Linked2Safety

FP7-288328

A Next-Generation, Secure Linked Data Medical Information Space for Semantically-Interconnecting Electronic Health Records and Clinical Trials Systems Advancing Patients Safety in Clinical Research

---

Deliverable D7.1

Linked2Safety Showcase Scope Definition

Editor(s):	Ann Gledson, John Keane, Xiao-Jun Zeng, Goran Nenadic, David Tian (UNIMAN) Iolie Nicolaidou (UCY)
Responsible Partner:	UNIMAN
Status-Version:	Final – v2.1
Date:	30/06/2013
EC Distribution:	Public

Project Number:	FP7-288328
Project Title:	Linked2Safety

Title of Deliverable:	D7.1 - Linked2Safety Showcase Scope Definition
Date of Delivery to the EC:	30/06/2013

Workpackage responsible for the Deliverable:	WP7 – Linked2Safety Clinical Trials Showcases Development and Performance Evaluation
Editor(s):	Ann Gledson, John Keane, Xiao-Jun Zeng, Goran Nenadic, David Tian (UNIMAN) Iolie Nicolaidou (UCY)
Contributor(s):	UNIMAN, UCY, CHUV, ZEINCRO, CING, LUH
Reviewer(s):	INTRASOFT
Approved by:	All Partners

Abstract:	This deliverable presents the detailed scope definition of the Linked2Safety showcases, including the identification of the realistic clinical research use-cases that are going to be executed, the AS IS analysis of clinical trials use-cases, and their detailed organization and planning. It outlines the protocol and procedures that will be followed during the evaluation of the performance of the showcases and quantifies the expected benefits and their assessment measures and metrics.
Keyword List:	Showcases, Scenario-based evaluation

Document Description

Document Revision History

Version	*Date*	*Modifications Introduced*	
		Modification Reason	*Modified by*
0.1	25/04/2013	Outline of the deliverable	UNIMAN
0.2	27/05/2013	Revised Outline Revision: added DM to Showcase Type 3	UNIMAN
0.3	28/05/2013	Removed references to who authors sections.	UNIMAN
0.4	30/05/2013	Added/Updated Section 5 sub-sections. Old Section 5 became 5.3.3	UNIMAN
0.5	06/06/2013	Added explanations to Sections 3 and 4 Added timelines and gantt chart for Section 5.3	UNIMAN
0.6	11/06/2013	Added UCY's scenario-based evaluation part to Section 6. Added Section 6.1 'Internal showcase evaluation' Re-structure of Section 5. Updated Section 5, added to the To-Be workflows section.	UNIMAN
0.7	14/06/2013	Updated Section 6 (and removed Section 6.1)	UCY
0.8	17/06/2013	Add UCY Section 6 (v0.7). Place Showcase templates into appendices, and check referencing. Add UCY's update to Section 4.4.5 'Scenarios of Showcase 3: Identification of relations between molecular fragments and	UNIMAN

		specific adverse side effect categories' Add new Section 4 as updated by CHUV, CING and ZEINCRO. (Ongoing) Add CHUV / ZEINCRO / CING / INTRASOFT edits. Rewritten introduction.	
1.0	21/06/2013	Added Executive Summary Added Conclusions Incorporated comments from CING/ZEINCRO/CHUV for Sections 1-5. Replaced ZEINCRO 'Pending feedback from the client…' statements in 'Background and expected outcome' section of type-2 templates with agreed statement. Added latest version of Section 6 from UCY Add Lists of Figures, Tables and Abbreviations	UNIMAN
1.1	24/06/2013	Corrected references to appendices in Section 3	UNIMAN
1.2	27/06/2013	Added new Section 6 from UCY Corrected mistake in data matrix Corrected data availability mistakes in Zeincro showcase templates	UNIMAN
2.0	27/06/2013	Revision from internal review	UNIMAN
2.1	28/06/2013	Updated template: Headers and footers	UNIMAN

Contents

List of Figures

List of Tables

Definitions, Acronyms and Abbreviations

Table 1: Definitions, Acronyms and Abbreviations

Acronym	Title
ADM	Anti-Depressant Medicine
ADE	Adverse Drug Event
AE	Adverse Event
BMI	Body Mass Index
CVD	Cardio-vascular Disease
CVRF	Cardio-vascular Risk Factors
DPI	Dry Powder Inhaler
EDC	Electronic Data Capture
EHR	Electronic Health Record
MDD	Major Depressive Disorder
NDA	New Drug Application
OR	Odds Ratio
PHR	Patient Health Record
RDF	Resource Description Framework
SSRI	Selective Serotonin Reuptake Inhibitors
TCA	Tricyclic Anti-depressants

1 Executive Summary

This document is Deliverable "D7.1 – Scope Definition" (henceforth referred to as D7.1) of the Linked2Safety project. The main objective of this document is to present a detailed scope definition of the Linked2Safety showcases, including the identification of the realistic clinical research use-cases that are going to be executed, the "AS IS" analysis of clinical trials use-cases, and their detailed organization and planning. It outlines the protocol and procedures that will be followed during the evaluation of the performance of the showcases and quantifies the expected benefits and their assessment measures and metrics.

Three Showcase types have been defined in the Linked2Safety description of work (DOW). In order to create the scope definition, a showcase scenario template has been designed, taking into consideration the requirements identified in WP1. These extended showcases have been defined by the three clinical partners taking into consideration the available data. The resulting showcases demonstrate the three showcase types as defined in the DOW:

a) Subject Selection: the unbiased randomized selection of subjects in clinical research;
b) Phase IV Clinical Trial: Post Marketing Surveillance of a drug after it receives
permission to be sold; and
c) Identification of relations between molecular fragments and specific adverse side effect categories.

In order to enable clarity, showcases in this and all future Linked2Safety documents will refer to these now developed extended showcases, while the abstract definitions contained in the initial proposal will be referred to as showcase types.

A full set of pre-requisites has been established that, alongside the WP1 requirements, contribute to the detailed planning of the Workpackage 7 Tasks 7.3 to 7.6:

- Task 7.3: Interoperable EHRs Deployment
- Task 7.4: Platform Deployment and Configuration
- Task 7.5: Showcase Execution
- Task 7.6: Performance Evaluation and Lessons Learnt

Finally, an outline of the protocol and procedures to evaluate the performance of the showcases is derived. A scenario-based assessment is to be used, based upon the scenarios developed by the clinical partners. Evaluation will involve extensive data collection with regard to the end-users' feedback on their interaction with the Linked2Safety platform and showcases. Example scenarios are presented for each of the three user types of interest to the project: medical science analysts, analytic methodology engineers and data providers. It should be mentioned that Tasks 7.6 'Performance Evaluation and Lessons Learnt' will be detailed in full,

along with all scenarios, in the following Workpackage 7 deliverable, due in month 24.

The scope definition process, reported in this document has resulted in the creation of 13 showcase scenarios, each detailing the data to be used, the objectives to be achieved, how it is currently performed ("AS IS"), how it will be performed using the Linked2Safety platform ("TO BE") and the Linked2Safety tools required. Detailed timelines have been produced for each task to execute the scenarios. In addition a proven, scenario-based evaluation methodology has been outlined, aimed at the three core end user types. The final result is a comprehensive strategy for the entire workpackage, with a carefully planned set of tasks and subtasks.

2 Introduction

The purpose of this deliverable is to document the detailed scope definition of the Linked2Safety showcases, including the identification of the realistic clinical research use-cases that are going to be executed, the "AS IS" analysis of clinical trials use-cases, and their detailed organization and planning. A definition of the protocol and procedures to be followed during the evaluation of the performance of the showcases is also included - which will be expanded upon in the next Task 7.2 deliverable (due month 24).

2.1 Workpackage 7

The Workpackage 7 showcases are one of the main indicators of success of the Linked2Safety project. They comprise 'a step-by-step cookbook including the adoption guidelines for leveraging the quality and the effectiveness of EHR reuse in clinical research for improving patient identification, selection and safety in the design of clinical trials, based on the lessons learnt through the evaluation of the performance of the clinical research showcases' [1].

The definition, realization and execution of the Linked2Safety clinical trials showcases leads to the stress-testing of the implemented ontologies, models, tools and services under pragmatic and realistic conditions. The monitoring and evaluation of the performance of the showcases will be used to generate a set of lessons learnt and methodological adoption guidelines for the reuse of semantically interoperable EHRs for advancing proactive patient' safety in clinical trials.

The main Workpackage 7 objectives are as follows:

- To define, scope and organize the Linked2Safety clinical trials showcases that focus on their applicability and usability in pragmatic conditions;
- To develop, operate and support the Linked2Safety clinical trials showcases;
- To test and monitor the performance of the Linked2Safety showcases' experimental environments;
- To define and validate the evaluation strategy of the performance of the Linked2Safety clinical trials showcases;
- To collect feedback from end-users and system engineers;
- To generate the lessons learnt. [2]

By the time the three main research and technological development results of the Linked2Safety project are developed and delivered, that is:

a) the Linked2Safety Reference Architecture (WP1),

b) the Semantic EHR Model (WP1), and

c) the integrated Linked2Safety Platform (WP6),

the definition, development, deployment and execution of the three Linked2Safety clinical research showcase types will be realized so as to demonstrate the applicability and usability of the Linked2Safety results in realistic setups and scenarios.

The purposes of the clinical research showcases: are:

a) to validate the concepts, models and tools resulting from the core technological R&D activities (carried out in WP1-WP6) in terms of addressing the identified requirements from stakeholders (in WP1),

b) to test and evaluate the developed genetic data analysis tools for the early detection and identification of genetic factors associated with adverse events (in WP5), and

c) to provide feedback to the second development cycle of the software development and integration work-packages, so as to refine the generated models and applications, and especially to validate and improve the functionality of the genetic data analysis tools implemented (in WP5).

All the clinical research end-user organizations participating in the Linked2Safety consortium, i.e. the Cyprus Institute of Neurology and Genetics (CING) maintaining its own EHR repository, the Lausanne University Hospital (CHUV) maintaining its own PHR repository, and ZEINCRO Hellas (ZEINCRO) providing quality consultancy services to the biopharmaceutical industry, as well as selected members of the Clinical Research and Patients Safety Special Interest Group (already established by the project partners) will actively participate in the Linked2Safety showcases, piloting the Linked2Safety tools and platform under pragmatic conditions.

The three showcases will involve:

a) Subject Selection: the unbiased randomized selection of subjects in clinical research;

b) Phase IV Clinical Trial: Post Marketing Surveillance of a drug after it receives
permission to be sold; and

c) Identification of relations between molecular fragments and specific adverse side effect categories.

The precise scope of each showcase will be jointly defined by the associated clinical research teams.

The activities, identified to facilitate the utilization of the Linked2Safety Platform to each pilot showcase, are grouped in the following objectives:

a) the AS IS analysis and the definition of the scope of each Linked2Safety experiment focusing in the applicability of the Linked2Safety project results in pragmatic collaborative research scenarios;

b) the centralized deployment and operation of the integrated Linked2Safety platform;

c) the application of the developed models and tools for the alignment of the EHR/PHR and EDC repositories existing in the pilot sites of the participating clinical research end-user organizations,

d) the execution and evaluation of the performance of the showcases, validating the overall Linked2Safety solution. [1]

2.2 Workpackage Progress

The following is a list of the tasks in this workpackage, highlighting current progress:

Task 7.1 Clinical Trials Showcases Scope Definition *(UNIMAN, Duration: M16 – M21)*

This task is completed and a full description of the scope definitions are outlined in Section 3 (showcase design) and Section 4 (showcase descriptions).

Task 7.2 Evaluation Strategy Definition and Validation (*UNIMAN*, *Duration: M19 – M24*)

The timelines of showcase evaluation are outlined in Section 5, and an outline is given of the evaluation strategy in Section 6, to be further detailed in the *Evaluation Strategy Definition and Validation* deliverable due in month 24.

Task 7.3 Interoperable EHRs Deployment *(SIVECO, Duration: M22 – M30)*

A detailed plan and timeline for this task is provided in Section 5.2.

Task 7.4 Linked2Safety Platform Deployment and Configuration *(INTRASOFT, Duration: M28 – M33)*

A detailed plan and timeline for this task is provided in Section 5.3.

Task 7.5 Clinical Trials Showcases Execution *(CHUV, Duration: M31 – M36)*

A detailed plan and timeline for this task is provided in Section 5.4 and examples of the scenario-based tasks to be executed by end-users are outlined in Section 6.

Task 7.6 Performance Evaluation and Lessons Learnt (*UNIMAN*, *Duration: M31 – M36*)

A detailed plan and timeline for this task is provided in Section 5.4 and the implementation of this evaluation framework is outlined in Section 6. This task is to be further detailed in the Evaluation Strategy Definition and Validation deliverable due in month 24.

2.3 Document Outline

This deliverable begins by documenting the scope definition of the Linked2Safety showcases. This includes the identification of the realistic clinical research use-cases that are to be executed. A showcase scenario template has been designed (Section 3), derived from the requirements identified in WP1. These were instantiated by the clinical partners, each producing details of at least three showcases (presented in Section 4 in brief and the appendices in full), using the available data from all three clinical partners.

Section 5 describes the detailed organisation and planning of the showcases, beginning with an outline of the pre-requisites for their successful execution (Section 5.1). The Interoperable EHR Deployment and timeline (Task 7.3) is described in Section 5.2, followed by the platform deployment and configuration (Task 7.4) in Section 5.3, and then the detailed workflows, execution (Task 7.5) and evaluation (Task 7.6) timelines in Section 5.4. A Gantt chart is displayed in Section 5.5 which illustrates each of the above workpackage 7 tasks.

Finally, Section 6 outlines the protocol and procedures to evaluate the performance of the showcases. A scenario-based assessment is proposed, based upon the scenarios developed by the clinical partners, and users are divided into the three core user types: Medical Science analysts, Analytic methodology engineers and Data providers. Evaluation is to involve extensive data collection with regard to the end-users' feedback on their interaction with the Linked2Safety platform and showcases.

3 Showcase Development Methodology

This section describes the design of the Showcase Scenario Template, explains step by step (i.e. item by item) the process of template completion and describes how the resulting template contains the information needed for the showcases and for D7.1.

It illustrates how the information required in D7.1 has been included in the template via sections such as 'clinical research use-cases, to specify the tasks that can be performed using the platform; 'feasibility', to describe whether the showcase is 'realistic'; and the "AS IS" analysis, to describe the methodology used before using the Linked2Safety platform. It also states how the information and material for Section 5 (Detailed Organisation and Planning) and Section 6 (Protocol and Procedures to Evaluate Performance of Showcases) has been obtained. Finally, it demonstrates how the template has provided the foundation for and is an effective way to develop the showcase.

3.1 Design of Showcases

3.1.1 Overview

The showcases are designed to demonstrate the entire cross-section of the Linked2Safety architecture and functionality. In its entirety, the set of showcase templates must therefore include:

1. In-depth modelling:
   - Development of a "showcase scenario template" to explain the task, associated data, methods to be run etc.
2. Creation of semantically-enriched RDF data cubes:
   - Data cubes generation from the raw clinical data
   - Use of Interoperable EHR Data Space components (common EHR Schema, semantic EHR Model, RDF) in off-line closed-world rooms
3. Semantic linking of RDF data cubes
   - Multiple data cubes are semantically linked with instances of the Linked Open Data (LOD) cloud.
   - The linked data cube is to be stored on a repository situated on the data providers' premises.
4. Application of statistical and data analytic techniques as appropriate
   - Data Analysis Space component tools will perform Subject Selection, Post Marketing Surveillance and Chemoinformatics Analysis.

5. Provision of evaluation and feedback

   - Expected outcomes, benefits and metrics for the evaluation.

3.1.2 Showcase Development Process

As shown in Figure 1, the showcase development is divided into the following steps:

Step 1. *Propose a showcase* which includes showcase type, showcase name, showcase proposer (owner), and background and expected outcome. Only if an identified showcase has the potential to be used to prove the concepts (based on the background information) and benefits (based on the expected outcome) of Linked2Safety, the case will be further developed;

Step 2. *Describe the data resource* which is to verify whether the information/data required for the showcase is available and whether it can be used to show the Linked2Safety benefits (such as data link and effectiveness). This is the first check whether the showcase is feasible and suitable;

Step 3. *Carry out the "AS IS" analysis* (without the Linked2Safety platform) which shows the current (i.e., "AS IS") clinic trial process for the identified showcase, the "AS IS" workflow and Current issues/Linked2Safety improvements. "AS IS" analysis will be used to identify the opportunities, on how the process can be improved and enhanced by the Linked2Safety platform and to act as the basis to develop the "TO BE" model (i.e., the new clinical trial process based on Linked2Safety);

Step 4. Based on the "AS IS" analysis and areas for improvement, *identify the objectives of this Linked2Safety showcase*;

Step 5. *Propose the "TO BE" Modelling based on the Linked2Safety platform*, which is a new clinical trial process including "TO BE" methods and tools and the "TO BE" workflow. "TO BE" Modelling will be proposed initially in D7.1 and will be refined and finalised based on an iterative process with the progress of Linked2Safety platform development. This is also the basis for developing the evaluation protocol and procedures in Task 7.2;

Step 6. *Carry out the feasibility analysis* which will check and verify the feasibility of the showcase. This is to ensure that the showcase is both *realistic and executable*. The feasibility analysis is performed from three aspects:

   a. Data availability feasibility: whether the data for the proposed showcase is available now or when it is available and whether the available data can support to achieve the identified objectives/benefits of the Linked2Safety platform;
   b. Data operation feasibility: If data is available, check whether the data cube generator will produce the semantically enriched and linked data as required. For example, I like to link and interoperate my data with another resource (such as a set of external data), whether Linked2Safety can do it correctly;

c. Data analysis feasibility: For given objectives (i.e., a showcase with the desired benefits) and availability of data, identify whether we have the appropriate data analysis or mining method capable to discover knowledge within the data and show the benefit.

Step 7. Identify risk and alternative which will be used to mitigate the risk in the showcase planning and ensure the success;

Step 8. Establish assessment measures and metrics for expected results, which is a list of measurable criteria to show and prove the functionality of the Linked2Safety platform. This is to ensure that each showcase will demonstrate, validate and/or exercise the effectiveness and efficiency of Linked2Safety platform.

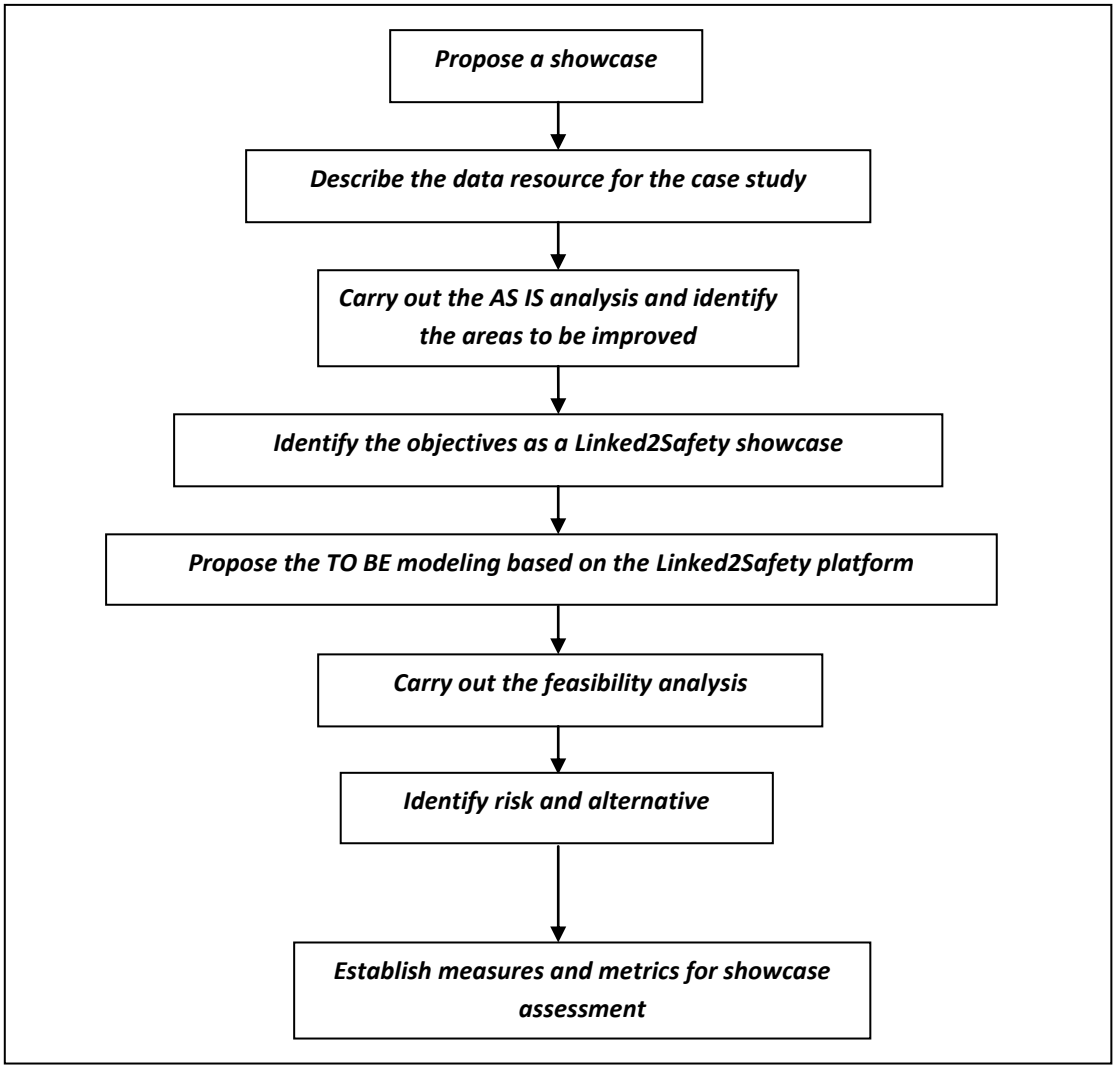

Figure 1 - Showcase Identification, development and verification process

3.1.3 Showcase Template

In order to fulfil the showcase development process outlined in Section 3.1.2, showcase templates were developed in collaboration with the clinical partners participating in Linked2Safety, to be completed by their expert users. These templates are shown below (Table 2), with explanations and/or possible answers shown in italics, in the right hand column.

Showcase Type	*Select from the Showcase List (currently):* *1. Subject Selection* *2. Post Marketing Surveillance* *3. Identification of relations between molecular fragments and specific adverse side effect categories*
Scenario Name	*Example: Subject Selection: Effect of drug A to Cardiovascular Disease*
Owner	*Who will conduct this trial showcase.*
Background and Expected Outcome	*Explain background and expected outcome of trial case*
Data to be used	*Describe what data is available for the proposed trial such as what kind of data is available internally (locally)? How big is the dataset? How will it be semantically enriched? What kind of external data would you wish to link to this data?*
"AS-IS" Methods and Tools used (without L2S platform)	*Describe existing methods and tools normally used in the proposed trial.*
"AS-IS" Workflow (without L2S platform)	*Describe existing process if the trial is carried out without the Linked2Safety platform, including normal flows and possible alternative flow(s). See WP1-D11_Requirement Analysis for examples of workflows.*
Current issues /L2S Improvements	*List problems and difficulties in the existing practice, paying particular attention to issues which can be overcome by the Linked2Safety platform.*
Objectives as an L2S showcase	*Reasons for proposing trial as Linked2Safety showcase: potential impact; how it can show applicability, usability and benefits of the Linked2Safety platform.*
"TO-BE" Methods and	*Describe Linked2Safety components to be used: the Interoperable EHR Data Space, Linked Medical Data Space*

Tools used with L2S platform	*and Data Analysis Space*
"TO-BE" Workflow with L2S platform	*Definition of "TO BE": "TO BE" can either be a revised process addressing the problems from the "AS IS" analysis or a new process that takes into account any "AS IS" analysis.* *Task: Describe process for trial to be carried out on Linked2Safety platform, including normal flow and possible alternative flow(s).*
Feasibility Analysis	*The feasibility analysis includes checking:* *1. Data availability – internal/external* *2. Data operation – will data cube generator produce required linked data* 3. *Data analysis – can tools generate expected result*
Risk Identification and alternative	*Based on feasibility analysis, identify any possible risks of the showcase.*
Assessment Measures and Metrics to Expected Results	*Provide a list of Assessment Measures and Metrics to evaluate the expected results to show that the objectives of the showcase are achieved and Linked2Safety benefits can be evidenced.*
Any Other Comments	

Table 2: Showcase Template as sent to Clinical Partners

4 Clinical research scenarios

4.1 Introduction

This section explains the different types of clinical research or trials, the phases of drug development, the aims of clinical research and then finally describes the completed showcase templates in the context of these descriptions / types.

