA.*2009; 301: 404-414.
[2] Breast Cancer after Childhood cancer: A Report from the Childhood Cancer Survivor Study, http://www.annals.org/cgi/reprint/141/8/590.pdf.
[3] Electronic medical records for clinical research: application to the identification of heart failure, Pakhomov et al., Am J Managed Care, 2007
[4] Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance, Hersh W., Am J Managed Care, 2007

# 3 EURECA APPROACH

The goal of the EURECA project is to enable seamless, secure, scalable and consistent linkage of healthcare information residing in EHR systems with information in clinical research information systems, such as clinical trial systems, supporting the two currently separated worlds of clinical research and clinical practice to connect and benefit from each other.

Main barriers of secondary use of EHR data for research and of enabling a consistent feedback loop to care are the lack of common technology standards and concept terminologies. While solving the interoperability issue in healthcare at generic level is not a realistic approach[5], EURECA aims at semantic interoperability on domains of concepts (i.e. describing specific clinical areas). We start from disease- and treatment-related sets of concepts in the oncology domain and demonstrate our solution in concrete clinical scenarios. On top of the achieved semantic interoperability we build software services and tools to support more efficient research, better care and improved patient safety.

The approach taken in EURECA is to rely when possible on existing initiatives and previous efforts in terminology development and standardization. We will demonstrate the viability of the solutions developed by implementing a set of loosely-coupled interconnected services that we will deploy in the context of several pilot demonstrators in the cancer area, at the sites of the healthcare organizations participating in EURECA.

The EURECA environment aims to provide several software services that help to securely interconnect the clinical trial systems (CTs) and the electronic health record systems (EHRs). This will bring several benefits among which early detection of patient safety issues and more efficient recruitment of eligible patients. Consistent linkage between CTs and EHRs will also help to significantly reduce the need for double data entry, which is currently often common practice.

Figure 1 depicts the envisioned services that will help to interconnect the two currently mostly disconnected worlds: the CTs and EHRs. Using the SemanticHEALTH classification of semantic interoperability (SIOp)5 we can observe that the current level of SIOp between CTS and EHRs is somewhere between level 0 (no interoperability at all) and level 1 (syntactic interoperability). In order to achieve the above objectives we will have to increase the SIOp level to at least 2b (bidirectional semantic interoperability of meaningful fragments) or 3 (full semantic interoperability, sharable context) on specific clinical areas.

The essential steps for achieving this SIOp improvement include the definition of sound information models describing the clinical trial systems, building on existing research results when possible[6]. Electronic health records too need to be properly modelled; to that end we will adopt the appropriate state-of-the-art representation formalisms such as HL7 CDA, the openEHR Reference Model, ISO/EN 13606, etc.

[5] Semantic Interoperability for Better Health and Safer Healthcare, SemanticHEALTH Report, pp 12-13, 2009

[6] G.Weiler et al., Ontology Based Data Management Systems for post-genomic clinical Trials within an European Grid Infrastructure for Cancer Research, Proc of the 29th Annual Int. Conf. of the IEEE EMBS, 2007

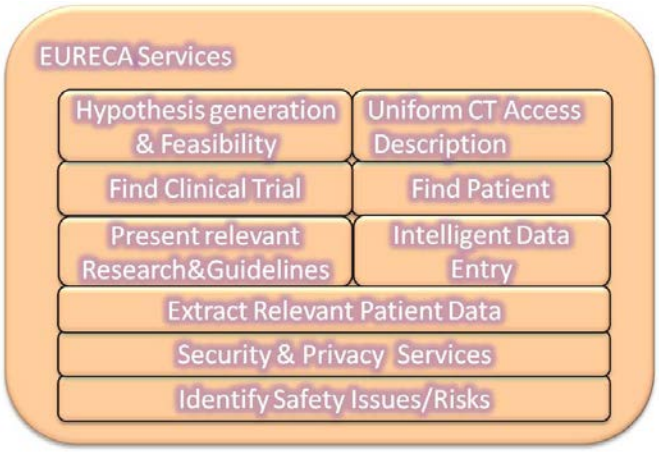


**Figure 1** An overview of the EURECA software services.

The semantics of the clinical terms will be captured by standard terminology systems such as SNOMED CT, ICD, LOINC. The scalability of the solution needs to be achieved by modularization, e.g. instead of aiming at inclusion of the complete SNOMED terminology (more than 300 thousand concepts) we will identify a core subset that covers the chosen clinical domain. Such core data set shall be validated both by clinical and knowledge engineering experts to assure proper coverage and soundness. Utilizing the core data set we will devise a mapping system (Figure 2) between the information models of CTS and those of EHRs. These mappings will too be verified by our clinical experts and validated in properly chosen scenarios/use cases focusing on those with a high potential for patient safety improvement and more efficient and effective research.