A clinical trial is an investigation on human subjects intended:

- to discover or verify the clinical, pharmacological and/or other pharmacodynamic effects of one or more investigational medicinal product(s), and/or
- to identify any adverse reactions to one or more investigational medicinal product(s) and/or
- to study absorption, distribution, metabolism and excretion of one or more investigational medicinal product(s) with the object of ascertaining its (their) safety and/or efficacy.

Clinical research is an investigation on human subjects intended:

- to discover associations between certain independent or predictor variables and a result/outcome that is meaningful in medicinal settings.

The showcases are designed to support each of these goals. The trials or research scenarios are also carried out in either one or multiple sites, and one or multiple countries to demonstrate that required data can be linked, and can still respect each institution's guidelines and regulations and adhere to regional/national laws.

4.2 Types and Aims of Clinical Trials

Depending on their focus and the intended objective, clinical trials are categorised as follows: Treatment Trials, Diagnostic Trials, Screening Trials and Quality of Life Trial (or Supportive Care trials).

All Clinical Trials are also classified as Interventional or non–Interventional. The definition of a non-interventional trial is a study where the medicinal product(s) is (are) prescribed in the usual manner in accordance with the terms of the marketing authorisation. The assignment of the patient to a particular therapeutic strategy is not decided in advance by a trial protocol but falls within current practice and the prescription of the medicine is clearly separated from the decision to include the patient in the study. No additional diagnostic or monitoring procedures shall be applied for the analysis of the collected data.

Lastly, clinical trials are classified according to phase: Phase I, II, II and IV. The four phases of clinical trials are described in the context of drug development, in the following section.

4.3 Phases of Drug Development

Clinical trials involving drug testing are classified into four phases, which are described below [3]. Each phase is designed to answer specific questions and includes several clinical trials where each one answers the questions of each phase.

Phase I

These are the first trials of a new active ingredient or new formulations in man, often carried out in healthy volunteers. Their purpose is to establish a preliminary evaluation of safety, and a first outline of the pharmacokinetic and, where possible, a pharmacodynamic profile of the active ingredient in humans.

Phase II

These trials are performed in a limited number of subjects and are often, at a later stage, of a comparative (e.g. placebo-controlled) design. Their purpose is to demonstrate therapeutic activity and to assess short-term safety of the active ingredient in patients suffering from a disease or condition for which the active ingredient is intended. This phase also aims at the determination of appropriate dose ranges or regimens and (if possible) clarification of dose- response relationships in order to provide an optimal background for the design of extensive therapeutic trials.

Phase III

These are trials on larger (and possibly varied) patient groups with the purpose of determining the short -and long -term safety/efficacy balance of formulation(s) of the active ingredient, and of assessing its overall and relative therapeutic value. The pattern and profile of any frequent adverse reactions must be investigated and special features of the product must be explored (e.g. clinically -relevant drug interactions, factors leading to differences in effect such as age).These trials should preferably be of a randomized double-blind design, but other designs may be acceptable, e.g. long-term safety studies. Generally, the conditions under which these trials are carried out should be as close as possible to normal conditions of use.

Phase IV

Studies performed after marketing of the pharmaceutical product. Trials in phase IV are carried out on the basis of the product characteristics on which the marketing authorization was granted and are normally in the form of post -marketing surveillance, or assessment of therapeutic value or treatment

strategies. Although methods may differ, these studies should use the same scientific and ethical standards as applied in premarketing studies. After a product has been placed on the market, clinical trials designed to explore new indications, new methods of administration or new combinations, etc. are normally considered as trials for new pharmaceutical products.

The drug development process will normally proceed through all these phases over many years. If the drug successfully passes through Phases I-III, it may be approved for use in the general population. Consequently, Phase IV deals with post-approval studies. Recently a supplementary phase, called Phase 0, has been introduced. Phase 0 Trials are a recent designation for exploratory, first-in-human trials conducted in accordance with the FDA 2006 Guidance on Exploratory Investigational New Drug (IND) Studies. The types, design, aims and timeframes for the different clinical trial phases are summarized in Figure 2 [4].

Stage of Development	Phase 1	Phase 2	Phase 3	Phase 4
Subjects	20-100 healthy volunteers	100-300 patient volunteers	1000-3000 patient volunteers	Real Population
End Points	Safety Profile	Efficacy, short - term safety, dose ranges or regimens	Safety/ efficacy balance	New indications, administration , combinations
Time Course	Year 3	Year 4-5	Year 6-8	Post marketing

Figure 2 – Design, aims and timeframes for the four clinical trial phases [4]

As the Linked2Safety showcases involve Phase III and Phase IV aspects these phases are further discussed in later sections. Linked2Safety is not designed to impact phase 1 and 2 because they involve very small numbers of subjects. In contrast, it can greatly impact phase 3 trials by helping in identifying institutions with sufficient number of subjects that meet the inclusion criteria of the study (although for access to the patients, further work will need to be performed by the actual institution). And for Phase 4 clinical trials, Linked2Safety could act as an early stage hypothesis generation and testing platform utilizing the available data in the platform in real time to discover associations with drugs on the market and adverse events.

4.4 Aims of clinical research

Clinical research is mostly classified as non-interventional. In other words, most of the research is carried out by assessing associations between clinically relevant phenomena which are either observed/measured objectively or reported by the patients themselves or by a third party (family members, health-care

professionals, subjects from the patient's natural environment). Clinical documentation may include information on various aspects of physical or mental health, potential causes of poor health or effects of treatment received (e.g. psychotherapy, pharmacological treatment) on health or well-being. An important aspect of pharmacological treatment that necessitates careful assessment is the effect of a specific drug or class of drugs on health in terms of adverse events. The Linked2Safety platform is also being developed to specifically allow for the assessment of the associations between such adverse events and clinically relevant phenomena that may have triggered the event.

4.5 Linked2Safety Showcases

The three showcases used in Linked2Saftey address the following areas:

I. Subject Selection
II. Phase IV Clinical Trial: Post Marketing Surveillance
III. Identification of relations between molecular fragments and specific adverse side effect categories.

Each showcase area is contextualised below and the associated scenarios detailed.

4.5.1 Showcase I: Subject Selection

Context: Phase III Clinical Trials

Phase III (interventional clinical) trials focus on the effectiveness of the studied drug in a variety of demographic and socio-economic subjects with variants of the disease under study. In addition, a comparison is usually made with gold standard drugs available on the market. It is imperative that the drug is shown to be effective and safe in this phase.

These trials are randomised controlled usually multi-center trials on large groups of human subjects (300–3,000 or more depending upon the disease/medical condition studied). Because of their size and comparatively long duration, Phase III trials are the most expensive, time-consuming and difficult to design and run, especially in therapies for chronic medical conditions. It is typically expected that there be at least two successful Phase III trials, demonstrating a drug's safety and efficacy, in order to obtain approval from regulatory agencies. When Phase III trials are completed with the drug being proved safe and effective, a New Drug Application (NDA) containing all manufacturing, pre-clinical and clinical data is filed. This collection of information is the basis for the marketing authorisation application dossier submitted to the regulatory authority/ies of the country/ies where the applicant (the pharmaceutical company "owner" of the drug) wishes to obtain approval for marketing the drug. The regulatory authority/ies will review the marketing authorisation application and, if all requirements (recommendations and guidelines) are met and the applicant has been able to

demonstrate that the drug is both safe and effective, the authority/ies will grant the marketing authorisation. This approval often comes with strict requirements for the company to conduct additional studies to keep the NDA active (usually involving paediatric trials and additional safety trials).

Context: Clinical research

The first showcase scenario demonstrates how the Linked2Safety platform avoids the amount of time (and cost) spent to locate subjects eligible to participate in a study or in a particular data analysis. It targets the identification of research groups that contain data on subjects with a given attribute. The platform provides overall frequency counts and stratified frequency counts by institutions enrolled with the requested information, so that researchers may contact these institutions to obtain the data to test the study hypothesis. If needed, these subjects could be re-contacted via the hosting institutions to obtain additional information that could semantically enrich each database in order to adequately test the study hypothesis with minimal time and cost. Examples of such scenarios are described in the following section.

4.5.2 Scenarios for Showcase I

The scenarios for this showcase are as follows:

I.I Anti-Depressant Drugs Evaluation (CHUV);
I.II Effect of Anti-hypertensive Drugs (CING);
I.III Asthma – Fluticasone/Salmeterol: Device A versus Device B (ZEINCRO);
I.IV Asthma - Formoterol DPI (A) versus Formoterol DPI (B) (ZEINCRO).

Scenario I.I: Anti-Depressant Drugs Evaluation (CHUV)

Scenario Name	Selection of participants with and without major depressive disorder (MDD) and hypertension in order to request additional information (if needed) on the use of antidepressant drugs.
Template	See Section A.1 Scenarios for Showcase 1

Scenario I.II: Effect of Anti-hypertensive Drugs (CING)

Scenario Name	Effect of anti-hypertensive drugs on Cardiovascular Disease in diabetes type II patients
Template	See Section A.1 Scenarios for Showcase 1

Scenario I.III: Asthma – Fluticasone / Salmeterol: Device A versus Device B (ZEINCRO)

Scenario Name	A study to compare the efficacy and safety of the combination of fluticasone / Salmeterol administered with device A against the innovator device B in patients with asthma.
Template	See Section A.1 Scenarios for Showcase 1

Scenario I.IV: Asthma: Formoterol DPI (A) versus Formoterol DPI (B) (ZEINCRO)

Scenario Name	A multicentre, randomized, double-blind, double-dummy, cross-over, single dose study, comparing the efficacy and safety of the test Formoterol DPI (A) versus the innovative Formoterol DPI (B) in patients with asthma.
Template	See Section A.1 Scenarios for Showcase 1

4.5.3 Showcase II: Phase IV Clinical Trial: Post Marketing Surveillance

Context: Phase IV Trials

The Phase IV trials (which may be interventional or non–interventional) ensure the safety surveillance and the ongoing technical support of a drug following approval. Specifically, the safety surveillance is designed to further explore and test the safety of the drug through the detection of rare or long-term adverse reactions over a much larger patient population and longer time period than was possible during earlier trials. Adverse Events identified during Phase IV trials may result in a drug being no longer sold, or restricted to certain indications or subpopulations.

Context: Clinical research:

The second showcase scenario demonstrates how the Linked2Safety platform avoids the amount of time (and cost) spent to perform an analysis of association testing on a given hypothesis in a single study. It targets the identification of research groups that contain data on subjects with a given attribute. If two or several institutions have data on similar attributes, the platform will allow for these sources to be merged in order to provide increased statistical power to test the study hypothesis. Therefore, a specific hypothesis may be tested on existing data without having to recruit new subjects or having to follow the subjects and collect the necessary information. Examples of such scenarios are described in the following section.

The scenarios for this showcase are as follows:

II.I Mediating role of BMI (CHUV)
II.II Glycaemic Effects Anti-depressant Medicines (CING)

II.III Titration Observation (ZEINCRO)
II.IV CLOpidogrel Clinical Knowledge (ZEINCRO)
II.V Does Simvastatin treatment enhance prevention (ZIENCRO)

4.5.4 Scenarios for Showcase 2

Scenario II.I: Mediating role of BMI (CHUV)

Scenario Name	Association between major depressive disorder (MDD) and type II diabetes: the mediating role of BMI
Template	See Section A.2 Scenarios for Showcase 2

Scenario II.II: Glycaemic Effects Anti-depressant Medicines (CING)

Scenario Name	Glycaemic Effects of anti-depressant medicines (ADMs)
Template	See Section A.2 Scenarios for Showcase 2

Scenario II.III: Titration Observation (ZEINCRO)

Scenario Name	Multicenter, non-interventional clinical study of titration observance of carvedilol in patients with hypertension and diabetes.
Template	See Section A.2 Scenarios for Showcase 2

Scenario II.IV: CLOpidogrel Clinical Knowledge (ZEINCRO)

Scenario Name	CLOpidogrel Clinical Knowledge
Template	See Section A.2 Scenarios for Showcase 2

Scenario II.V: Does Simvastatin treatment enhance prevention (ZEINCRO)

Scenario Name	Does Simvastatin treatment enhance prevention?
Template	See Section A.2 Scenarios for Showcase 2

4.5.5 Scenarios of Showcase III: Identification of relations between molecular fragments and specific adverse side effect categories

Context: Identification of relations between molecular fragments and specific adverse side effect categories

The purpose of this showcase is to demonstrate how innovative analytical techniques could be applied independently of Linked2Safety to generate results that are then utilized within Linked2Safety to test hypothesis that could prove to be critical in discovering novel adverse events to existing drugs on the market. For the implementation of this task, the initial analyses will be performed on Lisis, a platform developed as part of another FP7 ICT project GRANATUM [5]. Lisis tools will enable researchers to perform clustering of drugs based on their molecular fragments.

Specifically a set of clustering tools will be implemented providing functionality for hierarchical and partitional clustering. These tools will be integrated into the Lisis platform. Implementations of clustering algorithms will be taken from R statistical package [6]. Chemical similarity measures will be used for the distance metrics of the clustering algorithms.

The tools are going to be used for clustering sets of drugs. Furthermore the results of clustering are going to be used for data analysis of "Trait of Interest" among the drugs of each cluster. This analysis will provide valuable information about the biological effects of the drugs in correlation to their chemical/structural information.

The outcome of the analyses will be clusters of drugs, which will then be passed to the Linked2Safety platform and tested through both traditional statistical analyses techniques and data mining approaches for the association to adverse events.

4.5.6 Scenarios for Showcase III

The scenarios for this showcase are as follows:

III.I Anti-depressants and Cardio Issues (CHUV)
III.II Asthma: ADEs and Formoterol (ZEINCRO)
III.III Depression: Anti-depressant Medicine and Glycaemic Effects (CING)
III.IV Hypercholestrerolemia: ADEs and Simvastatin (ZEINCRO)

Scenario III.I: Anti-depressants and Cardio Issues (CHUV)

Scenario Name	Associations between anti-depressant medicine (ADM) and cardio-vascular risk factors (CVRFs) and cardio-vascular diseases (CVDs)
Template	See Section A.3 Scenarios for Showcase 3

Scenario III.II: Asthma: ADEs and Formoterol (ZEINCRO)

Scenario Name	Associations between adverse events and Formoterol in patients with asthma.
Template	See Section A.3 Scenarios for Showcase 3

Scenario III.III: Depression: Anti-depressant Medicine and Glycaemic Effects (CING)

Scenario Name	Associations between anti-depressant medicine (ADMs) and glycaemic effects in patients with depression
Template	See Section A.3 Scenarios for Showcase 3

Scenario III.IV: Hypercholestrerolemia: ADEs and Simvastatin (ZEINCRO)

Scenario Name	Associations between adverse events and simvastatin in patients with hypercholestrerolemia.
Template	See Section A.3 Scenarios for Showcase 3

5 Detailed Organisation and Planning

"The detailed organization and planning of the pragmatic clinical trials showcases that are going to be piloted in order to prove the applicability, the usability and the effectiveness of the Linked2Safety enabling technologies."

This section presents the detailed organization of the Clinical Trials Showcases and Performance Evaluation for the Linked2Safety. It is based upon the following main activities:

- Pre-verification (detailed and systematic) of the deployment of the Linked2Safety platform and showcase execution. The purpose of this activity is to confirm feasibility and hence ensure successful deployment, and execution and evaluation.
- Interoperable EHRs Deployment (Task 7.3): This activity involves the application of the first release of the tools and services of the Interoperable EHR Data Space (WP3) to align, transform and enrich the specific EHR and EDC repositories existing in each clinical research end-user organization participating in the Linked2Safety project. This will establish the semantically interoperable EHR-related information resources space for the implementation of the identified clinical trials showcases.
- Linked2Safety Platform Deployment and Configuration (Task 7.4): This activity will deploy the integrated Linked2Safety platform (WP6) for serving as the pilot environment for the Linked2Safety clinical research showcases. This will facilitate the semantic interoperation, access and reuse of medical datasets (from spatially distributed EHRs and EDCs repositories), for advancing proactive patients' safety in clinical trials.
- Clinical Trials Showcases Execution and Evolution (Task 7.5 and Task 7.6): In this activity, the clinical research organizations will proceed with the execution of the identified clinical trials showcases in order to evaluate their performance. A scenario-based evaluation methodology is to be utilised. If necessary, the additional activities such as workshops will be organized for potential users to assess and evaluate the Linked2Safety platform.

5.1 Pre-verification of Deployment and Showcase Execution

The pre-verification (detailed and systematic) of the deployment of the Linked2Safety platform and showcase execution is to confirm feasibility and hence ensure successful deployment, and execution and evaluation. It involves the following tasks and the responsibilities:

- Confirmation of data availability: the data providers check and confirm the data availabilities for the deployment and showcase execution. This is reported in Section 5.1.1, Tables 3-7.

- Detailed verification to allow successful execution and evaluation: for this pre-execution and after deployment task, the data providers, with the support of the technical partners, will test and monitor the performance of the Linked2Safety showcases' experimental environments and update the evaluation strategies based on deployment experiences. An associated verification list is given in Section 5.1.2.

5.1.1 Showcase data confirmation

The following tables (3-7) indicate for each showcase and date provider the data that is expected to be applicable and therefore input into the Linked2Safety platform. The focus in this section is given to the phenotypic data that vary both in terms of their definition and in the data collection methodology. With respect to genetic data, only two of the partners are expected to provide data, CHUV and CING. All of CHUV's data is based on the Affymetrix 500k genotyping platform that genotypes a total of about 500,000 markers across the whole genome. CING data however includes candidate gene studies, where only specific regions of the genome are genotyped and tested for their association to a disease (a few dozen markers per study). CING also has multiple datasets, each targeting a different disease, and therefore different candidate genes. Both types of studies however rely on the same type of genetic markers: single nucleotide polymorphisms (SNPs). To maximize the positive impact of the Linked2Safety approach, priority will be given to analyzing the genetic markers that CING has collected, which overlap with the genetic markers in CHUV.

Assessed By	Showcase	Description	Data Provider		
			CHUV	CING	ZEINCRO
CHUV	Type 1 (I.I)	MDD and hypertension and antidepressant drugs	Demographic characteristics (age, sex)		Demographic characteristics (age, sex)
			MDD Diagnosis		MDD Diagnosis: very little (mostly medical history)
			Hypertension / blood pressure	Hypertension / blood pressure	Hypertension / blood pressure
			Antidepressant drugs		
CHUV	Type 2 (II.I)	Association between MDD and type II diabetes: the mediating role of BMI	BMI/weight	BMI/weight	BMI/weight
			type II diabetes	type II diabetes	type II diabetes
			MDD		MDD Diagnosis: very little (mostly medical history)
CHUV	Type 3 (III.I)	Associations between ADM and CVRFs and CVDs	Myocardial infarcation		
			BMI/weight	BMI	BMI/weight
			type II diabetes	type II diabetes	type II diabetes
			Hypertension / blood pressure	Hypertension	
			lifetime use of ADM		MDDs: very little (mostly medical history)

Table 3 – Scenarios to be overseen by CHUV

Assessed By	Showcase	Description	Data Provider		
			CHUV	**CING**	**ZEINCRO**
CING	Type 1 (I.II)	Effect of anti-hypertensive drugs on CVD in type II diabetes	Diabetese type II (self reported)	Diabetes type II daignosis or not	*Diabetes type II diagnosis or not*
			Hypertension	Hypertension diagnosis or not	*Hypertension diagnosis or not*
			Intake of antihypertensive drugs (yes/no self reported) Providing enough participants take the drug.	Info on anti-hypertensive drugs (Self reported)	*Info on anti-hypertensive drugs: Dosage and treatment duration (Documented)*
CING	Type 2 (II.II)	Glycaemic Effects of ADMs	Drug use in patients with depression		
			Physical activity and metabolic syndrome (eg blood glucose levels)		
CING	Type 3 (III.III)	Association between ADMS and glycaemic effects in patients with depression	Glycaemic side effects		Blood glucose (if drug use info also available on same subjects)
			drug use (including antidepressants)		drug use

Table 4 – Scenarios to be overseen by CING

Assessed By	Showcase	Description	Data Provider		
			CHUV	CING	ZEINCRO
ZEINCRO	Type 1 (I.III)	Efficiacy and safety of the Formoterol DPI (A) vs innovative Formoterol DPI (B) in patients with asthma			Drug use in patients with asthma in relation to the efficacy of Formoterol DPI and the presence of adverse drug reactions
ZEINCRO	Type 1 (I.IV)	Compare efficacy and safety of combination of fluticasone / Salmetorol (device A) with device B in patients with asthma			Drug use in patients with asthma in relation to the efficacy of fluticasone / Salmeterol in relation ot adverse drug reactions.

Table 5 – Scenarios to be assessed by ZEINCRO (1)

Assessed By	Showcase	Description	Data Provider		
			CHUV	CING	ZEINCRO
ZEINCRO	Type 2 (II.III)	Titration observance of carvedilol in patients with hypertension and diabetes	*hypertension*	*hypertension*	hypertension
			diabetes	*diabetes*	diabetes
			Intake of antihypertensive drugs (yes/no self reported) Only a few who took Carvedilol but could extend research to similar treatments. Providing enough participants take the drug.		titration of Carvedilol
			adverse drug reaction		adverse drug reaction
ZEINCRO	Type 2 (II.IV)	CLOpidogrel clinical knowledge	*Drug use in patients treated with Clopidogrel (Yes/No Self reported). Providing enough subjects take the drug.*		Drug use in patients treated with Clopidogrel.
			demographics		demographics
					medical history
			clopidogrel therapeutic schema		clopidogrel therapeutic schema
					lab tests
			adverse drug reactions		adverse drug reactions
ZEINCRO	Type 2 (II.V)	Does Simvastatin treatment enhance prevention	*Cardiovascular system related drugs (yes/no self reported). Only a few subjects who took Simvastatin) Could provide subjects who take similar drugs.)*		Drug use in patients who are treated with Simvastatin.
			demographics		demographics
					medical history
			adverse drug reactions		adverse drug reactions

Table 6 – Scenarios to be assessed by ZEINCRO (2)

Assessed By	Showcase	Description	Data Provider		
			CHUV	CING	ZEINCRO
ZEINCRO	Type 3 (III.II)	Associations between adverse events and Formoterol in patients with asthma			Drug use in patients with asthma who are treated with Formoterol in relation to presence of adverse drug reactions.
ZEINCRO	Type 3 (III.IV)	Associations between adverse events and simvastatin in patients with hypercholesterolemia	*Level of cholesterol and intake of cardiovascular system related drugs (yes/no self reported) A few who took Simvastatin: Could extend to other similar treatment*		hypercholesterolemia patient drug use
					Compliance
			Adverse Drug reaction		Adverse drug reactions

Table 7 – Scenarios to be assessed by ZEINCRO (3)

5.1.2 Verification of successful showcase execution and evaluation

For each showcase, the check list for verification includes:

1. The interoperable EHRs Deployment was successful:
   - Data cubes have been successfully generated
   - The generated data cubes have been semantically enriched
2. The Linked Medical Data Space deployment has been successful
   - The required data linkage has been successfully achieved
3. The Data Analysis Space has been successfully deployed
   - The required data analysis methods and tools are ready to be executed and capable of providing the analysis results.
4. The Linked2Safety platform has been successfully configured.
5. Check all outcomes related to the identified risk(s) and check whether the showcases are ready to be further executed.
6. Check whether the evaluation strategies need updating based on the above deployment process.

5.2 Detailed Interoperable EHRs Deployment and Timeline

The Interoperable EHRs Deployment involves the application of the first release of the tools and services of the Interoperable EHR Data Space (WP3) (Figure 3) to align, transform and enrich the specific EHR and EDC repositories existing in each

clinical research end user organization participating in the Linked2Safety project. This will establish the semantically interoperable EHR-related information resources space for the implementation of the identified clinical trials showcases.

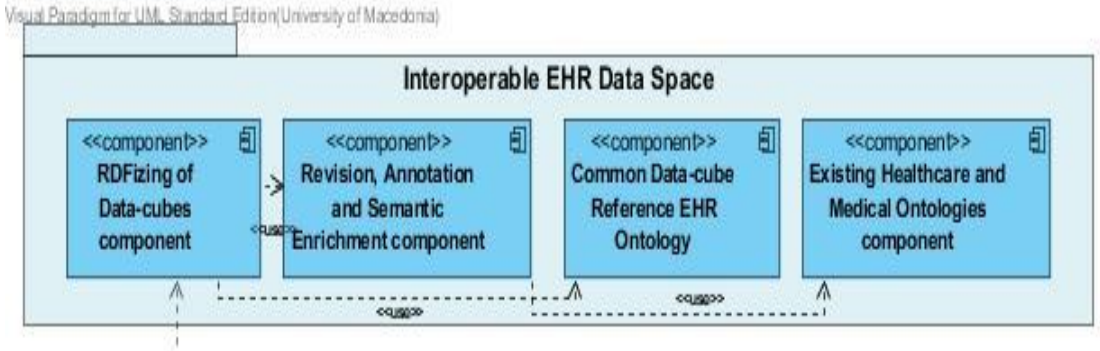

Figure 3 – Interoperable EHR Data Space

Months 22 – 24 (Weeks 1 – 12)

Task 7.3.1 *Clinical Partners*: Host deployment of the Interoperable EHR Prototype.

Task 7.3.2 *SIVECO*: Package binaries and providing deployment instructions for the clinical partners.

Task 7.3.3 *UBITECH*: Provide technical support for issues that arise during deployment.

Task 7.3.4 *LUH*: Overview the deployment of software tools from a legal perspective.

Months 25 – 27 (Weeks 13 – 24)

Task 7.3.5 *Clinical Partners*: Provision of data as specified in the clinical trials showcases.

Task 7.3.6 *Clinical Partners*: Assist with all work requiring domain expertise, by SIVECO and UBITECH regarding alignment, transformation and enrichment of the specific EHR and EDC repositories existing in each clinical research and end user organisation.

Task 7.3.7 *SIVECO* and *UBITECH*: Alignment, transformation and enrichment of the specific EHR and EDC repositories existing in each clinical research and end user organisation.

Task 7.3.8 *SIVECO* and *UBITECH*: Establish the semantically interoperable EHR-related information resources space for the implementation of the clinical trials showcases identified in Task 7.1.

Task 7.3.9 *SIVECO* and *UBITECH*: Provide technical support for issues that arise during usage.

Task 7.3.10 *SIVECO* and *UBITECH*: Implement required change requests and fixing bugs in the software.

Months 28 – 30 (Weeks 25 - 36)

Task 7.3.11 *SIVECO* and *UBITECH*: Assist with Task 7.4 (Platform deployment and configuration), for tools / components relating to the "Interoperable EHR Data Space", responsible for the RDFization of a data cube and its semantic enrichment.

Task 7.3.12 *SIVECO* and *UBITECH*: Complete relevant questionnaires and/or interviews, which will contribute to Task 7.6: Performance evaluation and lessons learnt.

5.3 Detailed Platform Deployment and Configuration Timeline

Platform Deployment and Configuration will deploy the integrated Linked2Safety platform (WP6) for serving as the pilot environment for the clinical research showcases. This will facilitate the semantic interoperation, access and reuse of medical datasets (from spatially distributed EHRs and EDCs repositories), for advancing proactive patients' safety in clinical trials.

The platform deployment and configuration will be achieved in four phases, and will integrate with Tasks 7.5 and 7.6. The Linked2Safety platform architecture comprises of four "spaces" that contain the basic software components, each space being required for the showcase scenarios:

a) The "Data Cube Generation Space" (part of Workpackage 3), where software components are responsible for producing non-identifiable data cubes from the EHR records provided;

b) The "Interoperable EHR Data Space" (part of Workpackage 3) is responsible for the RDFization of a data cube and its semantic enrichment;

c) The "Linked Medical Data space" (Workpackage 4) is responsible for the publication and the federation of data request queries;

d) The "Genetic Data Analysis Space" (Workpackage 5) is responsible for the software components that perform the analysis of data provided by the "Linked Medical Data Space".

The four phases involve the testing of the tools and components that are to be used for each showcase type, using the data that has been input during Task 7.3.

In Deliverable 6.1, the following tests were presented (Figure 4):

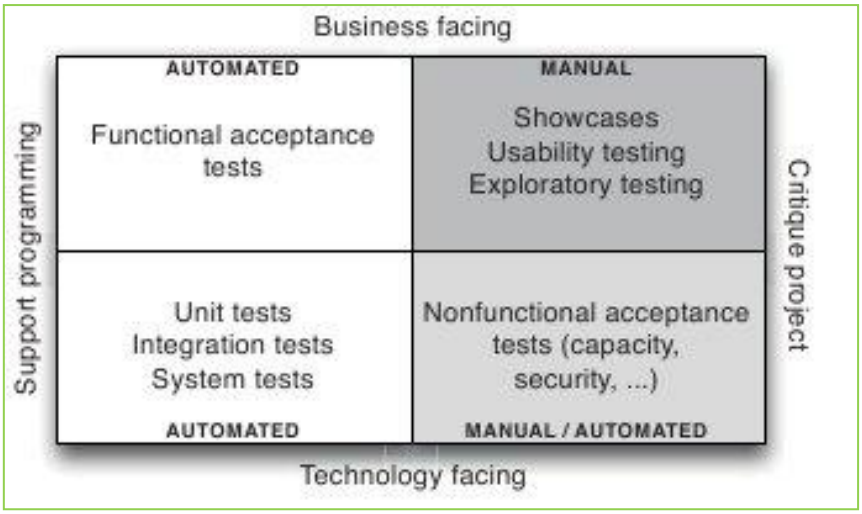

Figure 4 – Testing types facets

The testing to be performed during months 28-30 involve the 'technology facing' tests relating specifically to the showcase scenarios, to ensure that the showcase execution can begin without delay, in month 31.

Weeks 1-3

Task 7.4.1 *INTRASOFT*: Familiarisation with the data to be used for the showcases, and the tools required to perform each one, as specified in this deliverable.

Task 7.4.2 *INTRASOFT*: Check that the required showcase data has been added to the Linked2Safety data stores in Task 7.3 and is fully accessible from the Linked2Safety platform.

Weeks 4-6

Task 7.4.3 *INTRASOFT* / *UBITECH*: Deploy and test all platform tools/components to be used in the type I (Phase II - Subject Selection Criteria) showcase scenarios, using the agreed scenario data.

Weeks 7-9

Task 7.4.4 *INTRASOFT* / *UBITECH*: Deploy and test all platform tools/components to be used in the type II (Phase IV – Post Marketing Surveillance trials) showcase scenarios.

Weeks 10-12

Task 7.4.5 *INTRASOFT* / *UBITECH*: Deploy and test all platform tools/components to be used in the type III (Chemoinformatics: Identification of relations between Molecular Fragments and Specific Adverse Side Effects) showcase scenarios.

Weeks 13-24

Task 7.4.6 *INTRASOFT* / *UBITECH*: Assist with Task 7.5 showcase execution, which includes:

1. The 'business facing' type testing illustrated in Figure 4.
2. Making necessary updates to the Linked2Safety platform to accommodate any faults / problems highlighted during showcase execution.

Weeks 22-24

Task 7.4.7 *INTRASOFT* / *UBITECH*: Test platform/documentation updates relating to the previous Task 7.6 evaluation phases.

Task 7.4.8 INTRASOFT / UBITECH: Assist with identification/generation of lessons learnt.

5.4 Detailed Showcase workflows, execution and evaluation timelines

5.4.1 Detailed workflow of Showcase Execution

This section describes the use cases and expected interface mock-ups that must be in place for successful execution of the scenarios. For each showcase type, the use cases designed in Deliverable D1.2 are displayed along with user interface mock-ups. This is followed by the "TO BE" workflows as described in the scenario templates.

5.4.1.1 Showcase I: Subject Selection

As outlined in Deliverable D1.2, *'this use case describes the clinical trial coordinator's ability to search subjects for a new clinical trial based on specific search criteria. The users can view the number of subjects meeting the search criteria as well as additional information about the recruitment sites/hospitals. Moreover, the users can receive recommendations on how to improve the search results as well as view related associations.'*

Table 8 displays this use case and Figure 5 shows a mock-up of the user interface to be used. This is followed by the detailed workflows relating to each type 1 showcase.

Name	Search for Subjects
Summary	The actor wants to find subjects to enrol in a new clinical trial.
Primary Actor(s)	Clinical Trial Coordinator
Secondary Actor(s)	

Referred Requirements	F1, F2, F13, F15, F16
Results(s)	The number of subjects that match the actor's search criteria is returned. Moreover, the actor has the ability to view the specific number of subjects provided by each recruitment site/hospital as well as details related to the specific hospital/site.
Normal Flow:	
Step	**Action**
1	The actor initiates the "search for subjects" procedure. The "search for subjects" user interface is displayed (Figure 5).
2	The Linked2Safety platform provides the actor with a form to fill the search criteria (inclusive/exclusive criteria). Advanced search is also available.
3	The actor specifies the desired values for specific attributes in the search form.
4	A SPARQL query is formed and is sent to the Linked2Safety repositories. These repositories are set of data cubes provided by individual clinical sites.
5	The number of matching subjects obtained by each recruitment site/hospital (stated in bandwiths of 10, e.g. '40-50') is returned and presented to the actor. The ability to view additional details of recruitment sites/hospitals is available. The ability to receive recommendations on how to improve the search results (get more subjects) is also available. The ability to view rules/associations related to his/her search criteria is also available.
6	The Linked2Safety platform asks the actor if s/he wants to store the query and the search results for future reference.
7	The actor agrees. The query and the search results are stored in the Linked2Safety platform.
Alternative Flows:	
2a	(Alternative to 2: Advanced Search) The actor can select the sources from which subjects derive (e.g. the desired recruitments sites/hospitals) and/or select sub-datasets (e.g., specific medical problem like Diabetes or Breast Cancer) for improved search results.
5a	(Alternative to 5: Additional information) The actor can view details about the recruitment site/hospital (additional information, their contact details, sites/ hospitals'

	system reports so that problematic recruitment issues related to sites/ hospitals may be identified).
5b	(Alternative to 5: Receive recommendations) The actor can view statistics about the same or similar search criteria acquired by previous usage.
5c	(Alternative to 5: Recorded associations) The actor can view associations (detected AEs) referring to his/her inclusion/exclusion criteria in order to have a better knowledge for the design of his/her clinical protocol.
Exceptional Flows:	
5d	(Exceptional to 5) The Linked2Safety platform displays a message describing no matching data can be found and the process continues from step 2.
7a	(Exceptional to 7) The actor does not wish the query and the search results to be stored in the Linked2Safety platform. The search process is terminated.

Table 8 - Subject Selection Use Case (from D1.2)

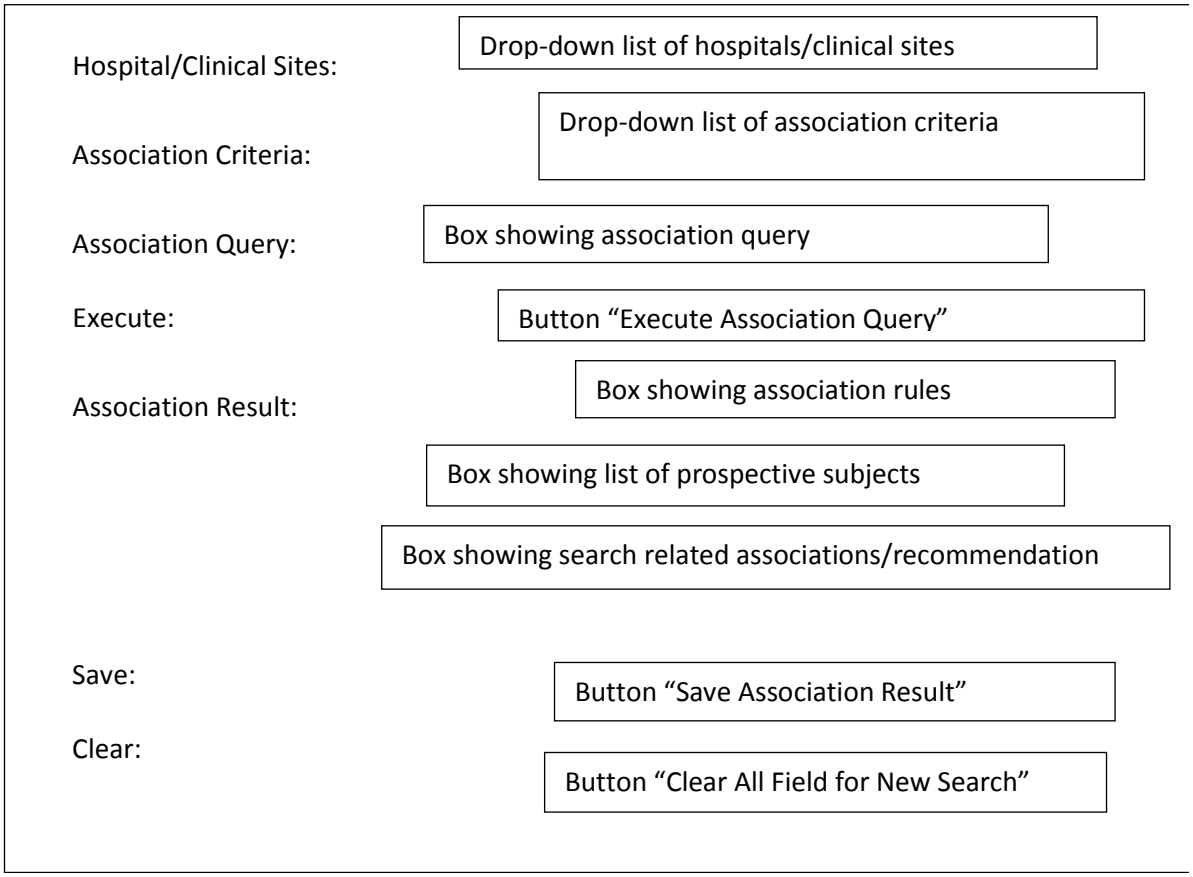

Figure 5 - User interface mock-up for subject selection

Scenario I.I: Anti-Depressant Drugs Evaluation (CHUV) – Detailed Workflow

1. Select Hospital/Clinical sites of interest. In this showcase, this would be all clinical provider sites and the expected results would be from CHUV and ZEINCRO, although it is unsure at this stage if much information on whether subjects have MDD is available from ZEINCRO. Also, no information on anti-depressant drug use is available.
2. Select attributes of interest. In this scenario the attributes are:
   1 MDD
   2 Blood pressure
   3 Hypertension
   4 Antidepressant drugs
3. Enter the association query details.
4. Click on the 'Execute Association Query' button
5. Obtain a breakdown of the frequency count (in bandwiths of 10, e.g. '40-50') by institutions and by the combination of available information
6. Request names and contact numbers of the different institutions with subjects that meet inclusion criteria to obtain the additional information.
7. Save the query and search results for future use.

Scenario I.II: Effect of Anti-hypertensive Drugs (CING) – Detailed Workflow

1. Select Hospital/Clinical sites of interest. In this showcase, this would be all clinical provider sites and the expected results would be from CHUV, CING and

ZEINCRO, although it is unsure at this stage if enough patients who take anti-hypertensive drugs are available from CHUV.

2. Select attributes of interest. In this scenario the attributes are:
   i. Diabetes type II diagnosis
   ii. Hypertension
   iii. Anti-hypertensive drug use.
3. Enter the association query details.
4. Click on the 'Execute Association Query' button
5. Obtain a breakdown of the frequency count (in bandwiths of 10, e.g. '40-50') by institutions and by the combination of available information.
6. Request names and contact numbers of the different institutions with subjects that meet inclusion criteria to obtain the additional information.
7. View any detected adverse events relating to the search criteria.
8. Save the query and search results for future use.

Scenario I.III Asthma – Fluticasone / Salmeterol: Device A versus Device B (ZEINCRO) – Detailed Workflow

1. Select Hospital/Clinical sites of interest. In this showcase, this would be all clinical provider sites and the expected results would be from ZEINCRO.
2. Select attributes of interest. In this scenario the attributes are:
   i. Asthma diagnosis
   ii. Drug use
   iii. Use of Fluticasone and/or Salmeterol
3. Enter the association query details.
4. Click on the 'Execute Association Query' button
5. Obtain a breakdown of the frequency count (in bandwiths of 10, e.g. '40-50') by institutions and by the combination of available information
6. View the results of a chi square test
7. Request names and contact numbers of the different institutions with subjects that meet inclusion criteria to obtain the additional information.
8. View any detected adverse events relating to the search criteria.
9. Save the query and search results for future use.

Scenario I.IV: Asthma: Formoterol DPI (A) versus Formoterol DPI (B) (ZEINCRO) – Detailed Workflow

1. Select Hospital/Clinical sites of interest. In this showcase, this would be all clinical provider sites and the expected results would be from ZEINCRO.
2. Select attributes of interest. In this scenario the attributes are:
   i. Asthma diagnosis
   ii. Use of Formoterol
3. Enter the association query details.
4. Click on the 'Execute Association Query' button
5. Obtain a breakdown of the frequency count (in bandwiths of 10, e.g. '40-50') by institutions and by the combination of available information
6. View the results of a chi square test
7. Request names and contact numbers of the different institutions with subjects that meet inclusion criteria to obtain the additional information.
8. View any detected adverse events relating to the search criteria.

9. Save the query and search results for future use.

5.4.1.2 Showcase II: Phase IV Clinical Trial: Post Marketing Surveillance

As outlined in Deliverable D1.1, 'A Phase IV study is a clinical trial, a quasi-experimental study, or an observational study to gather specific information about an approved drug, a biological product, a device, or a procedure. Post-approval research is typically initiated to better understand product use in real-world situations, to obtain evidence for higher reimbursement or submission for expanded labelling, to fulfil a specific requirement of regulatory authorities, or to monitor safety of a drug or device in a larger, non-clinical trial setting.' The use case for this scenario type is hypothesis testing, which is already executed by the clinical partners, but Linked2Safety will automate this process.

Table 9 displays this use case and Figures 6 and 7 show mock-ups of the user interfaces to be used. This is followed by the detailed workflows relating to each type 2 showcase.

Name	Perform Data Analysis – Hypothesis Testing
Summary	This use case describes the process of a hypothesis testing.
Primary Actor(s)	Clinical Researcher
Secondary Actor(s)	
Results(s)	The hypothesis testing computes the observed value of T from the samples and returns to the actor the outcome of the test (based on the observed value)
Normal Flow (Hypothesis Testing):	
Step	**Action**
1	The actor initiates the data analysis procedure.
2	The Linked2Safety platform provides the actor with four options: (i) to test a hypothesis, (ii) to perform data mining, (iii) to manually select association rules, and (iv) to find associations that can be used in a chemoinformatic analysis.
3.1	The actor selects option (i), to test a hypothesis on the Linked2Safety platform and the hypothesis testing user interface appears (Figure 6).

4.1	The actor defines the null hypothesis and the alternative hypothesis. A null hypothesis is a simple hypothesis associated with a contradiction to a theory which one would like to prove. An alternative hypothesis is a hypothesis associated with a theory which one would like to prove.
5.1	The actor defines the subject selection criteria by invoking the "Search for Subjects" use case on the Linked2Safety platform (Figure 6).
6.1	The Linked2Safety platform selects subjects (samples) according to the subject selection criteria (result of "Search for Subjects" use case).
7.1	The actor selects a test statistic T (e.g. chi-square) from a dropdown list of test statistics available on the Linked2Safety platform (Figure 6).
8.1	The actor selects the distribution of the test statistic using the Linked2Safety platform (Figure 6).
9.1	The Linked2Safety platform provides the actor with two options: (a) partition the possible values of test statistic T using the distribution of T and (b) set a significant level as threshold below which the null hypothesis is rejected.
10.1	The actor selects option (a).
11.1	The distribution of the test statistic partitions the possible values of T into those for which the null-hypothesis is rejected, called critical region, and those for which it is not rejected. Moreover, the Linked2Safety platform computes the observed value of T from the samples and displays it to the actor in the "Text field to display the value of test statistic".
12.1	The Linked2Safety platform returns the outcome of the test based on the observed value as follows (Figure 6); reject the null hypothesis in favour of the alternative hypothesis if the observed value is in the critical region and accept the null hypothesis otherwise. Moreover, the Linked2Safety platform asks the actor if s/he wants to save the hypothesis test outcome.
13.1	The actor selects to save the outcome of hypothesis testing into the Linked2Safety platform. The outcome is stored in the Knowledge Base.
Alternative Flows (Hypothesis Testing):	
10.1a	The actor selects option (b), to set a threshold below which the null hypothesis is rejected.
11.1a	The Linked2Safety platform pops up a user interface enabling the actor to select a significance level as a probability threshold below which the null hypothesis is rejected (Figure 7). Common values are 5% and 1%.
12.1a	The actor sets a significant level as threshold and starts the hypothesis testing.

13.1a	The Linked2Safety platform computes from the selected subjects the observed value of the test statistic T. Moreover, from the statistic it calculates a probability of the observation under the null hypothesis (p-value[1]).
14.1a	The Linked2Safety platform returns the outcome of the test based on the observed value as follows (Figure 6); reject the null hypothesis in favour of the alternative hypothesis if and only if the p-value is less than the significance level (the selected probability) threshold and accept the null hypothesis otherwise. Moreover, the Linked2Safety platform asks the actor if s/he wants to save the hypothesis test outcome.
15.1a	The actor selects to save the outcome of hypothesis testing into the Linked2Safety platform. The outcome is stored in the Knowledge Base.
Exceptional Flows (Hypothesis Testing):	
11.1a	(Exceptional to 11.1) The selected subjects are not appropriate for testing this hypothesis because these subjects do not contain all the relevant information. For example, a hypothesis is "drug A is likely to be associated with diabetes", but the selected subjects do not contain information on drug A. The Linked2Safety platform displays an error message describing the failure. The hypothesis testing is cancelled and the process continues from step 4.1.
11.1b	(Exceptional to 11.1) The hypothesis test outcome is invalid because assumptions are invalid. The process is terminated. The Linked2Safety platform displays an error message describing the failure. The hypothesis testing is cancelled and the process continues from step 4.1.
13.1b	(Exceptional to 13.1 and 15.1a) The actor does not wish to save the outcome of hypothesis testing into the Linked2Safety platform. The process is terminated.

Table 9 – Perform Hypothesis Testing Use Case (from D1.2)

[1] P-value is the probability assuming the null hypothesis is true, of observing a result at least as extreme as the test statistic.