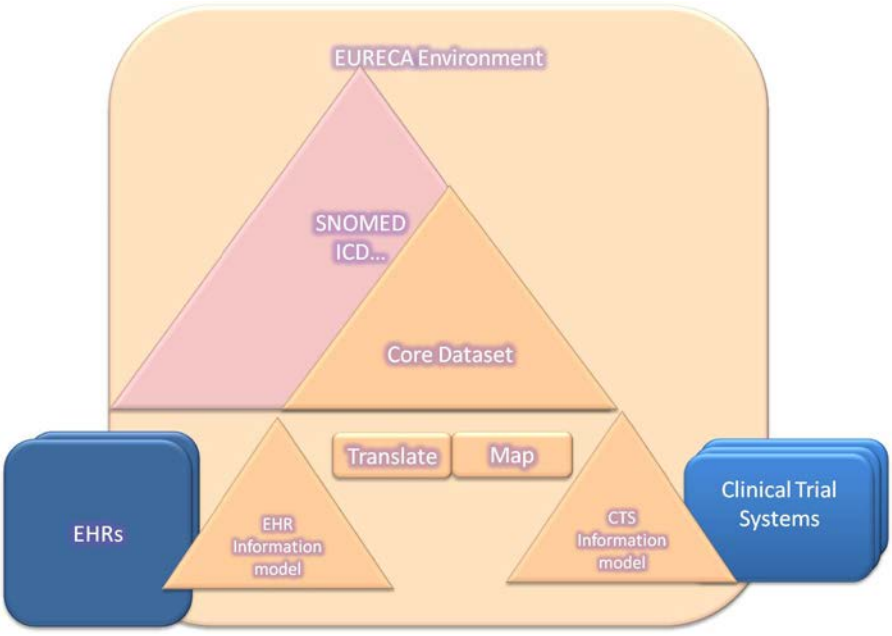


**Figure 2** EURECA mapping components for achieving semantic interoperability

For the scenarios focused on early detection of safety risks and on long term follow-up of patients beyond the end of a clinical trial we will provide uniform access to Patient

Health Record information, enabling the use of patient-generated information to improve the safety and quality of care.

The EURECA infrastructure will enable new ways of conducting meta-analysis over larger population samples, including long term patient follow-up as well as the results of previous clinical trials including those with negative results in order to be able to avoid known “dead-ends”.

The need and usefulness of tools that help to push forward the knowledge of risks associated with unique patient characteristics and treatment patterns has been well recognized[7]. EURECA will enable data mining for patient safety over the combined data from electronic medical health records and clinical research databases, thus also allowing extending the scope of data mining to a currently untapped resource which is the EHR system.

[7] V.N. Stroetmann et al., Impact of ICT on Patient Safety and Risk Management, eHealth for Safety Report, 2007

# 4 OBJECTIVES

### Enabling semantic interoperability among EHR and clinical trial systems

Achieving semantic interoperability among EHR and clinical trial systems is at the core of the EURECA project, as it is the basis for enabling many of the software services and tools developed in the project.

> … regardless of the type of vision one may develop, semantic interoperability is not a phenomenon to be expected overnight.[..] The SemanticHEALTH is characterized by a large number of changes at both the technical and the use case level. Note however, that even in this vision, no full semantic interoperability or a complete harmonization of either EHR models or terminologies can be expected.
>
> *"Semantic Interoperability for Better Health and Safer Healthcare"*
> *SemanticHEALTH Report, 2009*

From the technology viewpoint, in order to provide an efficient, robust and semantically interoperable solution, one needs to move from plain keyword matching to a combined approach where keywords are mapped to higher level concepts with clearly defined semantics. Such concepts are usually organized in concept hierarchies and include domain specific attributes and relations. Reasoning at this level, rather than at keyword level, is expected to enable us to move from error-prone lexical matching to more robust semantic-aware solutions.

The linking of clinical care information in the EHR systems with research data in the clinical trial systems will be achieved in EURECA by exposing the EHR information in a canonical clinical care model leveraging on existing standards and vocabularies, exposing the clinical trial information in a canonical clinical trial model, again leveraging on existing standards and vocabularies, and reaching semantic interoperability through the definition of a model and a core set of concepts that provide a mapping between the clinical care (EHR) model and the clinical trial model.

As some of the relevant data in the EHR is stored in free text format, its standards-based seamless reuse relies on the accurate extraction of the important concepts and their relations. The task of building canonical information models of the EHR and CT systems becomes more complex when we also deal with unstructured documents in those systems. The discovery of the underlying information model contained in free text reports requires additional effort to identify the relevant concepts and their relations and to understand their meaning. We will achieve this gradually, with the support of a mix of NLP techniques used to answer concrete questions. Whenever possible, existing tools will be used.

For the free text documents we will carry out information extraction based on our defined core dataset, with a clear search goal and not a random search in an infinite space. We start by scanning the text with concepts from the core dataset (tagging, synonyms, annotations, etc.) and progress to relations and meaning. As data and model quality is essential, and any error in the NLP task would propagate further into the model and into the application, we do not aim at a fully automated NLP process (that would replace the expert); instead, our goal is to make the tasks of the modelling and domain experts efficient and manageable.

### Enabling secondary use of care data for research

EURECA aims to support more effective and efficient execution of clinical research by providing access – in a legally compliant and secure manner – to the large amounts of

patient data collected in the EHR systems to be used for new hypotheses building and testing (e.g. to benefit rare diseases), cohort studies, as well as protocol feasibility.

Association studies on large volumes of EHR data can also reveal serious side effects to drugs and therapies and link those with patient characteristics (generate new research hypotheses for biomarkers identification). Data mining of the EHR data can be carried out based directly on the information model of the source or based on the EURECA core set of concepts when it is assumed that the end user does not need detailed knowledge of the source. We will demonstrate this secondary use in concrete scenarios and provide uniform access interfaces to the data and tools/services developed to support the execution of the scenarios. Our services will also provide uniform access to external repositories of data and knowledge relevant in the discovery scenarios.

Another important re-use of EHR data is to support patient recruitment and to test trial feasibility in a data-driven protocol design process. Semantic interoperability between EHR and CT systems can support faster recruitment of patients in clinical trials, but also help to improve protocols to target the right patient population, define the inclusion/exclusion criteria, identify sites with sufficient patients, etc. This can lead to more efficient enrolment, better protocols, and fewer protocol amendments (e.g. criteria that are too narrow can be detected earlier).