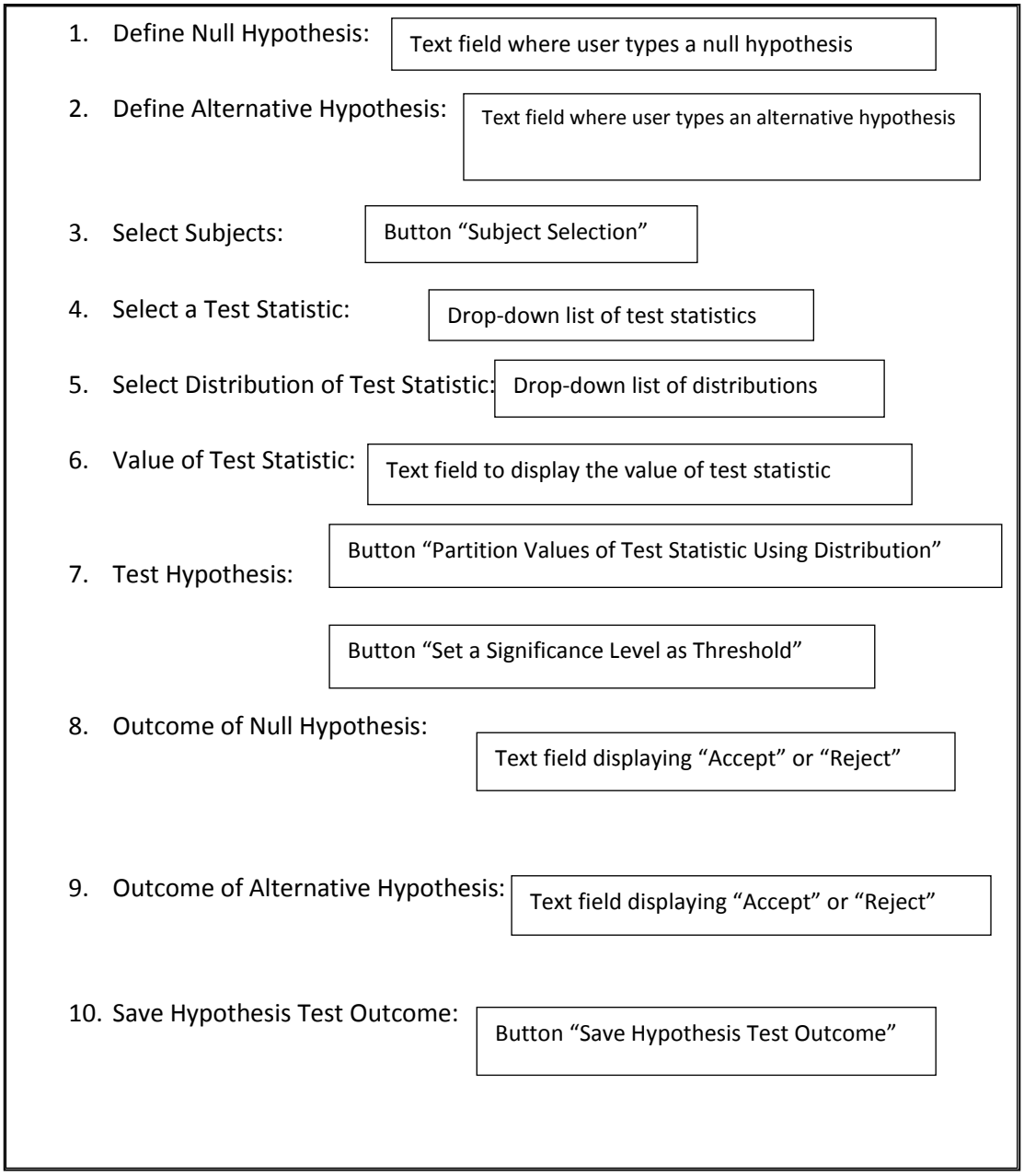

Figure 6 – User interface mock-up to perform hypothesis testing

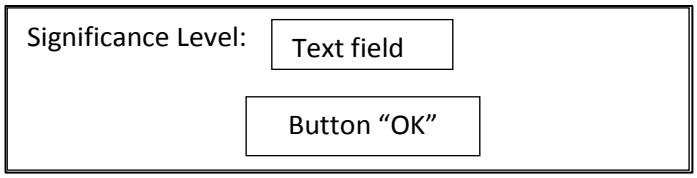

Figure 7 – User interface mock-up for setting a significance level as threshold

5.4.1.3 Scenario II.I Mediating role of BMI (CHUV) - Detailed Workflow

1. Define the null hypothesis and the alternative hypothesis. A null hypothesis is a simple hypothesis associated with a contradiction to a theory which one would like to prove. An alternative hypothesis is a hypothesis associated with a theory which one would like to prove.
2. Define the subject selection criteria by invoking the "Search for Subjects" use case on the Linked2Safety platform.
    a. Request the frequency count of subjects with available information on diabetes, BMI and MDD. It is expected that subjects will be available from CHUV, with a possibility of some data from ZEINCRO.
    b. Determine whether there are enough subjects from the available data sources to test the study hypothesis.
    c. Request a breakdown of the frequency table to show which institutions and where, have how many eligible subjects. Determine whether these data sources are compatible (for e.g. determine possible sources of incompatibility or bias regarding the type of sample (e.g. general population sample versus clinical sample), subject selection (e.g. recruitment of particularly severe cases of depression in hospital settings versus inpatient settings, etc.).
3. Select a test statistic T (e.g. chi-square) from a dropdown list of test statistics available on the Linked2Safety platform (Figure 6).
4. Select the distribution of the test statistic using the Linked2Safety platform.
5. The Linked2Safety platform provides two options: (a) partition the possible values of test statistic T using the distribution of T and (b) set a significant level as threshold below which the null hypothesis is rejected.
6. Select option (a).
7. The distribution of the test statistic partitions the possible values of T into those for which the null-hypothesis is rejected, called critical region, and those for which it is not rejected. Moreover, the Linked2Safety platform computes the observed value of T from the samples and displays it in the "Text field to display the value of test statistic".
8. The Linked2Safety platform returns the outcome of the test based on the observed value as follows (Figure 6); reject the null hypothesis in favour of the alternative hypothesis if the observed value is in the critical region and accept the null hypothesis otherwise. Moreover, the Linked2Safety platform asks if the user wants to save the hypothesis test outcome.
9. Select to save the outcome of hypothesis testing into the Linked2Safety platform. The outcome is stored in the Knowledge Base.

10. Follow alternative flows that can be tested in this showcase (see Table 9):

Scenario II.II: Glycaemic Effects Anti-depressant Medicines (CING) - Detailed Workflow

1. Define the null hypothesis and the alternative hypothesis. A null hypothesis is a simple hypothesis associated with a contradiction to a theory which one would like to prove. An alternative hypothesis is a hypothesis associated with a theory which one would like to prove.
2. Define the subject selection criteria by invoking the "Search for Subjects" use case on the Linked2Safety platform.
    - Request the frequency count of subjects diagnosed with depression who also have drug use, physical activity, and glycaemic index information available. It is expected that subjects will be available from CHUV, with a possibility of some data from ZEINCRO.
    - Determine whether the available sample sizes are enough for the statistical analysis to have enough power.
3. Select a test statistic T (e.g. chi-square) from a dropdown list of test statistics available on the Linked2Safety platform (Figure 6).
4. Obtain chi square test of association to determine whether the observed numbers of subjects with hyperglycaemia in patients prescribed TCAs and patients prescribed SSRIs differ from the expected numbers. This test does not take into account physical activity.
5. Select the distribution of the test statistic using the Linked2Safety platform.
6. The Linked2Safety platform provides two options: (a) partition the possible values of test statistic T using the distribution of T and (b) set a significant level as threshold below which the null hypothesis is rejected. (Figure 6)
7. Select option (a).
8. In order to correct for the effect of physical activity, logistic regression is carried out. The subjects will be divided into those who have developed hyperglycaemia and those who have not. Type of drug prescribed and physical activity will be entered into the model as exposure variables. If type of drug use (TCA vs. SSRI (baseline)) increases the risk of hyperglycaemia (significant OR above 1) while correcting for physical activity, we can then conclude that TCAs do have glycaemic side effects and can predispose to MetS and/or diabetes type II.
9. The Linked2Safety platform returns the outcome of the test based on the observed value as follows: reject the null hypothesis in favour of the alternative hypothesis if the observed value is in the critical region and accept the null hypothesis otherwise. Moreover, the Linked2Safety platform asks if the user wants to save the hypothesis test outcome.
10. Select to save the outcome of hypothesis testing into the Linked2Safety platform. The outcome is stored in the Knowledge Base.
11. Follow alternative flows that can be tested in this showcase (see Table 9):

Scenario II.III Titration Observation (ZEINCRO) - Detailed Workflow

1. Define the null hypothesis and the alternative hypothesis. A null hypothesis is a simple hypothesis associated with a contradiction to a theory which one would like to prove. An alternative hypothesis is a hypothesis associated with a theory which one would like to prove.
2. Define the subject selection criteria by invoking the "Search for Subjects" use case on the Linked2Safety platform.
    a. Request the frequency count of subjects with available information on drug use in patients with hypertension and diabetes who took Carvedilol. It is expected that subjects will be available from ZEINCRO, with a possibility of some data from CHUV if enough patients took this drug or the study is extended to similar treatments.
    b. Determine whether there are enough subjects from the available data sources to test the study hypothesis.
    c. Request a breakdown of the frequency table to show which institutions and where, have how many eligible subjects. Determine whether these data sources are compatible (for e.g. determine possible sources of incompatibility or bias regarding the type of sample (e.g. general population sample versus clinical sample), subject selection (e.g. recruitment of particularly severe cases of depression in hospital settings versus inpatient settings, etc.).
3. Select chi-square from a dropdown list of test statistics available on the Linked2Safety platform (Figure 6).
4. Select the distribution of the test statistic using the Linked2Safety platform.
5. The Linked2Safety platform provides two options: (a) partition the possible values of test statistic T using the distribution of T and (b) set a significant level as threshold below which the null hypothesis is rejected.
6. Select option (a).
7. The distribution of the test statistic partitions the possible values of T into those for which the null-hypothesis is rejected, called critical region, and those for which it is not rejected. Moreover, the Linked2Safety platform computes the observed value of T from the samples and displays it in the "Text field to display the value of test statistic".
8. The Linked2Safety platform returns the outcome of the test based on the observed value as follows (Figure 6); reject the null hypothesis in favour of the alternative hypothesis if the observed value is in the critical region and accept the null hypothesis otherwise. Moreover, the Linked2Safety platform asks if the user wants to save the hypothesis test outcome.
9. Select to save the outcome of hypothesis testing into the Linked2Safety platform. The outcome is stored in the Knowledge Base.
10. Follow alternative flows that can be tested in this showcase (see Table 9):

Scenario II.IV: CLOpidogrel Clinical Knowledge (ZEINCRO) - Detailed Workflow

1. Define the null hypothesis and the alternative hypothesis. A null hypothesis is a simple hypothesis associated with a contradiction to a theory which one would like to prove. An alternative hypothesis is a hypothesis associated with a theory which one would like to prove.

2. Define the subject selection criteria by invoking the "Search for Subjects" use case on the Linked2Safety platform.
    a. Request the frequency count of subjects with available information on drug use in patients treated with CLOpidogrel. Information regarding demographics, medical history, CLOpidogrel therapeutic schema, lab tests and adverse drug reactions are to be collected. It is expected that subjects will be from ZEINCRO, with a possibility of some data from CHUV if enough patients took this drug. CHUV has information on demographics and intake of CLOpidogrel (yes/no self reported) and other restrictions exist (see scenario template in Appendix A2).
    b. Determine whether there are enough subjects from the available data sources to test the study hypothesis.
    c. Request a breakdown of the frequency table to show which institutions and where, have how many eligible subjects. Determine whether these data sources are compatible (for e.g. determine possible sources of incompatibility or bias regarding the type of sample (e.g. general population sample versus clinical sample), subject selection (e.g. recruitment of particularly severe cases of depression in hospital settings versus inpatient settings, etc.).
3. Select chi-square from a dropdown list of test statistics available on the Linked2Safety platform (Figure 6).
4. Select the distribution of the test statistic using the Linked2Safety platform.
5. The Linked2Safety platform provides two options: (a) partition the possible values of test statistic T using the distribution of T and (b) set a significant level as threshold below which the null hypothesis is rejected.
6. Select option (a).
7. The distribution of the test statistic partitions the possible values of T into those for which the null-hypothesis is rejected, called critical region, and those for which it is not rejected. Moreover, the Linked2Safety platform computes the observed value of T from the samples and displays it in the "Text field to display the value of test statistic".
8. The Linked2Safety platform returns the outcome of the test based on the observed value as follows (Figure 6); reject the null hypothesis in favour of the alternative hypothesis if the observed value is in the critical region and accept the null hypothesis otherwise. Moreover, the Linked2Safety platform asks if the user wants to save the hypothesis test outcome.
9. Select to save the outcome of hypothesis testing into the Linked2Safety platform. The outcome is stored in the Knowledge Base.
10. Follow alternative flows that can be tested in this showcase (see Table 9):

Scenario II.V: Does Simvastatin treatment enhance prevention (ZEINCRO) - Detailed Workflow

1. Define the null hypothesis and the alternative hypothesis. A null hypothesis is a simple hypothesis associated with a contradiction to a theory which one would like to prove. An alternative hypothesis is a hypothesis associated with a theory which one would like to prove.

2. Define the subject selection criteria by invoking the "Search for Subjects" use case on the Linked2Safety platform.
    a. Request the frequency count of subjects with available information on drug use in patients treated with Simvastatin. Information regarding demographics, medical history and adverse drug reactions are to be collected. It is expected that subjects will be from ZEINCRO, with a possibility of some data from CHUV if enough patients took this drug, or if the study was broadened to similar treatments (see scenario template: Appendix A2).
    b. Determine whether there are enough subjects from the available data sources to test the study hypothesis.
    c. Request a breakdown of the frequency table to show which institutions and where, have how many eligible subjects. Determine whether these data sources are compatible (for e.g. determine possible sources of incompatibility or bias regarding the type of sample (e.g. general population sample versus clinical sample), subject selection (e.g. recruitment of particularly severe cases of depression in hospital settings versus inpatient settings, etc.).
3. Select chi-square from a dropdown list of test statistics available on the Linked2Safety platform (Figure 6).
4. Select the distribution of the test statistic using the Linked2Safety platform.
5. The Linked2Safety platform provides two options: (a) partition the possible values of test statistic T using the distribution of T and (b) set a significant level as threshold below which the null hypothesis is rejected.
6. Select option (a).
7. The distribution of the test statistic partitions the possible values of T into those for which the null-hypothesis is rejected, called critical region, and those for which it is not rejected. Moreover, the Linked2Safety platform computes the observed value of T from the samples and displays it in the "Text field to display the value of test statistic".
8. The Linked2Safety platform returns the outcome of the test based on the observed value as follows (Figure 6); reject the null hypothesis in favour of the alternative hypothesis if the observed value is in the critical region and accept the null hypothesis otherwise. Moreover, the Linked2Safety platform asks if the user wants to save the hypothesis test outcome.
9. Select to save the outcome of hypothesis testing into the Linked2Safety platform. The outcome is stored in the Knowledge Base.
10. Follow alternative flows that can be tested in this showcase (see Table 9):

5.4.1.4 Showcase III: Identification of relations between molecular fragments and specific adverse side effect categories- Detailed Workflow

As described in Deliverable D1.1, '*This scenario focuses on the identification of structural features in drugs that may be related to AEs in population sub-groups. The inspiration stems from the application of similar chemical structure-based techniques for the prediction of biological properties including potency, solubility, toxicity, etc. in the drug discovery field. To succeed in this, we need detailed information on patients' records including drugs administered and AEs. The availability of this information will allow the application of chemoinformatics*

methods that suggest potential chemical structure-adverse effect relationships. A subset of such potential relationships will be subjected to further analysis at a molecular level to elucidate the mechanism of action of the associated molecular sub-structure and the characteristics of the sub-population affected.'

Table 10 displays this use case and Figures 8 to 13 show mock-ups of the user interfaces to be used. This is followed by the detailed workflows relating to each type 3 showcase.

Name	Perform Data Analysis - Data Mining
Summary	This use case describes the process of data mining and finding associations that can be used in chemoinformatic analysis.
Primary Actor(s)	Clinical Researcher
Secondary Actor(s)	
Results(s)	The result of the data mining process is a set of rules mined by the data mining algorithm (decision trees algorithm, random forest algorithm, association rules algorithm). The outcome of manually selecting association rules is the selection of several association rules with supports greater than or equal to the minimum support and confidence (the threshold values). The rules that were not selected are removed. The outcome of finding associations that can be used in chemoinformatic analysis is a set of rules/relations between drugs with similar chemical structural features and AEs.

Normal Flow (Data Mining, Decision Trees):	
Step	**Action**
1	The actor initiates the data analysis procedure.
2	The Linked2Safety platform provides the actor with four options: (i) to test a hypothesis, (ii) to perform data mining, (iii) to manually select association rules, and (iv) to find associations that can be used in a chemoinformatic analysis.
3	The actor selects option (ii), to perform data mining. The data mining user interface pops up (Figure 8).
4	The actor defines the subject selection criteria by invoking the "Search for Subjects" use case on the Linked2Safety platform (Figure 8).

5	The Linked2Safety platform selects subjects (samples) according to the subject selection criteria (result of "Search for Subjects" use case).
6	The actor selects a data mining algorithm from a dropdown list of available data mining algorithms on the Linked2Safety platform (Figure 8).
7	The **decision tree algorithm** is selected, the Linked2Safety platform pops up the decision trees algorithm user interface (Figure 9).
8	The actor specifies the percentage split and starts running the algorithm on the samples (Figure 9).
9	The Linked2Safety platform displays in the "Data Mining Output" text area the decision tree constructed from the training set, the rules extracted from the decision tree and the accuracy of the decision tree and asks the actor if s/he wants to save the rules.
10	The actor selects to save the rules. The new rules are stored in the Knowledge Base.
Alternative Flows (Data Mining, Random Forest Algorithm):	
7.2a	In case **random forest algorithm** is selected, the Linked2Safety platform pops up the random forest algorithm user interface (Figure 10).
8.2a	The actor specifies the percentage split and starts running the algorithm on the samples (Figure 10).
9.2a	The Linked2Safety platform displays in the "Data Mining Output" text area the rules extracted from the random forest and the accuracy of the random forest and asks the actor if s/he wants to save the rules
10.2a	The actor selects to save the rules. The new rules are stored in the Knowledge Base.
Alternative Flows (Data Mining, Association Rules Algorithm):	
7.2b	In case **association rules algorithm** is selected, the Linked2Safety platform pops up the association rules algorithm user interface (Figure 11).
8.2b	The Linked2Safety platform displays in the "Data Mining Output" text area the association rules extracted from the algorithm and the accuracy of the association rules. Moreover, the Linked2Safety platform provides the actor with two options: (1) to select association rules or (2) to filter association rules.
9.2b	The actor presses the "select association rules" option (Continues from step 3.3: Manually Select Association Rules)
Alternative Flows (Data Mining, Filter Association Rules):	

9.2c	The actor presses the "filter association rules" option. The use case filter rules/associations are invoked to filter rules using relative Reporting Ratio (RR).
Alternative Flows (Manually Select Association Rules):	
3.3	The actor selects option (iii), to manually select association rules on the Linked2Safety platform. The select association rules user interface pops up (Figure 12).
4.3	The actor defines the selection criteria for association rules. Specifically, s/he selects the support and confidence metrics on the Linked2Safety platform, sets the threshold value i.e. the value of the minimum support and confidence, and starts the process.
5.3	The Linked2Safety platform selects the association rules with supports greater than or equal to the minimum support and confidence (the threshold values) and displays the selected rules in the "Selected Rules" text area. Moreover, the Linked2Safety platform asks the actor if s/he wants to save the rules.
6.3	The actor selects to save the selected rules onto the Linked2Safety platform. The rules are stored in the Knowledge Base.
Alternative Flows (Finding associations that can be used in chemoinformatic analysis):	
3.4	The actor selects option (iv), find associations that can be used in chemoinformatic analysis and the associations for chemoinformatic analysis user interface pops up (Figure 13).
4.4	The actor defines the subject selection criteria by invoking the "Search for Subjects" use case on the Linked2Safety platform (Figure 13).
5.4	The Linked2Safety platform selects subjects (samples) according to the subject selection criteria (result of "Search for Subjects" use case).
6.4	The Linked2Safety platform selects subjects (samples) according to the subject selection criteria (result of "Search for Subjects" use case).
7.4	The actor selects an algorithm data mining or the hypothesis testing algorithms from a dropdown list of the available algorithms on the Linked2Safety platform (Figure 13).
8.4	The Linked2Safety platform displays the output of the analysis and asks the actor if s/he wants to save the rules.
9.4	The actor selects to save the rules. The new rules are stored in the Knowledge Base.
Exceptional Flows (Data Mining):	
9.2c	(Exceptional to 9.2, 9.2a and 8.2b) The output rules are invalid because the data size is too small. The Linked2Safety platform displays a message describing that the accuracy of

	rules is too limited and the process continues from step 4.2.
10.2b	(Exceptional to 10.2 and 10.2a) The actor does not wish to save the rules into the Linked2Safety platform. The process is terminated.
Exceptional Flows (Manual Selection of Association Rules):	
5.3a	(Exceptional to 5.3) An inappropriate minimum support or minimum confidence is set such that all the rules are selected or no rules are selected. The Linked2Safety platform displays an error message describing the failure and the process continues from step 4.3.
6.3a	(Exceptional to 6.3) The actor does not wish to save the selected rules onto the Linked2Safety platform. The process is terminated.
Exceptional Flows (Finding associations that can be used in chemoinformatic analysis):	
8.4a	(Exceptional to 8.4) The relations discovered are not valid because the data sample size is too small. The process is terminated. The Linked2Safety platform displays a message describing that the accuracy of rules is too limited and the process continues from step 4.4.
9.4a	(Exceptional to 9.4) The actor does not wish to save the rules. The process is terminated.