**Enabling efficient recruitment by matching relevant patient data with eligibility criteria from clinical trials**

Good semantic interoperability among EHR and clinical trial systems can be exploited to allow more efficient patient enrolment in clinical trials, which is a need shared by both clinical care and clinical research as it would both improve care and speed up research. The eligibility and exclusion criteria of running clinical trials described in the Clinical Report Forms can be matched based on the EURECA core set of concepts with patient data in the EHR to find eligible patients for clinical trials, and in the context of a patient case to find relevant clinical trials. Additionally, we will provide support for efficient protocol design and refinement, enabling the interactive automatic evaluation of the various criteria based on EHR data.

**Enabling long term follow up of patients beyond the end of a clinical trial, for better research and improved safety**

It is relevant for clinical research to follow the patients long after the end of the trial, to monitor survival, recurrence, and serious side effects of treatments. Next to important benefits for the patient, this can also contribute to the generation of new research hypotheses and the identification of safety risks. EURECA will provide solutions to access the relevant information while adhering to all applicable privacy and security requirements. Relevant information can be extracted from the EHR based on the information model of the source and linked through the EURECA core set of concepts to clinical research concepts. Uniform access to patient-generated PHR data will add a new dimension to this objective: Leveraging this new source of information can support better (more accurate) assessment of the risks and benefits of drugs and treatments, at different time granularity (e.g. patient may describe side effects and symptoms when they happen), in a different context (patient at home) and on a time span beyond the duration of a clinical trial.

**Improving efficiency by reducing the need for multiple data entry in the information exchange between clinical research and clinical care**

We aim to avoid double data entry in care and research by establishing a primary source for the various types of data and providing the necessary data-feed loops to the EHR and the clinical trial systems (e.g. automatically filling in clinical trial eligibility criteria in the EHR based on the clinical trial CRFs) and preserving compliance with workflows in clinical care and research. Additionally, EURECA will enable a consistent and efficient reporting of SAEs and SUSARs (adverse events in clinical trials) and address the current issues of multiple and inconsistent reporting.
EURECA tools/software services helping avoid multiple data entry are:

- Recruitment (e.g. through automatically filling in clinical trial eligibility criteria in the EHR based on the clinical trial CRFs, and automatic EDC prefilling based on EHR data)
- Safety issue detection and reporting (primary source, single reporting)

In this objective, we include evaluation of the use of PHR data as potential source for detecting safety issues.

**Exposing a uniform presentation of clinical trial information, validated clinical trial results and other relevant external knowledge and data resources**

There are currently two types of databases collecting clinical trial-related data. Clinical trial registry databases are online catalogues of hypothesis-testing clinical trials conducted on human subjects. Clinical trial results databases are online repositories containing the results of clinical trials.

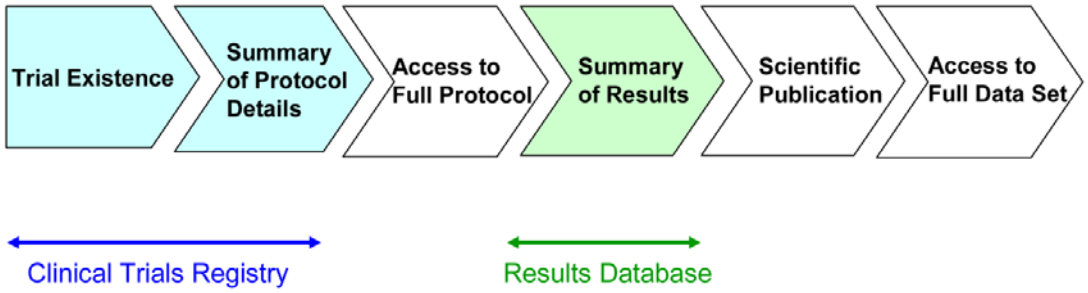


**Figure 3** Information available in the clinical trial registry and results database[8]

EURECA will provide uniform interfaces to access external repositories of clinical trial information and results. Based on existing initiatives and emerging standards such as WHO ICTRP[9] we will select a core set of information items that need to be provided to comprehensively describe a clinical trial (compliant with the set of 20 key elements specified by WHO) and its results (including the case of trials with negative results) and findings, propose an information model and build a prototype repository to instantiate this information model. Additionally, we will link to external relevant clinical trial registries and clinical trial databases to be used by services in the EURECA environment.

- **Repository of running trials and public data repositories**

  We will identify relevant existing external data repositories and provide uniform access to them. When available, EURECA will utilize the existing repository(ies) of

[8] D.A. Zarin and T. Tse, Moving Towards Transparency of Clinical Trials, Science. 2008 March 7; 319(5868): 1340–1342
[9] World Health Organization's International Clinical Trials Registry Platform, http://www.who.int/ictrp/en

running clinical trials including their focus/aim, patient inclusion criteria, and suitable location (taking into account the geographic proximity to the patients) . EURECA will also provide access to public biomedical databases, the combination of those with the EHR and clinical trial data will be of interest to broader meta-analysis and data mining for patient safety. Important for our scenarios will be the linkage to PHR systems to enable the reuse of patient-authored information.

- **Repository of validated results**

We also plan to identify those repositories which are not yet (widely) available and propose a suitable data model of the information that they should contain. This for instance includes a repository of validated results of completed clinical trials. We believe that structured access to validated results of clinical trials will bring tremendous benefits in speeding up the knowledge transfer and the rate of adoption of the latest research advances by the clinical practice.