Table 10 - Perform Data Mining or finding associations that can be used in chemoinformatic analysis Use Case (from D1.2)

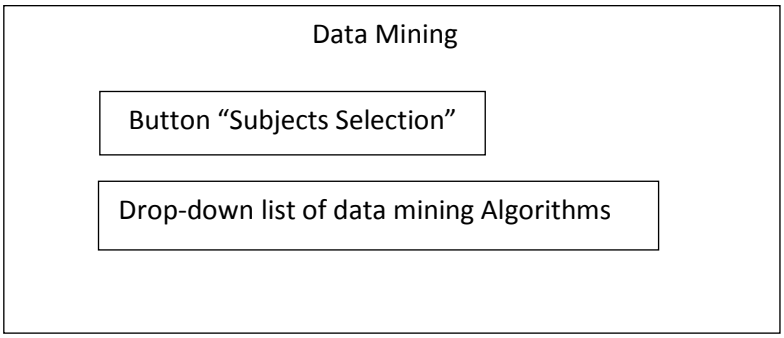

Figure 8 - User interface mock-up for data mining

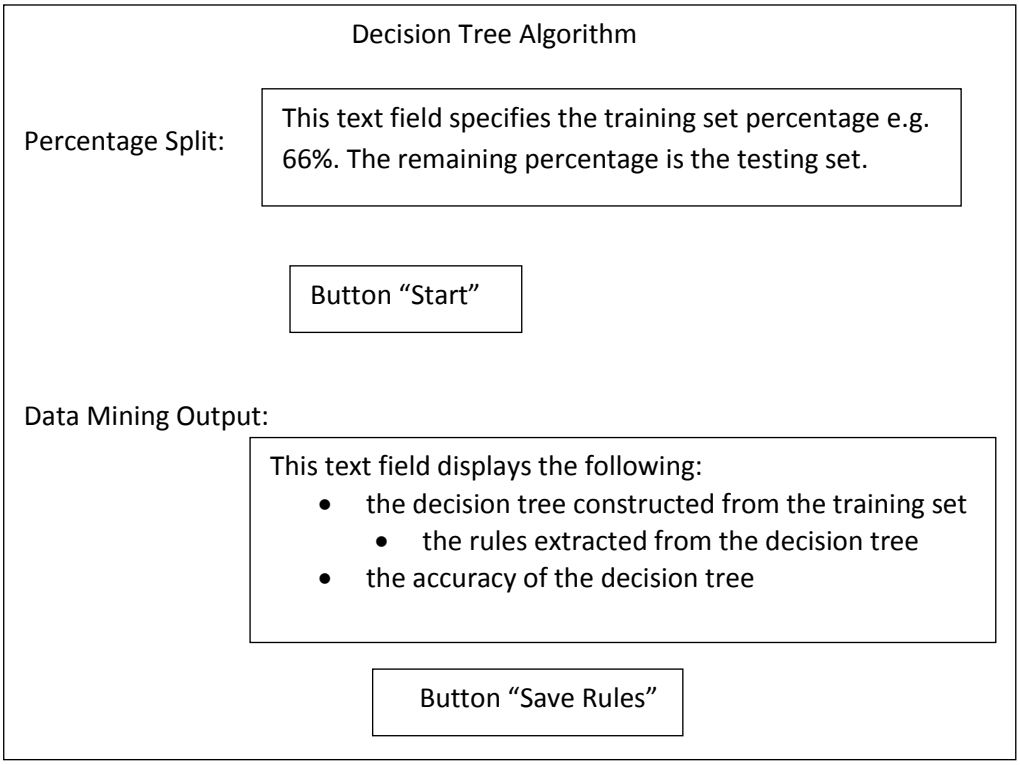

Figure 9 - User interface mock-up for decision trees

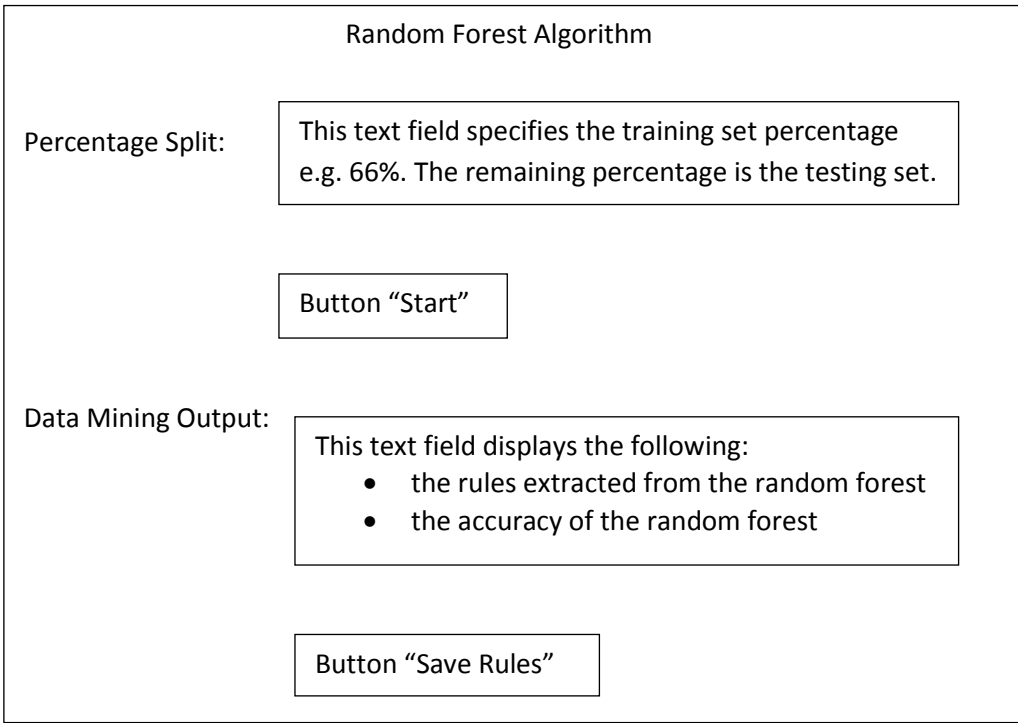

Figure 10 - User interface mock-up for random forest

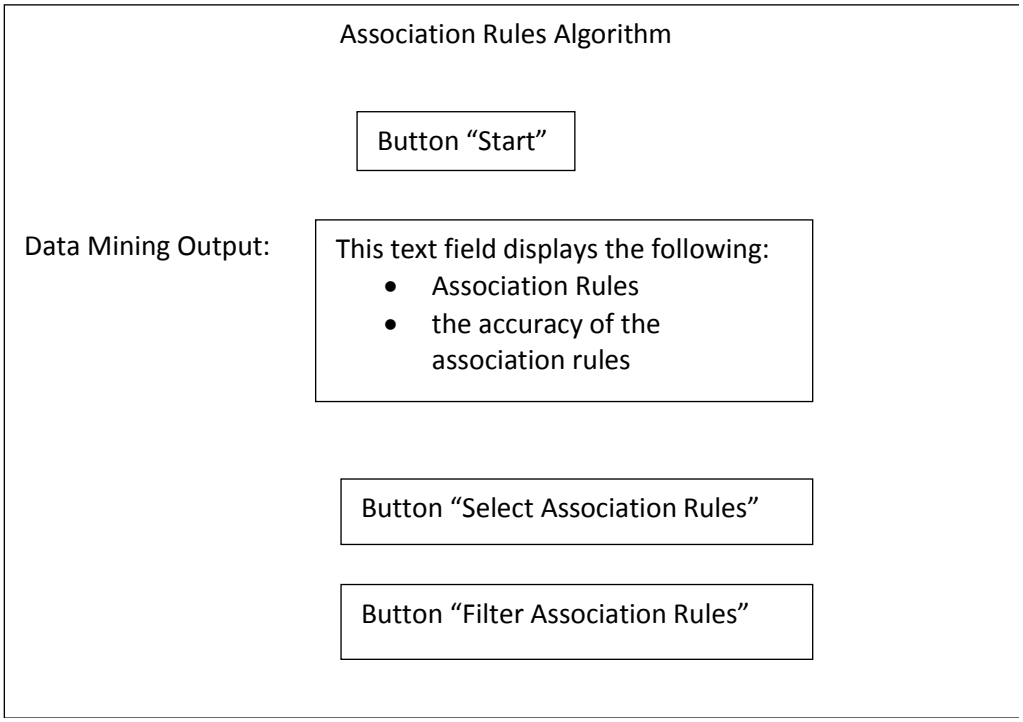

Figure 11- User interface mock-up for association rules mining

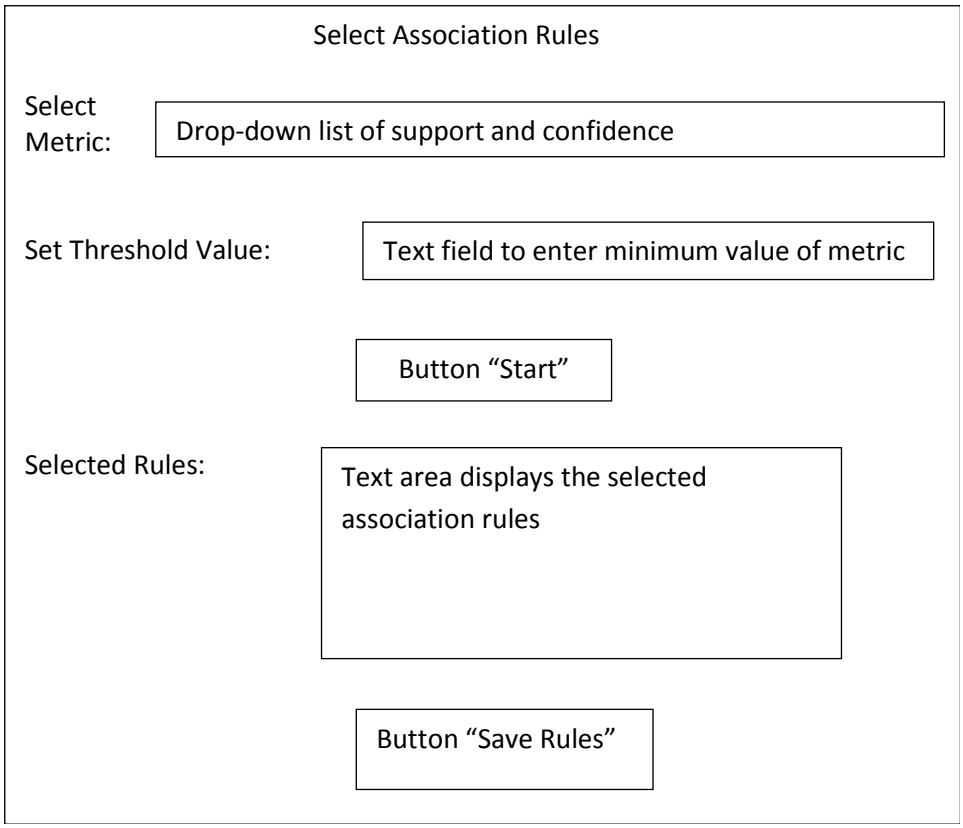

Figure 12 - User interface mock-up for manual selection of association rules

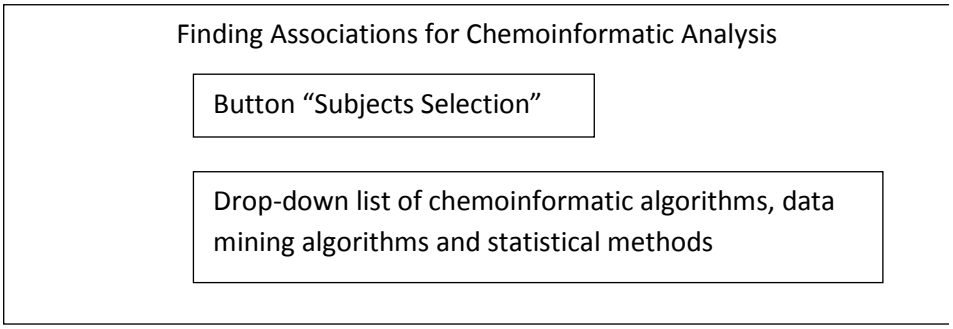

Figure 13 - User interface mock-up for finding associations that can be used in chemoinformatic analysis

Scenario III.I Anti-depressants and Cardio Issues (CHUV) - Detailed Workflow

1. Initiate the data analysis procedure.
2. The Linked2Safety platform provides the user with four options: (i) to test a hypothesis, (ii) to perform data mining, (iii) to manually select association rules, and (iv) to find associations that can be used in a chemoinformatic analysis.
3. Selects option (iv), find associations that can be used in chemoinformatic analysis and the associations for chemoinformatic analysis user interface pops up (Figure 13).
4. Define the subject selection criteria by invoking the "Search for Subjects" use case on the Linked2Safety platform (Figure 13).
   Select subjects with data myocardial infarction, BMI/weight, type II diabetes, hypertension and the lifetime use of ADM.
5. The Linked2Safety platform selects subjects (samples) according to the subject selection criteria (result of "Search for Subjects" use case).
6. Select an algorithm data mining or the hypothesis testing algorithms from a dropdown list of the available algorithms on the Linked2Safety platform (Figure 13).
7. The Linked2Safety platform displays the output of the analysis and asks the user if s/he wants to save the rules.
8. Select to save the rules. The new rules are stored in the Knowledge Base.
9. Follow alternative flows that can be tested in this showcase (see Table 10).

Scenario III.II: Asthma: ADEs and Formoterol (ZEINCRO) - Detailed Workflow

1. Initiates the data analysis procedure.
2. The Linked2Safety platform provides the user with four options: (i) to test a hypothesis, (ii) to perform data mining, (iii) to manually select association rules, and (iv) to find associations that can be used in a chemoinformatic analysis.
3. Select option (iv), find associations that can be used in chemoinformatic analysis and the associations for chemoinformatic analysis user interface pops up (Figure 13).

4. Define the subject selection criteria by invoking the "Search for Subjects" use case on the Linked2Safety platform (Figure 13).
    a. Search for subjects with asthma in relation to the presence of adverse drug reactions. It is expected that all data will come from ZEINCRO.
    b. Request a breakdown of the frequency table to show which institutions and where, have eligible subjects to test the study hypothesis. Determine whether these data sources are compatible (for e.g. determine possible sources of incompatibility or bias regarding the type of sample (e.g. general population sample versus clinical sample), subject selection (e.g. recruitment of particularly severe cases of depression in hospital settings versus inpatient settings, etc.), and methodology (e.g. the use of self-report versus biological measures).
5. The Linked2Safety platform selects subjects (samples) according to the subject selection criteria (result of "Search for Subjects" use case).
6. Select an algorithm data mining or the hypothesis testing algorithms from a dropdown list of the available algorithms on the Linked2Safety platform (Figure 13).
7. Conduct mining techniques and association testing to determine if Formoterol is associated with which adverse event in patients with asthma and which common molecular fragments might be responsible.
8. The Linked2Safety platform displays the output of the analysis and asks the user if s/he wants to save the rules.
9. Select to save the rules. The new rules are stored in the Knowledge Base.
10. Follow alternative flows that can be tested in this showcase (see Table 10):

Scenario III.III Depression: Anti-depressant Medicine and Glycaemic Effects (CING) - Detailed Workflow

1. Initiates the data analysis procedure.
2. The Linked2Safety platform provides the user with four options: (i) to test a hypothesis, (ii) to perform data mining, (iii) to manually select association rules, and (iv) to find associations that can be used in a chemoinformatic analysis.
3. Select option (iv), find associations that can be used in chemoinformatic analysis and the associations for chemoinformatic analysis user interface pops up (Figure 13).
4. Define the subject selection criteria by invoking the "Search for Subjects" use case on the Linked2Safety platform (Figure 13).
    a. Search for subjects with glyceamic side effects and drug use (including anti-depressant drug use) in relation to the presence of adverse drug reactions. It is expected that all data will come from CHUV. CHUV's data could also be enriched with data from ZEINCRO, if they have enough subjects with available information on blood glucose an reported ADM use.

   b. Request a breakdown of the frequency table to show which institutions and where, have eligible subjects to test the study hypothesis. Determine whether these data sources are compatible (for e.g. determine possible sources of incompatibility or bias regarding the type of sample (e.g. general population sample versus clinical sample), subject selection (e.g. recruitment of particularly severe cases of depression in hospital settings versus inpatient settings, etc.), and methodology (e.g. the use of self-report versus biological measures).
5. The Linked2Safety platform selects subjects (samples) according to the subject selection criteria (result of "Search for Subjects" use case).
6. Select an algorithm data mining or the hypothesis testing algorithms from a dropdown list of the available algorithms on the Linked2Safety platform (Figure 13).
7. Conduct mining techniques and association testing to determine if and which ADMs are associated with glyceamic side effects in patients with depression and which common molecular fragments might be responsible.
8. The Linked2Safety platform displays the output of the analysis and asks the user if s/he wants to save the rules.
9. Select to save the rules. The new rules are stored in the Knowledge Base.
10. Follow alternative flows that can be tested in this showcase (see Table 10):

Scenario III.IV: Hypercholestrerolemia: ADEs and Simvastatin (ZEINCRO) - Detailed Workflow

1. Initiates the data analysis procedure.
2. The Linked2Safety platform provides the user with four options: (i) to test a hypothesis, (ii) to perform data mining, (iii) to manually select association rules, and (iv) to find associations that can be used in a chemoinformatic analysis.
3. Select option (iv), find associations that can be used in chemoinformatic analysis and the associations for chemoinformatic analysis user interface pops up (Figure 13).
4. Define the subject selection criteria by invoking the "Search for Subjects" use case on the Linked2Safety platform (Figure 13).
   a. Search for subjects with drug use in patients with hypercholesterolemia in relation to the compliance of the patients and the presence of adverse drug reactions. It is expected that most data will come from ZEINCRO, but this could be enriched by CHUV data as they have information on cholesterol and cardiovascular system related drugs (yes/no self reported). In addition they also have a few subjects who took Simvastatin but the search could be extended to similar treatments. CHUV's contribution also depends on how many took the related drugs.
   b. Request a breakdown of the frequency table to show which institutions and where, have eligible subjects to test the study

hypothesis. Determine whether these data sources are compatible (for e.g. determine possible sources of incompatibility or bias regarding the type of sample (e.g. general population sample versus clinical sample), subject selection (e.g. recruitment of particularly severe cases of depression in hospital settings versus inpatient settings, etc.), and methodology (e.g. the use of self-report versus biological measures).

5. The Linked2Safety platform selects subjects (samples) according to the subject selection criteria (result of "Search for Subjects" use case).
6. Select an algorithm data mining or the hypothesis testing algorithms from a dropdown list of the available algorithms on the Linked2Safety platform (Figure 13).
7. Conduct mining techniques and association testing to determine if Simvastatin is associated with which adverse event in patients with hypercholestrerolemia and which common molecular fragments might be responsible.
8. The Linked2Safety platform displays the output of the analysis and asks the user if s/he wants to save the rules.
9. Select to save the rules. The new rules are stored in the Knowledge Base.
10. Follow alternative flows that can be tested in this showcase (see Table 10):

5.4.2 Timeline of Tasks 7.5 and 7.6 - Showcase Execution and Evaluation

The showcase execution and evaluation will be done in five phases, following a cyclical process. Each of the first three phases include the recording of showcase processes, completion of questionnaires and interviews, feedback evaluation, and optimisation and modification of the Linked2Safety framework and/or its documentation where necessary/possible, before the next phase begins. Once these phases are complete, examples of all showcase types (I-III) will have been run. An interim contingency phase is included to make any adjustments to the platform/documentation. Next, another showcase execution phase is completed where showcases can be re-run and/or remaining showcases run. This culminates in the generation of higher-level lessons learnt and adoption guidelines at the end of the final phase.

Each clinical partner will execute the showcases that they authored, and each execution/evaluation phrase will include a scenario from each scenario type, to be run by each data provider in parallel: 1) Subject selection, 2) Phase IV clinical trial post marketing surveillance and 3) Identification of relations between molecular fragments and specific adverse side effect categories. This is to allow focus on each showcase type by all partners involved.

Weeks 1-3

Task 7.6.1 *Clinical Partners:* Complete initial questionnaire and/or interview.

Task 7.5.1 *Clinical Partners:* Execute showcase scenario (Type 1)

Task 7.6.2 *Clinical Partners:* Monitoring of the performance and usability of the showcase via recording of processes/experiences, (questionnaire and/or interview) to provide useful feedback regarding further optimisations and other customisations required.

(See T7.4.6) *System Engineers:* Regular recording of processes / experiences.

(See T7.4.7) *System Engineers:* Update of Linked2Safety platform and/or documentation to fix any problems highlighted in this phrase of tests.

Task 7.6.3 *UCY / UNIMAN:* Data collection, regarding the end-users' and system engineers' feedback of their interaction with the Linked2Safety Platform and Showcases.

Weeks 4-6

Task 7.6.4 *Clinical Partners:* Complete initial questionnaire and/or interview.

Task 7.5.2 *Clinical Partners:* Execute showcase scenario (Type 2)

Tasks 7.6.5 *Clinical Partners:* Monitoring of the performance and usability of the showcase via recording of processes/experiences, (questionnaire and/or interview) to provide useful feedback regarding further optimisations and other customisations required.

(See T7.4.6) *System Engineers:* Regular recording of processes/experiences.

(See T7.4.7) *System Engineers:* Update of Linked2Safety platform and/or documentation to fix any problems highlighted in this phrase of tests.

Tasks 7.6.6 *UCY / UNIMAN:* Data collection, regarding the end-users' and system engineers' feedback of their interaction with the Linked2Safety Platform and Showcases.

Weeks 7-9

Task 7.6.7 *Clinical Partners:* Complete initial questionnaire and/or interview.

Task 7.5.3 *Clinical Partners:* Execute showcase scenario (Type 3)

Task 7.6.8 *Clinical Partners:* Monitoring of the performance and usability of the showcase via recording of processes/experiences, (questionnaire and/or interview) to provide useful feedback regarding further optimisations and other customisation required.

(See T7.4.6) *System Engineers:* Regular recording of processes/experiences.

(See T7.4.7) *System Engineers:* Update of Linked2Safety platform and/or documentation to fix any problems highlighted in this phrase of tests.

Task 7.6.9 *UCY / UNIMAN:* Data collection, regarding the end-users' and system engineers' feedback of their interaction with the Linked2Safety Platform and Showcases.

Weeks 10-12

(See T7.4.7) *System Engineers:* Test platform/documentation updates relating to the previous evaluation phases.

(See T7.4.8) *System Engineers:* Assist with generation of lessons learnt.

Task 7.5.4 *Clinical Partners:* Assist with System Engineers platform/documentation updates testing.

Task 7.6.10 *UCY / UNIMAN:* Identification/generation of lessons Learnt from first 3 phases.

Weeks 13-18

Task 7.6.11 *Clinical Partners:* Complete initial questionnaire and/or interview.

Task 7.5.5 *Clinical Partners:* Re-execute any required showcase scenarios that needed further development/documentation by the system engineers.

Run any remaining scenarios (ZEINCRO).

Task 7.6.12 *Clinical Partners:* Monitoring of the performance and usability of the showcase via recording of processes/experiences, (questionnaire and/or interview) to provide useful feedback regarding further optimisations and other customisations required.

Task 7.6.13 *UCY / UNIMAN:* Data collection, regarding the end-users' and system engineers' feedback of their interaction with the Linked2Safety.

Weeks 19-24

Task 7.6.14 *Clinical Partners:* Final Questionnaire and/or interview

Task 7.6.15 *Clinical Partners:* Assist with identification/generation of lessons learnt.

Task 7.6.16 *UCY / UNIMAN:* Identification/generation of lessons learnt from the Linked2Safety project (with regard to the development, deployment and execution of the showcases), formulating them as methodological adoption guidelines (including dos and don'ts).

The Gantt chart shown in Figure 14 displays all of the tasks 7.3 to 7.6 and where they run in parallel.

5.5 Tasks 7.3 – 7.6 Gantt Chart

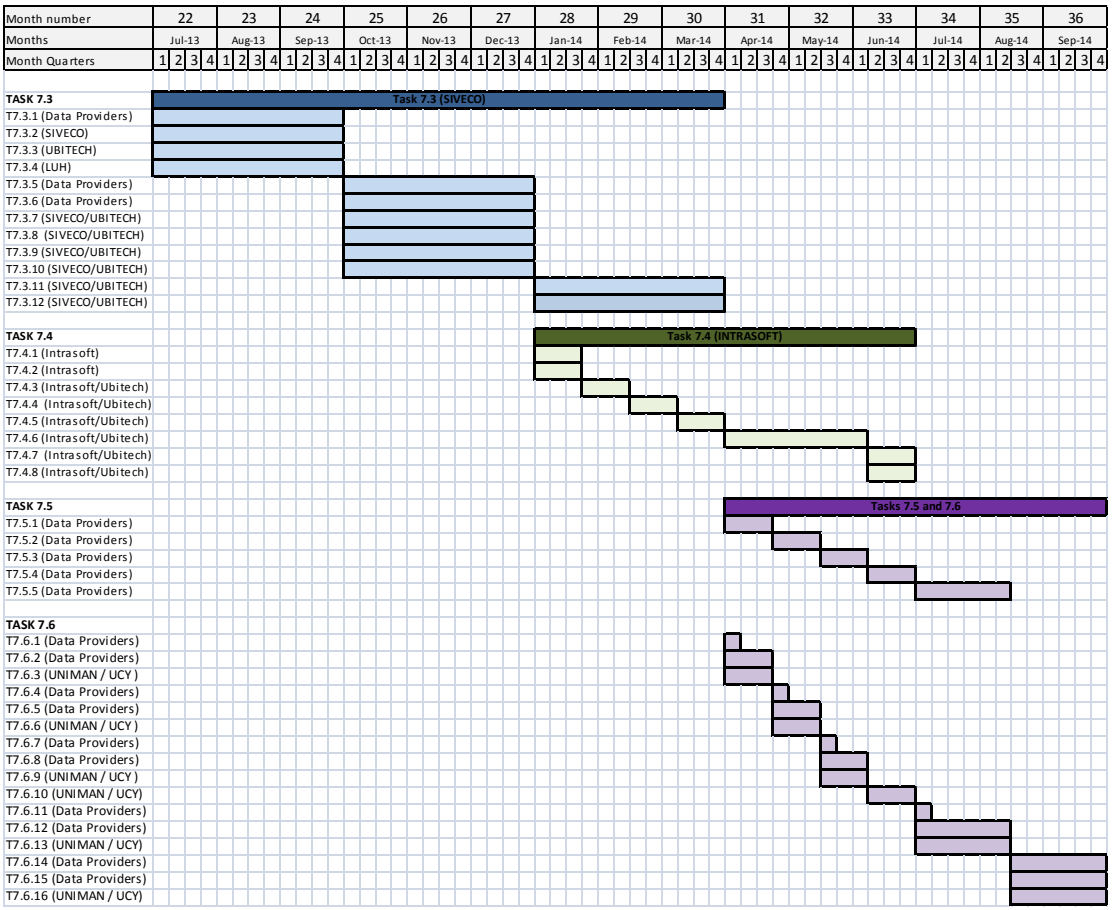

Figure 14 – Tasks 7.3 to 7.6 Gantt Chart

6. Protocol and Procedures to Evaluate Performance of Showcases

In this section, the evaluation protocol and methodology are defined stating the various practices for obtaining feedback from end-users. The evaluation protocol will be set in operation once the operation of the showcases starts, and will involve extensive data collection with regard to the end-users' feedback on their interaction with the Linked2Safety Platform and Showcases.

The implementation of this evaluation framework is expected to lead to valuable observations and conclusions about the viability and sustainability of the Linked2Safety platform and will conclude with the generation of the lessons learnt from the Linked2Safety project with regard to the development, deployment and execution of showcases. These are to be 'formulated as methodological adoption guidelines (including "dos and don'ts") enabling the reuse of semantically interoperable EHRs for advancing proactive patients' safety in clinical trials'.

6.1. Scenario-based assessment to evaluate the performance of showcases

Scenario-based assessment is proposed for the Linked2Safety technologies, as part of an attempt to address the needs of the main targeted user groups (stakeholders) of Linked2Safety. The targeted user groups serve as the point of departure for designing and executing evaluation. Evaluation is based on scenarios typical for each user group's context of using Linked2Safety. The suggestion to use activity scenarios is supported by previous research that showed that good assessment scenarios are particularly revealing and valuable because they ask learners to make decisions by applying their understanding of the system [7]. The literature documents several research attempts that focused on developing a scenario based assessment model in various diverse disciplines such as in: global software development for project management [8], online assessment of students' problem solving skills [9] and the assessment of clinical skills [10]. It is expected that the use of scenario-based evaluation will provide more accurate and detailed feedback from the end-users of the Linked2Safety system, compared to other types of evaluation.

Although three groups are presented, it is a key aspect to note that a specific institution will likely have individual scientists to cover all three roles. In many cases individual scientists will need to act in combinations of roles in their day to day activities also. For evaluation purposes we focus on evaluating each user group separately. The three main target groups of users (stakeholders) for the evaluation of the Linked2Safety platform are:

a) Medical Science analysts
b) Analytic methodology engineers
c) Data providers

a) **Medical science analysts** focus their efforts on analyzing data; they rely primarily on using existing statistical or computational methods to test pre-existing hypotheses or to generate new hypotheses depending on the problem they are working on. These will be typically associated with large pharmaceutical industry organizations, academic institutions interested in medical analyses and even hospitals and other medical care providers that perform analyses on data as part of their decision support, prognostics, or other efforts. Medical science analysts routinely seek for new sources of data to test their hypotheses with increased statistical power, using standardized analytical tools.
b) **Analytic methodology engineers** are focused on developing innovative analytical techniques to perform analyses on data and on evaluating those techniques. These may be derived from the fields of statistics, computational intelligence, data mining, software engineering and development, or other fields of study. Their focus is on the development of tools that can either introduce new analytic possibilities to solve medical problems through the analyses of medical data, or to introduce new iterations of analytic methodologies that are expected to have certain advantages over existing ones. Typically, an analytic methodology engineer can utilize the Linked2Safety platform as part of his efforts to evaluate newly developed tools, and once the tools are proven to be successful, the platform can serve for the quick deployment of his work to be made available to a large number of medical science analysts for use.
c) **Data providers** are institutions that hold medical data, these may be organizations that are responsible for data collected through clinical trials, epidemiological studies, health providers with patients' electronic health records, and others who have the ability to use that medical data in some form of research analyses. These users' primary focus will be to capitalize in terms of science on their work in collecting this data but they need to do this while following strict adherence to legal and ethical limitations.

It is worth noting here, that Medical Science Analysts and Analytic Methodology Engineers are what we previously collectively called "Expert Users". The term "expert users" has been replaced because the area of expertise is not sufficiently stated or implied in the term "expert users". It is also noted that an individual may have more than one role, and can therefore be a member of more than one of the three groups.

The idea behind scenario-based evaluation in this particular case is to request users to solve a problem, e.g. perform a type of task that is typical in their professional work, by utilizing the Linked2Safety platform. Users will then be asked to complete the Linked2Safety evaluation questionnaire. At this stage, for illustrative purposes, three scenarios are developed, one for each target group.

6.1.1. Example of scenario for target group 1 medical science analysts

[This scenario was based on Showcase 1: Phase III Clinical Trial: Subject Selection Scenario: Effect of Anti-hypertensive Drugs (CING)]

You are a medical science analyst, interested in examining the effect of anti-hypertensive drugs on Cardiovascular Disease (CVD) risk in diabetes type II patients. Specifically, you are interested in locating data and recruiting subjects in order to investigate the following three questions:

•Is diabetes type II highly correlated with high blood pressure?
•Does diagnosis with diabetes type II increase one's risk of CVD, while correcting for high blood pressure?
•In individuals with both diabetes type II and hypertension, does treatment with anti-hypertensive drugs, Candesartan and/or Hydroxhlorothiazide, decrease the increased CVD risk?

1. Approach this problem utilizing the Linked2Safety platform

[Users will receive detailed instructions including screenshots, if necessary, on how to use the basic functionality of the Linked2Safety platform that pertains to this task]

In this particular task, for the successful execution of this scenario, users are expected to request the frequency count of subjects with available information on diabetes and hypertension diagnosis as well as information on drug use and request a breakdown of the frequency table to show which institutions and where, have how many eligible subjects]

2. Complete the Linked2Safety evaluation questionnaire (questions are indicative)

6.1.2. Example of scenario for target group 2: Analytic methodology engineers

[This scenario was based on Showcase 3: Identification of relations between molecular fragments and specific adverse side effect categories: Anti-depressants and Cardio Issues (CHUV)]

You are an analytic methodology engineer, you have developed an innovative algorithm that structures compounds based on their chemical or other properties. You hypothesize that the clusters you have generated through your methodology will result in strong associations between the clusters you have created and adverse events experienced by subjects that undergo treatment with those compounds. You have focused your analyses based on the results of your algorithm into drugs typically administered to depressed patients.

1. Approach this problem utilizing the Linked2Safety platform

[Users will receive detailed instructions including screenshots, if necessary, on how to use the basic functionality of the Linked2Safety platform that pertains to this task]

2. Complete the Linked2Safety evaluation questionnaire [questions are indicative]

6.1.3. Example of scenario for target group 3: Data providers

[This scenario was based on Showcase 12: Phase III Glycaemic Effects Anti-depressant Medicines (CINGClinical Trial: Subject Selection)]

You are a researcher wishing to test an association between the tricyclic anti-depressant drug and adverse events of high blood sugar and diabetes in patients with depression. The number of subjects available at your institution is not large enough to provide you with sufficient statistical power to test this association. For this reason you wish to link your data with data from other researchers and you can approach this problem by becoming a data provider for Linked2Safety.

1. Approach this problem utilizing the Linked2Safety platform

[Users will receive detailed instructions including screenshots, if necessary, on how to prepare their data and use the basic functionality of the Linked2Safety platform that pertains to this task]

2. Complete the Linked2Safety evaluation questionnaire [questions are indicative]

6.2. Development of the "Linked2Safety evaluation questionnaire"

The Linked2Safety evaluation questionnaire was developed to be used an example of the proposed evaluation approach. The Linked2Safety evaluation questionnaire is structured in five main parts.

Part 1 of the evaluation questionnaire refers to users' personal information, such as gender, age, employment and experience. The second part allows users to evaluate the following five aspects of the Linked2Safety platform:

A. Analyses space: Questions that fall under the category of analyses space cover issues of subject selection, hypothesis testing, hypothesis generation, data mining, replication testing, time, cost and usability.

B. Linked Data Space

C. Usability

D. Legal and ethical issues

E. Value (for patients, future research)

The third part targets only members of Group 2 (analytic methodology engineers) and the fourth part targets only members of Group 3 (data providers). Lastly, the fifth part of the evaluation questionnaire provides users with the ability to express their opinion in a few open-ended questions that focus on ways to improve Linked2Safety. Questions are indicative, additional questions will be added, as necessary.

6.3. Data analysis for the evaluation of the performance of showcases

a) Quantitative analysis:

Descriptive statistics (frequencies) of participants' answers in parts 2, 3 and 4 (where applicable) will be provided to illustrate what the users' perceptions of the Linked2Safety system are, broken down into specific categories that can be analyzed in more detail (analyses space, linked data space, usability of the system, legal and ethical issues, value of the system).

Associations/correlations between the users' personal data (e.g. their experience, educational level, age, gender) and their perceptions of the Linked2Safety system (e.g. to what extent they value the system, to what extent they find specific tools user-friendly etc.) will be conducted.

A comparison of responses in questions that are common in the three groups (Part 2) will be conducted to identify whether there are any differences between users' perception of Linked2Safety depending on their role.

b) Qualitative analysis:

The open ended questions of part 5 will be analyzed qualitatively and themes emerging from the data will be coded, and translated into recommendations.

Statistical analyses combined with qualitative analyses will allow the generation of the lessons learnt from the Linked2Safety project with regard to the execution of showcases, and the formulation of methodological adoption guidelines to enable the reuse of semantically interoperable EHRs for advancing proactive patients' safety in clinical trials.

6.4. The Linked2Safety evaluation questionnaire

PART 1: Personal information

1. **Gender**:

a) Male
b) Female

2. **Age**:

a) Under 25 years old
b) 26-40 years old
c) 41-60 years old
d) Older than 61 years old

3. Completed Educational Level:

a) Secondary school
b) Undergraduate
c) Master's
d) PhD

4. **Employment:**

a) Pharmaceutical industry organization
b) Academic institution
c) Medical research institute
d) Hospital
e) Other: _____

5. Experience
a) No experience
b) 1-2 years of experience
c) 3-10 years of experience
d) More than 10 years of experience

PART 2:

Instructions: Compare the use of Linked2Safety to perform this task with what you would normally do to perform the same task by selecting the level of your agreement/disagreement with specific statements.

SD=Strongly disagree D=Disagree A=Agree SG=Strongly agree N/A=Not applicable

A. Analyses space					
1. The use of Linked2Safety assisted me to identify institutions with possible access to subjects	SD	D	A	SA	N/A
2. The use of Linked2Safety allowed me to discover a large pool of subjects fulfilling the desired inclusion criteria.	SD	D	A	SA	N/A
3. The use of Linked2Safety allowed unified analysis utilizing data from multiple institutions	SD	D	A	SA	N/A
4. The use of Linked2Safety allowed me to successfully test the hypothesis of the study.	SD	D	A	SA	N/A
5. With the use of Linked2Safety I was able to generate extra hypotheses worth investigating.	SD	D	A	SA	N/A
6. The use of Linked2Safety could increase the statistical power of my experiments.	SD	D	A	SA	N/A
7. I was able to use data mining to generate further hypotheses.	SD	D	A	SA	N/A
8. The use of Linked2Safety saves time when doing subject selection.	SD	D	A	SA	N/A
9. The use of Linked2Safety saves considerable money when doing subject selection.	SD	D	A	SA	N/A
10. With Linked2Safety I can find independent datasets that will enable testing my results for replication.	SD	D	A	SA	N/A
11. The Linked2Safety data mining tools were easy to use.	SD	D	A	SA	N/A
12. The "analytic space" was easy to use.	SD	D	A	SA	N/A
Please indicate which scenario or scenarios you based this evaluation on:					
If you feel strongly about a statement (e.g. you strongly disagree with an affordance of	Question number where I have strong				

Linked2Safety) please explain your reasons	disagreement: Reason:					
B. Linked Data Space						
13. With the use of Linked2Safety I was able to trace the result back to information about the original data (e.g. where they originated from, how they were collected).	SD	D	A	SA	N/A	
14. The technologies developed through Linked2Safety could provide a standardized approach to enable merging data from multiple sources for analysis.	SD	D	A	SA	N/A	
15. The technologies developed through Linked2Safety could provide a standardized approach to enable merging clinical terminologies from multiple sources for analysis	SD	D	A	SA	N/A	
16. The use of Linked2Safety reduced data integration problems when analyzing data from multiple institutions.	SD	D	A	SA	N/A	
17. The use of Linked2Safety allowed authorized access to restricted clinical resources.	SD	D	A	SA	N/A	
18. The use of Linked2Safety provided meaningful links of clinical terminologies.	SD	D	A	SA	N/A	
Please indicate which scenario or scenarios you based this evaluation on:						
If you feel strongly about a statement (e.g. you strongly disagree with an affordance of	Question number where I have strong					

Linked2Safety) please explain your reasons	disagreement: Reason:					
C. Usability						
19. The Linked2Safety platform was easy to use.	SD	D	A	SA	N/A	
20. The Linked2Safety platform was well documented with easy-to-use instructions.	SD	D	A	SA	N/A	
21. I would not use the Linked2Safety platform because learning the system requires a lot of time.	SD	D	A	SA	N/A	
22. I would not use the Linked2Safety platform because the user interface is too complex.	SD	D	A	SA	N/A	
23. I feel motivated to use Linked2Safety in the future.	SD	D	A	SA	N/A	
Please indicate which scenario or scenarios you based this evaluation on:						
D. Legal and ethical issues						
24. Linked2Safety enables the combined analysis of data from multiple institutions following strict adherence to legal and ethical limitations.	SD	D	A	SA	N/A	
25. I found the data privacy framework easy to understand.	SD	D	A	SA	N/A	
Please indicate which scenario or scenarios you based this evaluation on:						

E. Value (for patients, future research)					
26. Linked2Safety is a system that enables analysis that will maximize the positive effect to the patient.	SD	D	A	SA	N/A
27. I would recommend the Linked2Safety infrastructure to users who are not part of the current Linked2Safety project.	SD	D	A	SA	N/A
28. I would not use the Linked2Safety platform because the drawbacks outweigh the benefits.	SD	D	A	SA	N/A
29. The Linked2Safety platform will impact future research.	SD	D	A	SA	N/A
30. The Linked2Safety platform will impact future research if it is made publicly accessible.	SD	D	A	SA	N/A
31. The Linked2Safety platform will impact future research if it is adopted by many institutions (more than 100 000 total subjects).	SD	D	A	SA	N/A
Please indicate which scenario or scenarios you based this evaluation on:					

Table 11 – Part 2 of Evaluation Questionnaire

If you are an analytic methodology engineer complete Part 3. If not, skip to Part 4.

PART 3: Analytic methodology engineers

32. With Linked2Safety I can find independent datasets that will enable testing my results for replication.	SD	D	A	SA	N/A
33. The use of Linked2Safety saves time when seeking to deploy a new methodology on real data.	SD	D	A	SA	N/A
34. The use of Linked2Safety saves considerable money when seeking to deploy a new methodology on real data.	SD	D	A	SA	N/A
35. I was able to search for associations between drug substructures/molecular fragments and specific adverse side effect categories.	SD	D	A	SA	N/A
36. Linked2Safety would help me examine	SD	D	A	SA	N/A

innovative analytical techniques I develop.					
37. Linked2Safety would help me make innovative analytical techniques I develop available to a large number of medical science analysts for use.	SD	D	A	SA	N/A
Please indicate which scenario or scenarios you based this evaluation on:					
If you feel strongly about a statement (e.g. you strongly disagree with an affordance of Linked2Safety) please explain your reasons	Question number where I have strong disagreement: Reason:				

Table 12 – Part 3 of Evaluation Questionnaire

If you are a data provider complete Part 4. If not, skip to Part 5.

PART 4: Data providers

38. I was able to combine my data with similar data from an external institution using semantic linking to improve sample size.	SD	D	A	SA	N/A
39. The process of becoming a data provider saves time compared to other alternatives that enable data sharing.	SD	D	A	SA	N/A
40. The process of becoming a data provider allows a significant increase in sample numbers.	SD	D	A	SA	N/A
41. I found the mapping tools easy to use.	SD	D	A	SA	N/A
42. I found the transformation tools easy to use.	SD	D	A	SA	N/A
43. The process of becoming a data provider is simple.	SD	D	A	SA	N/A
44. I would recommend the Linked2Safety platform to data providers.	SD	D	A	SA	N/A
45. The benefits of becoming a data-provider outweigh the costs.	SD	D	A	SA	N/A
Please indicate which scenario or scenarios you based this evaluation on:					
If you feel strongly about a statement (e.g. you strongly disagree with an affordance of Linked2Safety) please explain your reasons	Question number where I have strong disagreement:				

	Reason:

Table 13 – Part 3 of Evaluation Questionnaire

PART 5:

Please share your opinion

1. If you strongly disagreed with specific questions in Parts 2-4, please explain your reasons.
2. What are the main limitations of Linked2Safety? How can they be overcome?
3. What are your suggestions/changes/recommendations to improve Linked2Safety?

7 Conclusion

This documents is "D7.1 – Scope Definition" of the Linked2Safety project and its main objective was to present the detailed scope definition of the Linked2Safety showcases, including the identification of the realistic clinical research use-cases that are going to be executed, the AS IS analysis of clinical trials use-cases, and their detailed organization and planning. It has also outlined the protocol and procedures to be followed during the evaluation of the performance of the showcases and quantified the expected benefits and their assessment measures and metrics [1].

A showcase scenario template has been presented, derived from the requirements identified in WP1. This template has been instantiated by the three clinical partners using the available data. The resulting showcases demonstrate all three required showcase types:

a) Subject Selection: the unbiased randomized selection of subjects in clinical research;
b) Phase IV Clinical Trial: Post Marketing Surveillance of a drug after it receives
permission to be sold;
c) Identification of relations between molecular fragments and specific adverse side effect categories.

Assessed By	Showcase Name	Description
CHUV	I.I	Selection of participants with and without major depressive disorder (MDD) and hypertension in order to request additional information (if needed) on the use of antidepressant drugs.
CING	I.II	Effect of anti-hypertensive drugs on Cardiovascular Disease in diabetes type II patients
ZEINCRO	I.III	A study to compare the efficacy and safety of the combination of Fluticasone / Salmeterol administered with device A against the innovator device B in patients with asthma.
ZEINCRO	I.IV	A multicentre, randomized, double-blind, double-dummy, cross-over, single dose study, comparing the efficacy and safety of the test Formoterol DPI (A) versus the innovative Formoterol DPI (B) in patients with asthma.

Table 14 – Showcase type 1 overview

Assessed By	Showcase Type	Description
CHUV	II.I	Association between major depressive disorder (MDD) and type II diabetes: the mediating role of BMI
CING	II.II	Glycaemic Effects of anti-depressant medicines (ADMs)
ZEINCRO	II.III	Multicenter, non-interventional clinical study of titration observance of carvedilol in patients with hypertension and diabetes.
ZEINCRO	II.IV	CLOpidogrel Clinical Knowledge
ZEINCRO	II.V	Does Simvastatin treatment enhance prevention?

Table 15 – Showcase type 2 overview

Assessed By	Showcase Type	Description
CHUV	III.I	Associations between anti-depressant medicine (ADM) and cardio-vascular risk factors (CVRFs) and cardio-vascular diseases (CVDs)
ZEINCRO	III.II	Associations between adverse events and Formoterol in patients with asthma.
CING	III.III	Associations between anti-depressant medicine (ADMs) and glycaemic effects in patients with depression
ZEINCRO	III.IV	Associations between adverse events and simvastatin in patients with hypercholestrerolemia.

Table 16 – Showcase type 3 overview

The detailed planning of the workpackage Tasks 7.3 to 7.6 was then presented.

Task 7.3: Interoperable EHRs Deployment

This task includes the packaging of binaries and provision of deployment instructions for the clinical partners and then the alignment, transformation and enrichment of the specific EHR and EDC repositories existing in each clinical

research and end user organisation. The clinical partners provide data as specified in the clinical trials showcases.

The next task is to establish the semantically interoperable EHR-related information resources space for the implementation of the clinical trials showcases identified in Task 7.1.

The final part of this task is to assist with Task 7.4 (Platform deployment and configuration), for tools / components relating to the "Interoperable EHR Data Space", responsible for the RDFization of a data cube and its semantic enrichment.

Task 7.4: Platform Deployment and Configuration
This task includes the deployment and testing of all platform tools/components to be used for each of the showcase types and assistance with Task 7.5 showcase execution. This includes the updating and testing of the Link2Safety platform where appropriate.

Tasks 7.5 and 7.6: Showcase Execution and Evaluation
Based on the requirements described in deliverable D1.2 (Linked2Safety Reference Architecture), a detailed description of the requirements relating to each of the showcase scenarios was presented. The showcase execution and evaluation will be done in five phases, following a cyclical process. Each of the first three phases include the recording of showcase processes, completion of questionnaires and interviews, feedback evaluation, and optimisation and modification of the Linked2Safety framework and/or its documentation where necessary/possible, before the next phase begins. This will culminate in the generation of higher-level lessons learnt and adoption guidelines at the end of the final phase.

Finally, an outline of the protocol and procedures to evaluate the performance of the showcases is derived. A scenario-based assessment is to be used, based upon the scenarios developed by the clinical partners. Evaluation is to involve extensive data collection with regard to the end-users' feedback on their interaction with the Linked2Safety platform and showcases. Examples have been given of three evaluation scenarios, each aimed at a different user type:

a) Medical science analysts
b) Analytic methodology engineers
c) Data providers

The complete set of scenarios for each user type will be detailed in the following deliverable 'Evaluation Strategy Definition and Validation' submitted in month 24.

The scope definition process, reported in this document has resulted in the creation of 13 showcase scenarios, each detailing the data to be used, the objectives to be achieved, how it is currently performed ("AS IS"), how it will be performed using the Linked2Safety platform ("TO BE") and the Linked2Safety tools required. Detailed timelines have been produced for each task to

successfully execute the scenarios and the evaluation methodology has been outlined, resulting in a comprehensive strategy for the entire workpackage.