Positive results are currently being published in journals and conference papers and the data which was used during the research is required to be made public so that the repeatability of the findings can be verified and new discoveries are made possible from the existing data. Despite these positive trends, the solutions that are currently used for sharing the published results and the underlying data do not allow for efficient machine processing and identification. By structuring these results (as well as data when appropriate), the information access will become easier and the impact on further research larger, e.g., by avoiding duplication in hypothesis generation, enabling efficient and effective data analysis. The research community will be able to take advantage of the uniform access to validated results of the completed trials, enabling efficient meta-analysis and accelerating the process of hypothesis generation.

If such repository also includes the negative results of clinical trials which are currently hardly ever published/shared, researchers can learn from the past by avoiding the already explored “dead-ends”, focusing their efforts and resources mainly on new promising directions. We are aware that building, but also maintaining and contributing to such repository will be a considerable effort which should be incentivized, but we also believe that this is one of the clear examples when a moderate investment in improving the sub-optimal status-quo will deliver considerable benefits and help to achieve in the near future more by better utilization of the research resources and available data.

- **Computerized Guidelines**

Uniform access to validated results will accelerate the development and deployment of decision support systems. Proper implementation and use of clinical decision support systems is regarded as an important recommendation for reducing the frequency and consequences of errors in medical care[10]. Having access to such repository will also help to make the process of creation of clinical guidelines much more systematic and ICT tools will be much more applicable in this, currently mainly manual, process. As a result, the clinical guidelines will become computerized from the very beginning and this will be another major enabler for a new generation of decision support systems which will be able to bring these guidelines to the point of care and ensure that the standard of care will be consistently improving.

[10] D.W. Bates et al., Reducing the Frequency of Errors in Medicine Using Information Technology, Journal of the American Medical Informatics Association, 2001

- **Warehouses of clinical trial data, data models, ontologies, etc.**

EURECA aims to build on existing initiatives focusing on standardization and sharing of data, models (e.g. risk models, predictive models of response, etc.) and knowledge in the clinical domain. We plan to achieve uniform access to relevant repositories and infrastructures. Partners in the project are currently active in several such initiatives in Europe and beyond, and we intend to build bi-directional links to those initiatives in order to achieve a critical mass. Interoperability is by definition a matter of critical mass, openness and team work and cannot be achieved in isolation.

**Extracting relevant clinical information from the EHR and contextualizing it to the patient case**

A proper semantic link between EHR and clinical trial systems will enable, among others, an efficient generation and adaptation of computerized clinical guidelines such that they reflect the latest validated research results. As a result, standard clinical care can use ICT tools to quickly adopt and personalize these guidelines, mapping them to the actual patient records, identifying optimal treatment plans based on the extracted relevant information from the EHR. These technological and conceptual advancements will lay the foundations for a new generation of clinical decision support systems that are both semantically aware and up-to-date with the latest validated research results.

**Building solutions for faster transfer and dissemination of relevant research results and literature to the care**

EURECA aims to research the contextualization of validated research data and literature to a patient case, to be used in standard care. Relevant external data, such as references to articles, applicable literature, guidelines, protocols and standardized descriptions of validated trials, will be selected based on the current patient case under investigation. In this process we will use the information model of the source EHR, the EURECA core of concepts and the external sources of information to which EURECA provides uniform access, such as computerized clinical guidelines, external research public databases and repositories of clinical trial information and validated results.

People making decisions about healthcare need access to knowledge derived from the findings of clinical trial research,

*"Reporting the findings of clinical trials: a discussion paper"*
*Ghersi D et al., Bulletin of the World Health Organization, 2008.*

# 5 Expected Results

EURECA targets a specific research objective in a sharply focused approach while at the same time it includes a coherent and integrated set of activities dealing with multiple related issues and provides state-of-the-art responses to the challenges identified in the call, i.e. integration of clinical research and care, semantic interoperability, re-use of care data for large scale epidemiology and cohort studies, and more efficient (faster recruitment, avoid multiple data entry) and safer research. Special emphasis is given in making sure that the new tools, services and applications to be developed in EURECA will also be evaluated on their effectiveness and validated in use cases with a high potential for improving patient safety in research and epidemiology.

The idea for the *EURECA* project has evolved over a significant period of time, as a result of experiences and R&D results from previous National and EU projects, but also as a result of urgent needs for technologies assisting the optimization of the process for new discoveries. An indication of this need is the fact that several EURECA partners have internally begun to locally address the problems of how best to "link their clinical patient data with their clinical research data, in an attempt to optimise both the care and new discoveries processes." However, achieving semantic interoperability at a scale that has impact requires EU-wide consolidated effort and collaboration.

Several members of the consortium (BIG, GBG/Luisenkrankenhaus Düsseldorf, IJB, UOXF) are top players in the global cancer research and treatment community. They will constitute a show case of EURECA, enabling us to prove the applicability of our solutions in a real multi-centric, multi-language (English, French, German) environment, also heterogeneous with respect to legal aspects, country-specific regulations, technology and ICT infrastructures. In this complex context we will validate the architectural framework, the services, and especially the viability of the core data set-based approach in an EU-wide setting. The success of this scenario will be a strong motivator for other large clinical networks (EORTC, SIOP, etc.) to adopt the EURECA solutions at a wide scale and convince our collaborators from the pharmaceutical industry of the significant added benefits of EURECA.

We are aware that interoperability is by definition a global issue that cannot be tackled in isolation, requiring both critical mass and openness. Therefore, it is our goal to join forces with other initiatives and contribute to common solutions in Europe and beyond.

# 6 IMPACT

*The EURECA* proposal focuses its research and development agenda on several pressing needs in the healthcare and in the bio-pharmaceutical industry. Their successful resolution will have significant and far-reaching implications.