8 References

[1] *Linked2Safety 288328 Project*, FP7-ICT-2011-7, Part B, 2011

[2] *Linked2Safety 288328 Project*, FP7-ICT-2011-7, Part A, 2011

[3] World Health Organization, *Good manufacturing practices: supplementary Guidelines for the manufacturer of investigational pharmaceutical products for clinical trials in humans*, No 863, 1996

[4] Friedman L.M., Furberg C. and DeMets D.L., *Fundamentals of Clinical Trials*, 4th edition, 2010

[5] Promponas V. et al. *GRANATUM: A Social Collaborative Working Space Semantically Interlinking Biomedical Researchers, Knowledge and Data for the Design and Execution of In-Silico Models and Experiments in Cancer Chemoprevention*, Intelligent Systems for Molecular Biology**,** ISMB/ECCB 2013, July 2013

[6] R Development Core Team. *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org, 2008

[7] Cooper B.C., Shepardson D.P. and Harber J.M., *Assessments as teaching and research tools in an environmental problem-solving program for in-service teachers*. Journal of Geoscience Education, 50(1), pp. 64-71, 2002

[8] Välimäki, A., Kääriäinen, J. and Koskimies, K, *Global Software Development Patterns for Project Management.* Procs. EuroSPI, Springer, pp. 137-148, 2009

[9] Mayotte S., *Online assessment of problem solving skills*, Computers in Human Behavior, vol. 26, pp. 1253–1258, 2010

[10] Nestel, B., Kneebone, R., Nolan, C., Akhtar, K. and Darzi, A., *Formative assessment of procedural skills: students' responses to the Objective Structured Clinical Examination and the Integrated Performance Procedural Instrument.* Assessment & Evaluation in Higher Education, *vol.* 36 (2), pp. 171–183, 2011

Appendices

A.1 Scenarios for Showcase 1

Scenario I.I: Anti-Depressant Drugs Evaluation (CHUV)

Scenario Outline	Selection of participants with and without major depressive disorder (MDD) and hypertension in order to request additional information (if needed) on the use of antidepressant drugs.
Background and Expected Outcome	Background: Both cardiovascular diseases (CVD) and psychiatric disorders are major public health issues leading to increased mortality and disability. They have an economic impact and are a major threat to individuals' lives and their quality of life. A positive association between MDD and hypertension is known from the literature. In addition, patients suffering from depression have reported lower levels of systolic blood pressure (BP) compared to controls. With a better understanding of the nature of the association between MDD and hypertension/levels of blood pressure, interventions with patients suffering from psychiatric disorders could be more specific and would help to prevent CVD. A potential mechanism which could explain the increased risk of hypertension in MDD patients is the intake of antidepressant drugs. However, findings on this are conflicting. Aim: To recruit a large sample of participants with and without MDD and hypertension in order to request additional information (if needed) on the use of antidepressant drugs. Expected Outcome: The expected outcome from the Linked2Safety platform is a list of hospitals/institutions where information on either MDD and/or hypertension/blood pressure are available. Listed sites will then be contacted to ask whether information on antidepressant drugs is available and, if not, whether we can collaborate in order to collect such information.
Data to be used	Available information needed for participants of this scenario is: • Demographic characteristics (age, sex): CHUV, ZEINCRO • MDD diagnosis: CHUV, ZEINCRO • Hypertension / blood pressure: CHUV, CING, ZEINCRO

	• Antidepressant drugs: CHUV CHUV will contribute approx. 3700 subjects of which 40% are participants with MDD. Information on hypertension/blood pressure and medications are available for almost all of them. Both CING and ZEINCRO collected information on hypertension/blood pressure. Moreover, ZEINCRO can provide MDD diagnosis (limited information, from medical history).
"AS-IS" Methods and Tools used (without L2S platform)	Without the Linked2Safety platform, a researcher would perform a systematic review of the existing literature to find other study groups working on the topic to contact.
"AS-IS" Workflow (without L2S platform):	1 Literature search: Search appropriate terms in MEDLINE, EMBASE, COCHRANE. Not restricted to published information but also to unpublished studies. No language restriction. 2 Selection of studies: Screening of all papers to identify possible collaborating centres. 3 Extract methodological information from candidates • Population (inpatients, outpatients, general population); • Sample size; • Assessments (self-reported questionnaires, interview, objective measurements); • Collected data. 4 Selection of research groups which assess information of interest 5 Contact of selected groups to obtain additional information from their subjects
Current issues /L2S Improvements	Difficulties in existing practice can be: 1 Literature search • Terms to define the same concept can differ from one study to another. Then by searching by terms one can miss some research groups working on the area of interest; • Some publications are still available only in foreign language (not English). It would be necessary to find a collaborator who is able to translate the paper. Thanks to the Linked2Safety platform, attributes for the same concept are mapped within the same entity (using the

	mapping tool) independently of the terms the studies used in practice. Language used is English independently of the country of origin of the study. 2 Selection of studies The same study group can publish several slightly different papers according to the aim of each paper. It becomes a loss of time to screen papers of the same study groups. Thank to the Linked2Safety platform collected data of a same study are already aggregated. 3 Extract methodological information from candidates • The same study group can publish results with different sample sizes according to the aim of the publication. It becomes difficult to have a clear idea of the exact number of participants in a study with specific criteria. • Adjustment is specific according to the specific topic of the publication. Not all available information in this study is presented in each publication. Thanks to the Linked2Safety platform, a better overview of the available data collected in a site is possible.
Objectives as an L2S showcase	This showcase demonstrates how the Linked2Safety platform avoids the amount of time (and cost) spent to perform the systematic review. It also demonstrates how the Linked2Safety platform can ease and speed the process of collaboration by targeting the groups interested in the same topic.
"TO-BE" Methods and Tools used with L2S platform	The Linked2Safety Medical Data Space will be used to combine date from CHUV and ZEINCRO. The data analysis space will be used to provide overall frequency counts and stratified frequency counts by institutions, available information.
"TO-BE" Workflow with L2S platform	1 Select attributes of interest • MDD • Blood pressure • Hypertension • Antidepressant drugs 2 Query to obtain frequency count of subjects with available information

	3 Query to obtain breakdown of the frequency count by institutions, combination of available information 4 Request names and contact numbers of the different institutions with subjects that meet inclusion criteria to obtain the additional information
Feasibility Analysis	The feasibility analysis includes the following checks: 1. Data availability: data are available as soon as the data-cubes can be transferred in the public access server (depending on availability of PC, DTA signature). 2. Data operation: N/A. 3. Data analysis: data analysis tools required for this scenario are available on the Linked2Safety platform.
Risk Identification and alternative	It is possible that data are not available on time since PC for the public access room is not available at CHUV and CHUV has not signed the DTA.
Assessment Measures and Metrics to Expected Results	Number of eligible subjects successfully identified. Enrichment of subject counts from CHUV with counts from ZEINCRO and CING. Successful creation of list of potential sites with existing available data.

Scenario I.II: Effect of Anti-hypertensive Drugs (CING)

Scenario Name	Effect of anti-hypertensive drugs on Cardiovascular Disease in diabetes type II patients
Background and Expected Outcome	Diabetes is known to be associated with circulatory problems since in subjects with diabetes, there is increased likelihood of build up of atherosclerotic plaques which in turn increase the risk of cardiovascular disease (CVD). One of the factors contributing to this higher CVD risk is that diabetes patients or patients with insulin resistance often also have a high blood pressure. In this scenario, we will investigate the following: • Is diabetes type II highly correlated with high blood pressure? • Does diagnosis with diabetes type II increase one's risk of CVD, while correcting for high blood pressure? • In individuals with both diabetes type II and hypertension, does treatment with anti-hypertensive

	drugs, Candesartan and/or Hydroxhlorothiazide, decrease the increased CVD risk? For the above investigations, we need to recruit the following groups of individuals: - Individuals not diagnosed with Diabetes type II, with or without hypertension - Individuals diagnosed with Diabetes type II, with or without hypertension For those individuals diagnosed with hypertension, we will need information on whether or not they were prescribed anti-hypertensive drugs
Data to be used	In order to recruit the appropriate subjects for this scenario, we need subjects with the following information available: - Diagnosis with diabetes type II or not - Diagnosis with hypertension or not - Information on anti-hypertensive drug use CING will contribute 1000 subjects, half with and half without diabetes. For all these subjects, we also have information on whether they were also diagnosed with hypertension. However, for CING subjects, drug use is self-reported. CHUV also has relevant data but drug use is also self-reported, and the number of subjects reporting taking the drug might be limited. For this reason, CING data can be enriched with ZEINCRO data, where drug use is documented in detail..
"AS-IS" Methods and Tools used (without L2S platform)	Without the Linked2Safety platform, any researcher wanting to carry out this investigation would have to recruit subjects into their study who have diabetes and hypertension diagnosis information available, as well as drug use information. In order to locate and recruit these individuals, researchers would have to contact collaborating research centres and/or hospitals that treat or deal with the diseases under study. Researchers would also have to ensure that drug use data for the same subjects is also available. If the sample sizes of available subjects from cooperating research centres and hospitals are not enough, then researchers would have to recruit more individuals from the general population.
"AS-IS"	1. A clinical researcher formulates a hypothesis and designs

Workflow (without L2S platform):	the study 2. The clinical researcher requests approval from the Health Authorities to perform the proposed research/study 3. (After receiving approval) The clinical researcher contacts via e-mail or telephone collaborating physicians, research centres, institutions and hospitals to identify sources of healthy and diseased subjects with available information on diabetes diagnosis, hypertension diagnosis, and drug use. 4. After identification of the centres that treat/recruit subjects who fit the selection criteria, the clinical researcher will determine whether the available sample sizes – considering the expected participation rate in such a study – are enough to give the study sufficient power so as to make the investigation worth-while. 5. Given that the sample sizes are enough, the clinical researcher will then request the collaborating hospitals, institutions, and research centres to identify those subjects with available information that are eligible, and invite them to participate in the study. 6. The clinical researcher will recruit those subjects that have responded to the invitation and given their informed consent to participate. 7. If the number of subjects available through the collaborating institutions is not enough to achieve sufficient power, the researcher may need to abandon the study or recruit more individuals, fulfilling the inclusion criteria, from the general population
Current issues /L2S Improvements	If the number of subjects available through the collaborating institutions is not enough to achieve sufficient power, the researcher may need to abandon the study or recruit more individuals, fulfilling the inclusion criteria, from the general population. However, going through the steps before getting a clear idea of the available and eligible sample sizes requires time and money. Thus abandonment of the study would result in waste of both time and money. With Linked2Safety, the process of determining which institutions/hospitals/research centres have eligible subjects is automated and saves tremendous amount of time and thus costs. In addition, a notion of whether there are enough available and eligible subjects is evident right away, preventing any wastes in terms of time and cost. Lastly, on Linked2Safety, only data from subjects who have given their consent is available. Therefore, if no further data is needed from the subjects (like in this case scenario), one can perform the analysis on the Linked2Safety platform without having to contact the subjects and receive explicit

	consent.
Objectives as an L2S showcase	This showcase scenario demonstrates the time and cost-efficient way of locating subjects eligible to participate in a study or in a particular data analysis, through Linked2Safety. It also demonstrates how Linked2Safety data can produce results of great interest/impact to the scientific community, without the need to contact in person the particular subjects.
"TO-BE" Methods and Tools used with L2S platform	The Linked2Safety Linked Medical Data Space will be used to combine data from CING and ZEINCRO where appropriate. In addition, the data analysis space could be used if we were going to carry out this scenario to completion, instead of focusing only on subject selection.
"TO-BE" Workflow with L2S platform	• Request the frequency count of subjects with available information on diabetes and hypertension diagnosis as well as information on drug use • Request a breakdown of the frequency table to show which institutions and where, have how many eligible subjects
Feasibility Analysis	The feasibility analysis includes the following checks: 1. Data availability: the required data will be available on Linked2Safety by mid-summer 2. Data operation: This will need to be done after the data becomes available 3. Data analysis: If this scenario was to be carried out to completion – beyond recruitment of subjects – correlation statistics and logistic regression models would be utilized, which are available on the Linked2Safety platform
Risk Identification and alternative	It is possible that linking self-reported data (CING and CHUV) and documented data (ZEINCRO) is not successful thus precluding enrichment.
Assessment Measures and Metrics to Expected Results	Number of subjects successfully identified Enrichment of subject counts from CING with counts from ZEINCRO
Any Other Comments	The scenario described above is one of subject selection for a research study/data analysis other than a clinical trial. The scenario demonstrates how Linked2Safety data can be used for subject recruitment in any research design/study

	and not just for clinical trials. This demonstration widens the audience/market for such a platform and indicates variability in possible usage. It also demonstrates how Linked2Safety data can produce results of great interest/impact to the scientific community, without the need to contact in person the particular subjects.

Scenario I.III Asthma – Fluticasone / Salmeterol: Device A versus Device B (ZEINCRO)

Scenario Name	A study to compare the efficacy and safety of the combination of fluticasone / Salmeterol administered with device A against the innovator device B in patients with asthma.
Background and Expected Outcome	Background: Asthma is a chronic inflammatory disorder of the airways in which many different types of cells and various cellular components are involved. The chronic inflammation causes an increased hyper-responsiveness of the airways, which leads to recurrent episodes of wheezing, breathlessness, chest tightness, and coughing, particularly at night or in the early morning hours. Regarding the pharmacological treatment of asthma, inhaled corticosteroids (CS) are the basis of maintenance therapy while the beta2-agonists are the long-term preferred additional therapy. The main clinical advantages of administering the medication directly into the lungs are associated with safety and efficacy: the side effects associated with the systemic circulation are zero, while high concentrations of the active substance can be directly attributed to the action points. Expected Outcome: to establish the therapeutic equivalence between the combination of Fluticasone / Salmeterol Sponsors administered with device A to the innovative device B, as to their bronchodilator effect on lung function.
Data to be used	For this scenario, Zeincro data on drug use in patients with asthma in relation to the efficacy of the study drug and the presence of adverse drug reactions will be investigated. This showcase cannot be enriched from CING or CHUV data. CING does not have data to contribute and CHUV cannot contribute to this showcase as it is not in line with their consent (only CVD/CVRF and Psychiatric disorders).
"AS-IS" Methods and	The method used without the Linked2Safety platform is that the investigator has to recruit subjects with asthma that are

Tools used (without L2S platform)	in line to the inclusion criteria of the study. In order for the investigator to recruit these subjects he would have to contact collaborating research centres and/or hospitals that treat or deal with asthmatic patients. Monitoring of these subjects must be performed throughout the study duration.
"AS-IS" Workflow (without L2S platform):	1. An investigator formulates a hypothesis for a study 2. Creation of the study protocol, statistical plan and all relevant safety documentation (Clinical safety reports, SAE forms etc) 3. Requesting approval from the Health Authorities and Ethics Committee to perform the proposed study 4. (After receiving approval) The investigator contacts physicians, hospitals/sites and fellow researchers in order to identifies sites that can provide the desired subjects 5. Invitation of subjects 6. Recruitment of subjects and retrieval of the inform consent forms 7. The investigator has to proceed with the randomization process (usually by an external vendor) since the study is double blind. 8. The study is being conducted and the subjects are being monitored throughout the study duration 9. Statistical plan is carried out again usually by external vendors 10. Review of outcomes 11. Submission of all documentation to Health Authorities
Current issues /L2S Improvements	Recruitment of subjects is a process that usually can take a lot of time, cost and effort especially if the investigator wants to conduct a clinical trial without external vendors or a sponsor. With Linked2Safety an investigator can have access to a large pool of subjects fulfilling the desired inclusion criteria without having to contact fellow researchers or sites. Additionally the potential gain from joining the Linked2Safety platform would be to combine own data with similar data from multiple institutions in order to increase the sample size without having to resort to interviewing more subjects or training additional collaborators. If enough data were available, some study hypotheses could even be tested without collecting any own data at all, using the data which is

	available on the Linked2Safety platform.
Objectives as an L2S showcase	This showcase demonstrates the time and effort that is required in order for a clinical trial to be conducted by an investigator. Through this showcase it has been noted that Linked2Safety can make the recruitment procedure a lot easier and cost and time effective through the ability of locating subjects through the Linked2Safety platform.
"TO-BE" Methods and Tools used with L2S platform	Data Cubes generated in closed room with components from the Interoperable EHR data space will be created and processed in the analysis space trough algorithms (such as chi-square)
"TO-BE" Workflow with L2S platform	Request the frequency count of subjects fulfilling the inclusion criteria (asthmatic patients) Determine whether there are enough subjects to test the study hypothesis. Conduct a chi square test of associating an adverse event with a suspect drug 1. An investigator formulates a hypothesis for a study 2. Creation of the study protocol, statistical plan and all relevant safety documentation (Clinical safety reports, SAE forms etc) 3. Requesting approval from the Health Authorities and Ethics Committee to perform the proposed study 4. (After receiving approval) The investigator through Linked2Safety does not have to contact physicians, hospitals/sites in order to identify the sites that can provide the desired subjects. He can request the frequency count of the desired subjects directly through querying on the Linked2Safety platform. 5. Based on the above identification and given the fact that the investigator would easily see if there are enough subjects (as opposed to waiting for individual responses from different sites/hospitals on the number of subjects) to test his hypothesis, the invitation of subjects would take place. 6. Recruitment of subjects and retrieval of the inform consent forms 7. The investigator has to proceed with the randomization process (usually by an external vendor) since the study

	is double blind. 8. The study is being conducted and the subjects are being monitored throughout the study duration 9. Statistical plan is carried out again usually by external vendors 10. Review of outcomes 11. Submission of all documentation to Health Authorities
Feasibility Analysis	The feasibility analysis includes the following checks: 1. Data availability: Data are already available from Zeincro. The available data should allow for the adequate testing of the study hypothesis. 2. Data operation: Manual test was conducted in order to verify that the outcome from the data cubes is close, In terms of numbers, to the actual data. 3. Data analysis: data analysis tools required for this scenario are available on the Linked2Safety platform
Risk Identification and alternative	Given the nature of data and the fact that the datasets do not have many common attributes since they originate from different studies with completely different CRF structure & design, it is possible that the results are not as rich as foreseen.
Assessment Measures and Metrics to Expected Results	Number of eligible subjects successfully identified.

Scenario I.IV: Asthma: Formoterol DPI (A) versus Formoterol DPI (B) (ZEINCRO)

Scenario Name	A multicentre, randomized, double-blind, double-dummy, cross-over, single dose study, comparing the efficacy and safety of the test Formoterol DPI (A) versus the innovative Formoterol DPI (B) in patients with asthma.
Background and Expected Outcome	Background: Bronchial asthma and chronic obstructive pulmonary disease (COPD) are among the world's most prevalent diseases. Inhaled therapy delivering bronchodilator and corticosteroid drugs in various doses is the mainstay of treatment for patients with asthma and COPD. The major advantage of

	delivering drugs directly into the airways via inhalation is that high concentrations can be delivered more effectively to the airways, and systemic side effects are avoided or minimized. The aim of inhaled therapy is to reverse and prevent airway inflammation and constriction and to minimise symptoms. Expected Outcome: to establish the therapeutic equivalence between the test and reference formoterol in terms of their bronchodilator effect in lung function.
Data to be used	For this scenario, Zeincro data on drug use in patients with asthma in relation to the efficacy of the study drug and the presence of adverse drug reactions will be investigated. This showcases cannot be enriched by CING or CHUV data. CING does not have data to contribute and CHUV cannot contribute to this showcase as it is not in line with their consent (only CVD/CVRF and Psychiatric disorders).
"AS-IS" Methods and Tools used (without L2S platform)	The method used without the Linked2Safety platform is that the investigator has to recruit subjects with asthma that are in line to the inclusion criteria of the study. In order for the investigator to recruit these subjects he would have to contact collaborating research centres and/or hospitals that treat or deal with asthmatic patients. Monitoring of these subjects must be performed throughout the study duration.
"AS-IS" Workflow (without L2S platform):	1. An investigator formulates a hypothesis for a study 2. Creation of the study protocol, statistical plan and all relevant safety documentation (Clinical safety reports, SAE forms etc) 3. Requesting approval from the Health Authorities and Ethics Committee to perform the proposed study 4. (After receiving approval) The investigator contacts physicians, hospitals/sites and fellow researchers in order to identifies sites that can provide the desired subjects 5. Invitation of subjects 6. Recruitment of subjects and retrieval of the inform consent forms 7. The investigator has to proceed with the randomization process (usually by an external vendor) since the study is double blind. 8. The study is being conducted and the subjects are being monitored throughout the study duration

	9. Statistical plan is carried out again usually by external vendors 10. Review of outcomes 11. Submission of all documentation to Health Authorities
Current issues /L2S Improvements	Recruitment of subjects is a process that usually can take a lot of time, cost and effort especially if the investigator wants to conduct a clinical trial without external vendors or a sponsor. With Linked2Safety an investigator can have access to a large pool of subjects fulfilling the desired inclusion criteria without having to contact fellow researchers or sites. Additionally the potential gain from joining the Linked2Safety platform would be to combine own data with similar data from multiple institutions in order to increase the sample size without having to resort to interviewing more subjects or training additional collaborators. If enough data were available, some study hypotheses could even be tested without collecting any own data at all, using the data which is available on the Linked2Safety platform.
Objectives as an L2S showcase	This showcase demonstrates the time and effort that is required in order for a clinical trial to be conducted by an investigator. Through this showcase it has been noted that Linked2Safety can make the recruitment procedure a lot easier and cost and time effective through the ability of locating subjects through the Linked2Safety platform.
"TO-BE" Methods and Tools used with L2S platform	Data Cubes generated in closed room with components from the Interoperable EHR data space will be created and processed in the analysis space trough algorithms (such as chi-square)
"TO-BE" Workflow with L2S platform	Request the frequency count of subjects fulfilling the inclusion criteria (asthmatic patients) Determine whether there are enough subjects to test the study hypothesis. Conduct a chi square test of associating an adverse event with a suspect drug *1. An investigator formulates a hypothesis for a study* *2. Creation of the study protocol, statistical plan and all relevant safety documentation (Clinical safety reports, SAE forms etc)* *3. Requesting approval from the Health Authorities and*

	Ethics Committee to perform the proposed study 4. *(After receiving approval) The investigator through Linked2Safety does not have to contact physicians, hospitals/sites in order to identify the sites that can provide the desired subjects. He can request the frequency count of the desired subjects directly through querying on the Linked2Safety platform.* 5. *Based on the above identification and given the fact that the investigator would easily see if there are enough subjects (as opposed to waiting for individual responses from different sites/hospitals on the number of subjects) to test his hypothesis, the invitation of subjects would take place.* 6. *Recruitment of subjects and retrieval of the informed consent forms* 7. *The investigator has to proceed with the randomization process (usually by an external vendor) since the study is double blind.* 8. *The study is being conducted and the subjects are being monitored throughout the study duration* 9. *Statistical plan is carried out again usually by external vendors* 10. *Review of outcomes* 11. *Submission of all documentation to Health Authorities*
Feasibility Analysis	The feasibility analysis includes the following checks: 1. Data availability: Data are already available from Zeincro. The available data should allow for the adequate testing of the study hypothesis. 2. Data operation: Manual test was conducted in order to verify that the outcome from the data cubes is close, In terms of numbers, to the actual data. 3. Data analysis: data analysis tools required for this scenario are available on the Linked2Safety platform
Risk Identification and alternative	Given the nature of data and the fact that the datasets do not have many common attributes since they originate from different studies with completely different CRF structure & design, it is possible that the results are not as rich as foreseen.
Assessment Measures and Metrics to Expected	Number of eligible subjects successfully identified.

Results	

A.2 Scenarios for Showcase 2

Scenario II.I Mediating role of BMI (CHUV)

Scenario Name	Association between major depressive disorder (MDD) and type II diabetes: the mediating role of BMI
Background and Expected Outcome	Background: General population studies have shown a link between type II diabetes and BMI, which are both independent risk factors for cardiovascular diseases (CVD). Moreover, there are some studies that have documented an association between type II diabetes and MDD and between BMI and MDD. Indeed, both diabetes and increased BMI have been documented as adverse events of depression. However, the mechanisms underlying these associations are still poorly understood. Indeed, does MDD increase the risk for diabetes independently from increased BMI, or is BMI the mediating factor between depression and diabetes? Expected outcome: it is likely that the increased likelihood of type II diabetes among depressed patients is partially mediated by BMI.
Data to be used	For this showcase, data from CHUV and ZEINCRO will be used. Both data providers have data on BMI/weight and type II diabetes, as well as on MDD (limited information in ZEINCRO dataset, mostly from medical history). The data from the non depressed subjects from CHUV and ZEINCRO will be used to constitute the control group. Any external data on depression, BMI/weight and type II diabetes would be welcome. However CING has data on type II diabetes and BMI but not on MDD.
"AS-IS" Methods and Tools used (without L2S platform)	Any researcher who would want to study these associations would need subjects with and without measures of depression, BMI/weight and type II diabetes/glucose levels. In order to obtain these subjects, researchers would need to contact research centres or clinical trial units that study these topics. Scientists studying this question would need to constitute variables with categorical outcomes (with and without depression, high versus low weight/ obesity versus non-obesity, with and without type II diabetes). First, statistical associations between all variables would need to be established using bivariate chi-square analyses. We could then conduct a logistic regression analysis with diabetes (yes/no) as the outcome variable and MDD as the independent variable, controlling for sex and age of subjects.