..Integrating EHRs with clinical trials has major potential to increase recruitment rate, as could be demonstrated in first studies…

*Prof. Dr. Ohmann, Kuchinke:*
*Meeting the challenges of patient recruitment.*
*A role for electronic health records.*
*Int J Pharm Med 2007; 21: 263-270*

The EURECA project also supports many of the EU policy and social objectives. The potential impact of a successful implementation of the current project is briefly presented in the sections to follow.

## 6.1 Common platform for a wide range of ICT-based healthcare services

EURECA aims to build a seamless, consistent, standards-based semantic interoperability framework linking care-specific information systems with clinical research information systems, to enable efficient and large scale reuse of care data for research. On top of the interoperability platform, we will build a set of software services for more efficient research and better care. The long term vision is to enable easy development and deployment into the environment of other ICT-based healthcare services.

EURECA envisages an open, scalable Service Oriented Architecture platform and will therefore focus on interoperability and interfacing. The main objective is to define an interoperability framework, based on a reference architecture and of sound definition of service interfaces and integration profiles, in which the tools and services created by the technical WPs can be fit and subsequently provide a reference implementation of this framework. This approach reuses existing systems and services whenever possible (speeding up initial deployment), ensures extensibility on the long run, allows for integration of (future) third party solutions. It will enable interfacing with other platforms from similar initiatives in order to rely on each other for scale and critical mass.

We will provide support for early adopters who would like to develop based on our reference architecture, join our environment, make use of our interoperability layer and services, and provide services on top of EURECA. The goal of EURECA is to provide an open architecture specification (interoperability framework) that can be used by collaborative initiatives to set up easily their data integration projects, in a way that such collaborations can interconnect with EURECA compliant data sources, re-use EURECA compliant services, etc. We will provide detailed guidelines concerning the minimum requirements and the steps to follow by an organization that would like to join or deploy a EURECA-compliant environment, make use of the EURECA services, or build new services on top of the existing EURECA environment.

Several ICT-based services relevant to the clinical research and care communities will be implemented in the project, such as: efficient automatic recruitment, protocol feasibility evaluation and support for protocol design; services enabling data mining of large volumes of data to generate new hypotheses, identify safety risks, and carry out epidemiology studies; integrated reporting of serious side effects; use of PHR data for reporting of side effects; and contextualization of information. All these services were

identified as highly relevant by our communities of users. They are supported by the scalable and consistent semantic interoperability platform among EHRs and clinical trial systems, and by uniform access within the EURECA environment to a variety of external knowledge and data repositories.

Lowering the complexity threshold for the integration of service components to the minimum is key to the success of the EURECA interoperability framework. We will setup a certification programme checking services for “EURECA compliance” (similar to caBIG). This programme will be coordinated by the EuroRec Institute of whom certification and quality labeling is the core business. The “EURECA compliant” label will guarantee system integrators that a service can be deployed in a EURECA compliant environment without needing additional development.

## 6.2 Improve sustainability of healthcare services by enabling better use of resources

Significant effort and financial investments in the biomedical research and in the healthcare industry have resulted into a wealth of data and information with the potential to bring along large qualitative improvements in patient outcome. However, the heterogeneity of data, the low adoption of shared standards, the fragmentation with respect to methodology, infrastructure and tools, the duplication of efforts, and the insufficient collaboration across disciplines, organizations and industries limit the impact of these investments.

Paradoxically, despite the huge volumes of data generated in clinical practice, very little of it is used for pushing the frontiers of medical science forward. Lack of data still hampers basic research, the secondary use of the large amounts of patient data collected in clinical practice is low, and the link between clinical research and clinical practice is often broken, creating a widening knowledge gap between research and care. As the cost of healthcare in Europe becomes almost unaffordable, reducing the expenses while both accelerating research and significantly increasing the quality, safety and efficiency of care is a necessity.

The goal of the EURECA project is to enable seamless, secure, scalable and consistent linkage of healthcare information residing in EHR systems with information in clinical research information systems to support the two currently separated worlds of clinical research and clinical practice to connect and benefit from each other. Next to benefits concerning patient outcomes and safety, this integration has an important potential to bring along significant cost reduction.

On top of the achieved semantic interoperability we build software services and tools to support more efficient research, better care and improved patient safety. The reuse of EHR data for research can reduce the costs and duration of clinical trials and solve the current shortage of data in clinical and biomedical research. Access to large amounts of high quality data from clinical practice will support large scale data mining for epidemiology, better hypotheses generation, better study design and execution.

Besides saving money on the research side through more efficient and better designed clinical trials, we also save the time of the healthcare professionals and help reduce the number of errors by eliminating the need for multiple data entry in several scenarios: trial enrolment, trial execution, reporting of serious side effects. Early detection and reporting of safety issues and better assessment of the benefits of drugs versus risks (including through the use of patient-managed data out of PHRs) can save the patients a lot of suffering (and sometimes even save lives), can save in

hospitalization costs, and can reduce the costs of trials incurred by the pharmaceutical industry.

## 6.3 Guidance on healthcare information systems issues in "green field" member states

The absence of significant (digital) legacy systems can be an advantage for a healthcare organization that aims to deploy an efficient and cost-effective healthcare information system. Taking into account the goal towards an integrated Europe-wide healthcare information system, adopting from the start a standards-based approach can provide huge benefits to the healthcare organization in the future.

The fragmentation in available EHR solutions and the lack of use of standards by vendors are major issues, hampering research and generating unnecessary costs. At the same time, efficient collaboration with the pharmaceutical industry and attracting research funding should be a priority for "green field" member states. In this context, the best solution is the implementation of standards-based, integrated healthcare information systems, and this is where EURECA can help.

Little digital legacy also makes the information extraction task simpler. For the paper based legacy data it needs to be considered whether the effort to digitize the information pays for the benefits and only digitize or store in a structured way data that is likely to be used again.

Additionally, when starting fresh with the development of a new information system it is crucial to introduce proper information models and core data sets based on standardized terminologies, enabling to describe the information in a structured, meaningful way. Not needing to map to prior information systems will make this structuring effort easier and less resource intensive. The creator of the systems can be inspired by current standards and show cases demonstrating best practices. EURECA can provide a bootstrapping solution by delivering core datasets, an open reference architecture and comprehensive guidelines.

EURECA envisages an open, scalable SOA. This approach reuses existing systems and services when possible (speeding up initial deployment), ensures extensibility on the long run, allows for integration of (future) third party solutions and will enable interfacing with other platforms.

We will provide support to users who would like to develop based on our reference architecture, join our environment, make use of our interoperability layer and services, and provide services on top of EURECA. We will provide detailed guidelines concerning the minimum requirements and the steps to follow by an organization that would like to join or deploy a EURECA-compliant environment, make use of the EURECA services, or build new services on top of the EURECA environment.

Additionally we will follow open source developments concerning EHR and CT systems with the goal to provide easy integration of such solutions. There are at this moment already two open source CT systems to which interoperability is envisioned: OpenClinica and ObTiMA. We will setup a certification programme checking services for "EURECA compliance". The "EURECA compliant" label will guarantee system integrators that a service can be deployed in a EURECA-compliant environment without needing additional development.

The large clinical research networks that have joined EURECA have a Europe-wide and International reach. This enables us to promote our solutions and approaches in all European countries and far beyond.

## 6.4 Accelerated establishment of interoperability standards and of secure, seamless communication of health data between all involved partners

At the core of the EURECA project is achieving semantic interoperability among EHR and clinical trial systems, consistent with existing standards (HL7 v3, CEN 13606, CDISC, etc.), while managing the various sources of heterogeneity: technology, medical vocabulary, language, etc. This requires building sound information models describing the EHR and the clinical trial systems, and capturing the semantics of the clinical terms by standard terminology systems such as SNOMED CT, ICD, LOINC.

The scalability of the solution will be achieved by modularization, identifying core data sets covering the chosen clinical domains. Achieved semantic interoperability among EHR and clinical trial systems opens the way for services and tools that support the health professionals in avoiding double data entry, enable automatic identification of patients for new clinical trials and, on the clinical care side, automatic identification of potentially relevant trials for patients, improve patient safety, and support epidemiology and cohort studies, and large scale collaboration.

EURECA will research and develop solutions to fulfil the data protection and security needs and the legal, ethical and regulatory requirements related to linking research and EHR data, enabling users to exchange data and collaborate in a trusted, secure environment.

EURECA focuses on open scalable standards-based solutions addressing all security, privacy and regulatory requirements. The wide adoption of our standards-based open framework will contribute to accelerate the implementation of Europe-wide interoperability and of fully secure information exchange.

## 6.5 Faster medication innovation and lower costs through a more efficient research process

EURECA will develop solutions to improve the efficiency of the research process taking into account all major stakeholders: independent investigators, clinical researchers, the pharmaceutical industry, patients and healthcare professionals. Gathering support for the large scale adoption of our solutions from both communities will help us achieve EURECA's full potential.

**Improving the Competiveness of European Biopharmaceutical Industry**

The Pharmaceutical Industry is a very important pillar of European Industry. Competition within the pharmaceutical industry continues to intensify, and competing for clinical investigators and patients is no exception.

While pharmaceutical companies invest heavily in marketing approved drugs, they often do not employ that same market research and marketing expertise when it comes to targeting, positioning and communicating the value of clinical trials to study sites and patients.

> ..European legislation governing clinical trials is slowing research and may even be costing lives, leading scientists have warned…
>
> *D. Cressey*
> *European clinical trials rules under fire*
> *Nature News, Published online 13 March 2009, Nature, doi:10.1038/news.2009.163*

The research data underscores the high stakes and urgent need for pharmaceutical companies to improve the clinical trial process. Effective communications can result in better-selected study sites and patients who will remain with a study until it is completed, saving time and money in clinical trials[11].

In Europe a number of new regulations, directives and corresponding guidelines have been issued over the last decade. In particular, the EU Directive 2001/20/EC on Implementing Good Clinical Practice in the conduct of clinical trials on medicinal products for human use has dramatically affected the regulation, oversight and practice of clinical research in Europe. The intention of the Directive was to protect patients and make the European pharmaceutical industry more competitive by ensuring that all Member States had the same rules. The weight of the directive at least doubled the costs of clinical trials[12] and increased the administrative burden so much that many academic researchers could no longer perform such trials.

The transposition of the Directive into the legislature of Member States has led States to make additional changes, with often tougher requirements on trial designs. The variety of interpretations of the Directive by Member States has thus ended up in many countries having different rules, hence increasing rather than decreasing disharmonies between EU Member States [13]As a consequence, Sweden, for instance, has seen a 25% decrease in academic clinical trials, Ireland 60% and Poland a staggering 90% reduction[14]. This is also an issue that needs to be quickly addressed.

The pharmaceutical industry has identified the need to integrate clinical trial data with healthcare data to improve efficiency of the clinical trials. By providing standard-based semantic interoperability on domains of concepts (in specific clinical areas) between the two types of systems EURECA supports enhanced clinical trial conduct through improved efficiency and accuracy. The sponsor of the trial will benefit from the achieved interoperability by having access to a larger population, being able to compare safety data from a clinical trial to a much larger baseline, and accessing larger volumes of data for analysis.

The EURECA service that aims to identify potential candidates for clinical trials based on EHR data has the potential to increase the patient accrual reducing the trial duration and the time-to-market of the pharmaceutical products.

Access to longitudinal patient data beyond the end of a clinical trial can allow monitoring of the newly released drugs and support the early detection of serious side effects of treatments and drugs, not only reducing the negative effects on patients, but also reducing the risk of damaging the reputation of the pharmaceutical company and of decreased public trust.

Eliminating the need for multiple data entry along the workflow of a clinical trial, the EURECA environment will improve the ease and efficiency of study execution, reduce the transcription errors and the necessary verification steps, and eliminate redundant processes. All these have the potential to save time and lower the costs of trials, additionally freeing resources and enabling an increase in the number of trials.

Better access to increased patient populations, by accessing EHR data and by achieving more efficient enrolment in trials, may enable research in rare diseases or addressing population sub-groups that are often excluded (e.g. elderly patients).

[11] Communicating the Value of Clinical Studies: Recruiting and Retaining Investigators & Patients. http://www.benchmarkingreports.com/ businessoperations/op85_clinical_communications.asp

[12] J. Hear, R. Sullivan. The impact of the 'Clinical Trials' directive on the cost and conduct of non-commercial cancer trials in the UK. *European Journal of Cancer*, vol. 43, pp. 8-13, 2007

[13] Editorial. Striking the right balance between privacy and public good. Lancet, vol. 367, pp. 275, 2006.

[14] R. Hoey. The EU clinical trials directive: 3 years on. Lancet, vol. 369, pp. 1777-8, 2007

The EURECA adverse events reporting support aims to eliminate independent data entries in the reporting and description of adverse events in clinical trials, improving the efficiency of the reporting and reducing the number of inconsistencies (e.g. due to different terminology) and errors. The sponsor of the trial can access pertinent information faster and assess sooner the causes and the outcome.

**Improving the Capability of Healthcare professionals and Patients**

The fact that many physicians, surgeons, and other members of the healthcare team do not encourage their patients to consider participation in clinical trials also contributes to low enrolment. Results from one study showed that a recommendation by their physician was the primary factor influencing patients' decisions to enrol in a trial.

According to a study survey conducted in 2000 by the American Society of Clinical Oncology (ASCO), the most significant barriers to patient enrolment included the intensity of paperwork collection and filing, and the extra time needed to train staff in the completion of enrolment and data collection forms. "Recruitment and adherence are very closely linked since those recruited must be followed to study completion as the inception cohort. …There is a clear impact on recruitment in terms of cost and both screening and staff burden"[15].

In responding to this challenge the EURECA project will deliver:

- Tools to assist clinicians in understanding their clinical data. Thus they should be quickly supported in identifying appropriate patient cohort for participation in a specific study and identifying relevant trials for a specific patient.
- Simplified and more efficient data entry and improved data retrieval. Tools and services that will ease the administrative burden of participating in CTs, by avoiding duplication of data entry, automating the reporting process, etc. A one-time data entry will replace the current multiple entries of the same data in different systems.
- Improved data quality and consistency, fewer errors and fewer verification steps due to single data entry
- Tool to simplify and increase the efficiency and reliability of serious adverse events reporting and management.
- Increased efficiency in trial management, allowing the contribution to more trials with the same resources
- Efficient presentation of a patient's relevant history, including data from clinical trial participation, contextualized to the current health episode.
- A semantically rich model for describing CTs for publication into CT repositories and a prototype of such a repository.

More physicians could become involved in clinical research due to lower costs and resource requirements, increased efficiency and simplified process of including patients in trials and of data collection.

[15] Wright JR, Crooks D, Ellis PM, Mings D, Whelan TJ; Factors that influence the recruitment of patients to Phase III studies in oncology. Cancer. 2002; 95(7):1584-1591

**Improving the Capability of Independent Investigators**

Investigator-driven clinical trials are clinical trials that are initiated by academic researchers and are aimed at acquiring scientific knowledge and evidence to improve patient care. Such studies deal with potential diagnostic and therapeutic innovations that do not attract or could even be against commercial interest. Typical examples are proof of concept studies, studies on orphan diseases, comparison of diagnostic or therapeutic interventions, surgical therapies or novel indications for registered drugs. They have a much broader scope and potential impact than industry-driven clinical trials, form a key part of patient-oriented clinical research, and create the basis for continually improving patient care[16].

Independent investigators are faced with all the main issues encountered in pharma-led trials, such as high costs, low recruitment, duplication of trials and complex and inefficient execution, however with a higher impact due to their access to very limited resources. Lack of funding, difficulties with collecting and managing sufficient data and unavailability of appropriate infrastructure have been identified as important obstacles[16].

In responding to this challenge the EURECA project will deliver:

- Services to automatically identify candidates for trials, improving enrolment
- Semantically interoperable linkage between EHR and CT systems to avoid duplicate tasks that currently increase the costs of clinical research
- Semantically-aware uniform access to extensive amounts of data from clinical care to be used for research (new hypotheses generation, research in rare diseases, cohort and association studies, etc.)
- Possibility to seamlessly follow up patients beyond the end of a clinical trial
- Services to assist in the analysis of their clinical data and recommendation services for appropriate patients for participation in specific CTs.
- Dissemination and educational material promoting the benefits of participating in CT, both for the sake of research but also in order to receive better care.

## 6.6 Social Impact

Allowing for discoveries in the laboratory to be quickly transferred to the clinical management and treatment of patients and obtaining societal benefits.

A standardized representation of a trial would promote the ability to determine the applicability of a trial result in the treatment of an individual patient.

The ultimate purpose of a clinical trial is to discover a method to improve the health or quality of life for an individual patient. One of the major challenges that repeatedly arises relates to the determination of whether or not the outcome of a particular trial or set of trials is applicable to a given patient. The decision often involves a review and evaluation of the structure of the trial with particular attention to the eligibility criteria, treatment regimen, and observed results.

[16] Investigator-Driven Clinical Trials, Forward Look, European Science Foundation, 2009

The lack of a standard trial description makes this task more difficult both in terms of simply locating the necessary information within a published trial description, but also in the interpretation and understanding of that information[17]. Additionally, the lack of access to information on trials with negative results may mean that instead of spending money efficiently on new research directions, research dead-ends are repeated with negative consequences for patients.

An IOM report concluded that "it is simply not acceptable for patients to be harmed by the same health care system that is supposed to offer healing and support"[18]. This statement refers to the avoidable harm brought to patients both by errors of omission and errors of commission and by undetected serious side effects of treatment and drugs.

Critical elements for providing safe care to patients include assembling a comprehensive picture of the patient case, applying this knowledge appropriately and effectively, monitoring the effects of disease and therapies on the patient over time and detecting and preventing errors that could harm the patient[18].

Improved patient enrolment in clinical trials does not only hold the potential of improving the treatment for future patients, but it may also provide better outcome for the participating patients as statistics show better outcomes of patients treated in clinical trials.

Other important aspects of patient safety are the integration of all the data that is known about a patient from all previous care encounters and from all systems that hold data about that particular patient, and the enhancement of the adherence to clinical protocols and guidelines[19].

By achieving semantic interoperability between the EHR and clinical trial systems and by building software services/modules to:

- Automatically match potentially-eligible patients with relevant clinical trials,
- Provide access and mine longitudinal care data to detect potential safety risks,
- Carry out data mining and build research hypotheses,
- Contextualize patient and clinical information to the patient case,
- Improve efficiency and communication by avoiding double data entry

The EURECA environment supports improvement in patient recruitment, targeted selection of patient-case information and clinical guidelines for better outcome, higher data quality and better safety. Faster recruitment also leads to improved research and time-to-market of new drugs.

Improved clinical trial accrual and better patient outcome is also supported by providing uniform access to repositories of running clinical trials information, enabling the physician and the patient to find suitable clinical trials.

[17] Zarin DA, Tse T. Medicine: moving toward transparency of clinical trials. Science 2008;319:1340 –2
[18] LT. Kohn et al., To Err is Human: Building a Safer Health System. Washington DC: National City Press, 2000
[19] PM. Kilbridge et al., The Informatics Oportunities at the Intersection of Patient Safety and Clinical Informatics, JAMIA 15 (4), 2008

# 7 SUMMARY

Despite large investments in IT, the healthcare domain is currently unable to obtain the desired benefits in quality, safety and efficiency of care and to use the IT systems at their full potential. The lack of integration and of semantic interoperability among the systems deployed is a significant source of inefficiency, data inconsistencies, unnecessary costs and an unacceptably large number of medical errors. Furthermore, the pharmaceutical industry faces low recruitment rates of patients and extremely high costs of running clinical trials due to lack of interoperability and complex and inefficient study execution, while having a strong need to reduce research expenses and the time-to-market of new drugs.

The EURECA project aims to build an advanced, standards-based and scalable semantic integration environment enabling seamless, secure and consistent bi-directional linking of clinical research and clinical care systems to:

1. **Support more effective and efficient execution of clinical research** by:
   a. Allowing faster eligible patient identification and enrolment in clinical trials,
   b. Providing access – in a legally compliant and secure manner – to the large amounts of patient data collected in the EHR systems to be re-used in clinical research, for new hypotheses building and testing (e.g. to benefit rare diseases), study feasibility, as well as for epidemiology studies,
   c. Enabling long term follow up of patients, beyond the end of a clinical trial,
   d. Avoid the current need for multiple data entry in the various clinical care and research systems during the execution of a study.
2. **Allow data mining of longitudinal EHR data for early detection of patient safety issues** related to therapies and drugs that would not become manifest in a clinical trial either due to limited sample size or to limited trial duration, and eliminate duplicate reporting of identified serious side effects,
3. **Allow for faster transfer of new research findings and guidelines** to the clinical setting (from bench-to-bedside),
4. **Enable the healthcare professionals to extract, in the context of each patient's case, the relevant data** out of the overwhelmingly large amounts of heterogeneous patient data and treatment information.

> Integrating EHRs and clinical trial data would provide a significant step towards increasing the efficiency and effectiveness of clinical trials […]. However, differences in terminology, classifications, standards, roles of clinicians, data quality management and software technology between the worlds of medical care and medical research need to be addressed.
>
> "Meeting the Challenges of Patient Recruitment. A Role for Electronic Health Records" C. Ohmann and W. Kuchinke,
> Int. J. Pharm. Med., 2007

At the core of the project is achieving **semantic interoperability among EHR and clinical trial systems, consistent with existing standards**, while managing the various sources of heterogeneity: technology, medical vocabulary, language, etc. This requires building sound information models describing the EHR and the clinical trial systems, and capturing the semantics of the clinical terms by standard terminology systems. The scalability of the solution will be achieved by modularization, identifying **core data sets covering the chosen clinical domains**. While the core vocabulary and the data models will initially describe sub-domains in oncology, the tools and software services developed in EURECA will be extendable to a different disease space by developing a new set of models and identifying core concepts specific to that disease.