	A second model could be run, correcting the association between MDD and diabetes (if there is an association) for the risk of obesity. If the size of the association between diabetes and MDD in the first model decreases after adjustment for obesity in the second model, it could mean that BMI is a mediator, or partial mediator, of this association.
"AS-IS" Workflow (without L2S platform):	A clinical researcher designs the study, based on an extensive literature search, and defines the study hypothesis. He/she would need to identify sources of potential study subjects and make power calculations to determine the adequate sample size needed to prove his/her hypothesis. He/she would specify what the impact of this knowledge gain could be, and what the potential ramifications would be for public health policy (in this case, introduce more intensive clinical efforts to impede weight gain in depressed subjects as this could lead to the onset of type II diabetes). Usually, the researcher would need to put all this information into a study grant in order to obtain funding for the study from a governmental or private organization. Then, once the funding has been granted, the clinical researcher would request ethics approval from the competent Health Authorities to perform the study. Recruitment of subjects that fulfil inclusion criteria (here, MDD or not, low and high BMI and diabetes or not) for the study and data collection would take place after obtaining their informed consent to participate. Data analysis would then take place to assess the study hypotheses. The results of these analyses would be published in scientific journals. The researcher would determine what the limitations of the study were (e.g. self-reported biological measures or rating scales for depression), and how improvements could be made in the future (e.g. use objective biological measures such as blood samples and actual measures of weight and height, and direct interviews to assess MDD). These limitations and suggestions for improvement would figure in the publications.
Current issues /L2S Improvements	One of the biggest issues in this type of research is the difficulty to recruit enough subjects meeting inclusion criteria to conduct meaningful analysis (power analysis). Indeed, some subjects may drop out of the study, others may be unavailable or refuse to participate, and some subjects may have missing information (unable to answer some questions because of recall problems, etc). In other cases, problems with data collectors are relatively common (difficulty in finding and training enough collaborators to collect data, trained collaborators leaving the study unit prematurely, or providing insufficient work). The potential gain from joining the Linked2Safety platform

	would be to combine own data with similar data from multiple institutions in order to increase the sample size without having to resort to interviewing more subjects or training additional collaborators. If enough data were available, some study hypotheses could even be tested without collecting any own data at all, using the data which is available on the Linked2Safety platform.
Objectives as an L2S showcase	There are a lot of studies that have been completed on small groups of subjects, for instance for study limitations as stated above. It would be beneficial to clinical researchers to be able to address their study questions by combining their data with similar data from other sources. Having enough power would enable researchers to respond to questions of high scientific relevance. In addition, using available data from other sources would have an obvious financial gain in that recruiting additional subjects and obtaining additional resources would not be needed to answer the study question.
"TO-BE" Methods and Tools used with L2S platform	The data space will allow for the combination of variables provided by CHUV and ZEINCRO. In particular, data cubes that combine information on type II diabetes, BMI and MDD (versus controls) will be created using data from both sources. These data cubes will be processed in the analysis space using the proposed tests (chi-square and logistic regression analyses).
"TO-BE" Workflow with L2S platform	Request the frequency count of subjects with available information on diabetes, BMI and MDD. Determine whether there are enough subjects from the available data sources to test the study hypothesis. Request a breakdown of the frequency table to show which institutions and where, have how many eligible subjects. Determine whether these data sources are compatible (for e.g. determine possible sources of incompatibility or bias regarding the type of sample (e.g. general population sample versus clinical sample), subject selection (e.g. recruitment of particularly severe cases of depression in hospital settings versus inpatient settings, etc.). Conduct analyses for association testing (see above).
Feasibility Analysis	The feasibility analysis includes the following checks: 1. Data availability: as far as we know, these data are already available from CHUV and ZEINCRO. The available data should allow for the adequate testing of the study hypothesis. 2. Data operation: the question regarding whether the data cube generator will produce the semantically enriched and linked data as required will be examined this year. 3. Data analysis: We have the data analysis method capable

	to discover knowledge and show benefits.
Risk Identification and alternative	Based on the feasibility analysis above, the outcomes from data analysis might not produce the expected results (BMI is not a mediator between MDD and diabetes). In this case, it would be more difficult to explain and publish these results, and alternative mechanisms that would explain the association between MDD and diabetes would need to be tested. The role of other important variables (e.g. psychotropic medication including anti-depressants and anti-psychotics) on the risk of diabetes in MDD patients with obesity cannot be assessed due to the unavailability of these data in the ZEINCRO dataset. Another example is that the effect of physical exercise on the risk of obesity and diabetes cannot be assessed as these data have only been collected by CHUV.
Assessment Measures and Metrics to Expected Results	See above.
Any Other Comments	Potential data providers would need to provide a description of their recruitment methods and sample selection in order to establish the compatibility of their respective data sets. The study question that could be answered depends largely on the data that is made available on the Linked2Safety platform.

Scenario II.II: Glycaemic Effects Anti-depressant Medicines (CING)

Scenario Name	Glycaemic Effects of anti-depressant medicines (ADMs)
Background and Expected Outcome	Background: It is known from the literature that there exists a positive relationship between a history of depression and insulin resistance syndrome (MetS), but the degree to which ADMs are involved in this association is not well established. Some ADMs were evidenced to have glycaemic effects. For example, tricyclic antidepressants (TCAs) were shown to contribute to hyperglycaemia whereas other ADMs such as Selective serotonin reuptake inhibitors (SSRIs), were not associated with glycaemic effects.

	Expected Outcome: It is hypothesized that specific ADMs, such as TCAs cause glycaemic side effects, predisposing to the development of MetS and diabetes type II in patients with depression. However, it is also shown in the literature that depression is associated with lower physical activity, hence the degree to which a lower physical activity status in patients with depression could be contributing to the glycaemic effects of ADMs needs to be accounted for For this scenario, whether glycaemic side effects are more common in patients with depression taking TCAs compared to patients with depression taking SSRIs is investigated, with and without correction for levels of physical activity.
Data to be used	For this scenario, CHUV data on drug use in patients with depression will be investigated. In addition, physical activity and metabolic syndrome information (such as blood glucose levels) will be used. ZEINCRO has very little data on ADM use mostly in terms of medical history. However, if there are even a few subjects that have used TCAs or SSRIs, then CHUV data can be enriched.
"AS-IS" Methods and Tools used (without L2S platform)	Without the Linked2Safety platform, any researcher wanting to carry out this investigation would have to recruit subjects into their study who are diagnosed with depression. Additional eligibility criteria such as gender and age can also be applied. The eligible subjects will be divided into two groups – those being prescribed TCA drugs and those receiving SSRI drugs. Both groups will be followed for a period of time to determine adverse effect reactions. For this scenario, blood glucose measurements will need to be obtained at various time points following recruitment. In addition, information on the physical activity of the subjects will need to be obtained. In order to locate and recruit these individuals, researchers would have to contact collaborating research centres and/or hospitals that treat or deal with mental disorders such as depression. Researchers would also have to ensure that drug use data for the same subjects is also available so as to be able to determine drug exposure, separate the groups, and determine sample sizes. If the sample sizes of available subjects from cooperating research centres and hospitals are not enough, then researchers would have to contact more research/hospital centres and/or recruit more individuals from the general population something which would be very time consuming.

"AS-IS" Workflow (without L2S platform):	1. A clinical researcher formulates a hypothesis and designs the study 2. The clinical researcher requests approval from the Health Authorities to perform the proposed research/study 3. (After receiving approval) The clinical researcher contacts via e-mail or telephone collaborating physicians, research centres, institutions and hospitals to identify sources of subjects diagnosed with depression with available information on drug use. 4. After identification of the centres that treat/recruit subjects who fit the selection criteria, the clinical researcher will determine whether the available sample sizes – considering the expected participation rate in such a study – are enough to give the study sufficient power so as to make the investigation worth-while. 5. Given that the sample sizes are enough, the clinical researcher will then request the collaborating hospitals, institutions, and research centres to identify those subjects with available information that are eligible, and invite them to participate in the study. 6. The clinical researcher will recruit those subjects that have responded to the invitation and given their informed consent to participate. 7. These subjects will be followed for a particular time period. Blood glucose and other glycaemic index measurements as well as physical activity information will be collected at various time points, depending on the study design. 8. Statistical analyses (logistic regression) will be performed to determine the relationship between drug use and blood glucose as well as other glycaemic index levels, while correcting for physical activity status. The analyses will need to be performed independently for each time point since any adverse reactions may take some time to develop. 9. If the number of subjects available through the collaborating institutions is not enough to achieve sufficient power, the researcher may need to abandon the study or recruit more individuals, fulfilling the inclusion criteria, from the general population or from other institutions
Current issues /L2S Improvements	If the number of subjects available through the collaborating institutions is not enough to achieve sufficient power, the researcher may need to abandon the study or recruit more individuals, fulfilling the inclusion criteria, from the general

	population. However, going through the steps before getting a clear idea of the available and eligible sample sizes requires time and money. Thus abandonment of the study would result in waste of both time and money. With Linked2Safety, the process of determining which institutions/hospitals/research centres have eligible subjects is automated and saves tremendous amount of time and thus costs. In addition, a notion of whether there are enough available and eligible subjects is evident right away, preventing any wastes in terms of time and cost. Lastly, on Linked2Safety, the investigation of the hypothesis can be completely automated, without the need to even carry out the study. On Linked2Safety platform, only data from subjects who have given their consent is available. Therefore there is no need to contact subjects to obtain informed consent. In addition, if all necessary variables/information are on the Linked2Safety platform and no further data is needed from the subjects (like in this case scenario), one can perform the analysis on the Linked2Safety platform without having to follow the subjects and collect the necessary information themselves. This provides an opportunity to investigate a hypothesis, with increased study power but with minimal time and monetary costs.
Objectives as an L2S showcase	This showcase scenario demonstrates the time and cost-efficient way of locating subjects eligible to participate in a study or in a particular data analysis, through Linked2Safety. It also demonstrates how Linked2Safety data can produce results of great interest/impact to the scientific community, just by aggregating information already collected from other institutions, without the need to contact in person the particular subjects and carry out a study.
"TO-BE" Methods and Tools used with L2S platform	The Linked2Safety Linked Medical Data Space will be used to combine data from CHUV and ZEINCRO where appropriate. In addition, the data analysis space will be used to carry out the statistical analyses for this scenario. More explicitly, logistic regression and chi-square analysis tools will be used to determine whether there is an association between ADMs use and glycaemia.
"TO-BE" Workflow with L2S platform	• Request the frequency count of subjects diagnosed with depression who also have drug use, physical activity, and glycaemic index information available. • Determine whether the available sample sizes are enough for the statistical analysis to have enough power • Request a chi square test of association to determine

	whether the observed numbers of subjects with hyperglycaemia in patients prescribed TCAs and patients prescribed SSRIs differ from the expected numbers. This test does not take into account physical activity • In order to correct for the effect of physical activity, logistic regression can be carried out. The subjects will be divided into those who have developed hyperglycaemia and those who have not. Type of drug prescribed and physical activity will be entered into the model as exposure variables. If type of drug use (TCA vs. SSRI (baseline)) increases the risk of hyperglycaemia (significant OR above 1) while correcting for physical activity, we can then conclude that TCAs do have glycaemic side effects and can redispose to MetS and/or diabetes type II.
Feasibility Analysis	The feasibility analysis includes the following checks: 1. Data availability: data on drug use, physical activity, and blood glucose levels is already provided by CHUV, and ZEINCRO. CING data will also become available mid-summer. 2. Data operation: This will need to be done after the data becomes available 3. Data analysis: Data analysis tools required for this scenario are available on the Linked2Safety platform
Risk Identification and alternative	It is possible that linking self-reported data (ZEINCRO) and documented data (CHUV) is not successful thus precluding enrichment. In addition, if ZEINCRO subjects who reported ADM drug use are too few then linking of data will not be beneficial for this scenario.
Assessment Measures and Metrics to Expected Results	Number of eligible subjects successfully identified Enrichment of subject counts from CHUV with counts from ZEINCRO and CING Successful performance of chi-square test for association – biologically realistic effect estimates obtained Successful run of the logistic model to correct for physical activity – biologically realistic effect estimates obtained

Scenario II.III Titration Observation (ZEINCRO)

Scenario Name	Multicenter, non-interventional clinical study of titration observance of carvedilol in patients with hypertension and diabetes.

Background and Expected Outcome	This is a phase IV multicenter, non-interventional clinical study of titration observance of carvedilol in patients with hypertension and diabetes. Medication titration is the adjustment of the dose until the medication has achieved the desired effect. The study collects patient demographics, smoking habits, duration of hypercholesterolemia and arterial blood pressure, medical history of diabetes mellitus, renal failure and other medical conditions including treatments as well as the adverse events throughout the study duration. In addition blood and biochemical test results are collected every time they occur.
Data to be used	For this scenario, Zeincro data on drug use in patients with hypertension and diabetes in relation to the titration of the study drug and the presence of adverse drug reactions will be investigated. This showcase could be enriched by CHUV data (CHUV has information on hypertension, diabetes and intake of antihypertensive drugs (yes/no self-reported). We have a few subjects who took Carvedilol but we could extend our search to other similar treatment. CHUV could contribute providing enough participants take the drug.). CING does not have data on Carvediol intake or adverse drug reactions.
"AS-IS" Methods and Tools used (without L2S platform)	The method used without the Linked2Safety platform is that the investigator has to recruit subjects with hypertension and diabetes that are in line to the inclusion criteria of the study. In order for the investigator to recruit these subjects he would have to contact collaborating research centres and/or hospitals that treat or deal with patients with hypertension and diabetes. Monitoring of these subjects must be performed throughout the study duration.
"AS-IS" Workflow (without L2S platform):	1. An investigator formulates a hypothesis for a study 2. Creation of the study protocol, statistical plan and all relevant safety documentation (Clinical safety reports, SAE forms etc) 3. Requesting approval from the Health Authorities and Ethics Committee to perform the proposed study 4. (After receiving approval) The investigator contacts physicians, hospitals/sites and fellow researchers in order to identifies sites that can provide the desired subjects 5. Invitation of subjects 6. Recruitment of subjects and retrieval of the inform

	consent forms 7. The investigator has to proceed with the randomization process (usually by an external vendor) since the study is double blind. 8. The study is being conducted and the subjects are being monitored throughout the study duration 9. Statistical plan is carried out again usually by external vendors 10. Review of outcomes 11. Submission of all documentation to Health Authorities
Current issues /L2S Improvements	Recruitment of subjects is a process that usually can take a lot of time, cost and effort especially if the investigator wants to conduct a clinical trial without external vendors or a sponsor. With Linked2Safety an investigator can have access to a large pool of subjects fulfilling the desired inclusion criteria without having to contact fellow researchers or sites. Additionally the potential gain from joining the Linked2Safety platform would be to combine own data with similar data from multiple institutions in order to increase the sample size without having to resort to interviewing more subjects or training additional collaborators. If enough data were available, some study hypotheses could even be tested without collecting any own data at all, using the data which is available on the Linked2Safety platform.
Objectives as an L2S showcase	This showcase demonstrates the time and effort that is required in order for a clinical trial to be conducted by an investigator. Through this showcase it has been noted that Linked2Safety can make the recruitment procedure a lot easier and cost and time effective through the ability of locating subjects through the Linked2Safety platform.
"TO-BE" Methods and Tools used with L2S platform	The data provided by CHUV will be semantically linked with the ones from Zeincro by the Linked2Safety Linked Medical Data Space. Data Cubes generated in closed room with components from the Interoperable EHR data space will be created and processed in the analysis space trough algorithms (such as chi-square)
"TO-BE" Workflow with L2S platform	Request the frequency count of subjects fulfilling the inclusion criteria. Determine whether there are enough subjects to test the

	study hypothesis. Conduct a chi square test of associating an adverse event with a suspect drug
Feasibility Analysis	The feasibility analysis includes the following checks: 1. Data availability: Data are already available from Zeincro. The available data should allow for the adequate testing of the study hypothesis. 2. Data operation: Manual test was conducted in order to verify that the outcome from the data cubes is close, In terms of numbers, to the actual data. 3. Data analysis: data analysis tools required for this scenario are available on the Linked2Safety platform
Risk Identification and alternative	Given the nature of data and the fact that the datasets do not have many common attributes since they originate from different studies with completely different CRF structure & design, it is possible that the results are not as rich as foreseen.
Assessment Measures and Metrics to Expected Results	Number of eligible subjects successfully identified. Enrichment of subject counts from Zeincro with counts from CHUV.

Scenario II.IV: CLOpidogrel Clinical Knowledge (ZEINCRO)

Scenario Name	CLOpidogrel Clinical Knowledge
Background and Expected Outcome	This is a phase IV study for the Clopidogrel Knowledge. Clopidogrel is indicated in adults for the prevention of atherothrombotic events in patients suffering from myocardial infarction, ischaemic stroke or established peripheral arterial disease and in patients suffering from acute coronary syndrome. The collected data are: patient demographics, smoking habits, hypercholesterolemia, diabetes mellitus, description of event leading to hospitalization, actions taken and adverse event recording. Furthermore Clopidogrel dosage and aspirin co-administration per day. Additional concomitant medication including treatment. Monitoring of Clopidogrel dosages 7±5 days and 30 days after event. Additionally blood and

	biochemical test results are collected every time they occur.
Data to be used	For this scenario, Zeincro data on drug use in patients who are treated with Clopidogrel will be investigated. Information regarding demographics, medical history, Clopidogrel therapeutic schema, lab tests and adverse drug reactions are collected. These showcases could be enriched by CHUV data (CHUV has information on demographics and intake of Clopidogrel (yes/no self-reported). This analysis would be restricted to this drug (thrombosis which is CVRF) and not extended to other drugs from the class B (of ATC). CHUV could contribute providing enough participants take the drug.). CING does not have data to contribute here.
"AS-IS" Methods and Tools used (without L2S platform)	The method used without the Linked2Safety platform is that the investigator has to recruit subjects with hypercholesterolemia that are in line to the inclusion criteria of the study. In order for the investigator to recruit these subjects he would have to contact collaborating research centres and/or hospitals that treat or deal with patients with hypercholesterolemia. Monitoring of these subjects must be performed throughout the study duration.
"AS-IS" Workflow (without L2S platform):	1. An investigator formulates a hypothesis for a study 2. Creation of the study protocol, statistical plan and all relevant safety documentation (Clinical safety reports, SAE forms etc) 3. Requesting approval from the Health Authorities and Ethics Committee to perform the proposed study 4. (After receiving approval) The investigator contacts physicians, hospitals/sites and fellow researchers in order to identifies sites that can provide the desired subjects 5. Invitation of subjects 6. Recruitment of subjects and retrieval of the inform consent forms 7. The investigator has to proceed with the randomization process (usually by an external vendor). 8. The study is being conducted and the subjects are being monitored throughout the study duration 9. Statistical plan is carried out again usually by external vendors

	10. Review of outcomes 11. Submission of all documentation to Health Authorities
Current issues /L2S Improvements	Recruitment of subjects is a process that usually can take a lot of time, cost and effort especially if the investigator wants to conduct a clinical trial without external vendors or a sponsor. With Linked2Safety an investigator can have access to a large pool of subjects fulfilling the desired inclusion criteria without having to contact fellow researchers or sites. Additionally the potential gain from joining the Linked2Safety platform would be to combine own data with similar data from multiple institutions in order to increase the sample size without having to resort to interviewing more subjects or training additional collaborators. If enough data were available, some study hypotheses could even be tested without collecting any own data at all, using the data which is available on the Linked2Safety platform.
Objectives as an L2S showcase	This showcase demonstrates the time and effort that is required in order for a clinical trial to be conducted by an investigator. Through this showcase it has been noted that Linked2Safety can make the recruitment procedure a lot easier and cost and time effective through the ability of locating subjects through the Linked2Safety platform.
"TO-BE" Methods and Tools used with L2S platform	The data provided by CHUV will be semantically linked with the ones from Zeincro by the Linked2Safety Linked Medical Data Space. Data Cubes generated in closed room with components from the Interoperable EHR data space will be created and processed in the analysis space trough algorithms (such as chi-square)
"TO-BE" Workflow with L2S platform	Request the frequency count of subjects fulfilling the inclusion criteria. Determine whether there are enough subjects to test the study hypothesis. Conduct a chi square test of associating an adverse event with a suspect drug.
Feasibility Analysis	The feasibility analysis includes the following checks: 1. Data availability: Data are already available from Zeincro. The available data should allow for the adequate testing of the study hypothesis. 2. Data operation: Manual test was conducted in order to verify that the outcome from the data cubes is close, In

	terms of numbers, to the actual data. 3. Data analysis: data analysis tools required for this scenario are available on the Linked2Safety platform
Risk Identification and alternative	Given the nature of data and the fact that the datasets do not have many common attributes as they originate from different studies with completely different CRF structure & design, it is possible that the results may not be as rich as foreseen.
Assessment Measures and Metrics to Expected Results	Number of eligible subjects successfully identified. Enrichment of subject counts from Zeincro with counts from CHUV.

Scenario II.V: Does Simvastatin treatment enhance prevention (ZEINCRO)

Scenario Name	Does Simvastatin treatment enhance prevention?
Background and Expected Outcome	This is a phase IV study monitoring the potential prevention by Simvastatin treatment. Simvastatin is indicated for Hypercholesterolemia and Cardiovascular prevention. Patient demographics, smoking habits, family history of hypercholesterolemia and Simvastatin treatment. Arterial blood pressure and lipid profile is recorded as well as coexisting medical conditions, past medications and adverse events recording at each visit. Additionally blood and biochemical test results are collected every time they occur.
Data to be used	For this scenario, Zeincro data on drug use in patients who are treated with Simvastatin will be investigated. Information regarding demographics, medical history and adverse drug reactions are collected. These showcases could be enriched by CHUV data. (CHUV has information on demographics and intake of cardiovascular system related drugs (yes/no self-reported). We have a few subjects who took Simvastatin but we could extend our search to other similar treatment. CHUV could contribute providing enough participants take the drug.). CING does not have data to contribute here.
"AS-IS" Methods and Tools used (without L2S	The method used without the Linked2Safety platform is that the investigator has to recruit subjects with hypercholesterolemia that are in line to the inclusion criteria

platform)	of the study. In order for the investigator to recruit these subjects he would have to contact collaborating research centres and/or hospitals that treat or deal with patients with hypercholesterolemia. Monitoring of these subjects must be performed throughout the study duration.
"AS-IS" Workflow (without L2S platform):	1. An investigator formulates a hypothesis for a study 2. Creation of the study protocol, statistical plan and all relevant safety documentation (Clinical safety reports, SAE forms etc) 3. Requesting approval from the Health Authorities and Ethics Committee to perform the proposed study 4. (After receiving approval) The investigator contacts physicians, hospitals/sites and fellow researchers in order to identifies sites that can provide the desired subjects 5. Invitation of subjects 6. Recruitment of subjects and retrieval of the inform consent forms 7. The investigator has to proceed with the randomization process (usually by an external vendor). 8. The study is being conducted and the subjects are being monitored throughout the study duration 9. Statistical plan is carried out again usually by external vendors 10. Review of outcomes 11. Submission of all documentation to Health Authorities
Current issues /L2S Improvements	Recruitment of subjects is a process that usually can take a lot of time, cost and effort especially if the investigator wants to conduct a clinical trial without external vendors or a sponsor. With Linked2Safety an investigator can have access to a large pool of subjects fulfilling the desired inclusion criteria without having to contact fellow researchers or sites. Additionally the potential gain from joining the Linked2Safety platform would be to combine own data with similar data from multiple institutions in order to increase the sample size without having to resort to interviewing more subjects or training additional collaborators. If enough data were available, some study hypotheses could even be tested without collecting any own data at all, using the data which is

	available on the Linked2Safety platform.
Objectives as an L2S showcase	This showcase demonstrates the time and effort that is required in order for a clinical trial to be conducted by an investigator. Through this showcase it has been noted that Linked2Safety can make the recruitment procedure a lot easier and cost and time effective through the ability of locating subjects through the Linked2Safety platform.
"TO-BE" Methods and Tools used with L2S platform	The data provided by CHUV will be semantically linked with the ones from Zeincro by the Linked2Safety Linked Medical Data Space. Data Cubes generated in closed room with components from the Interoperable EHR data space will be created and processed in the analysis space trough algorithms (such as chi-square)
"TO-BE" Workflow with L2S platform	Request the frequency count of subjects fulfilling the inclusion criteria. Determine whether there are enough subjects to test the study hypothesis. Conduct a chi square test of associating an adverse event with a suspect drug
Feasibility Analysis	The feasibility analysis includes the following checks: 1. Data availability: Data are already available from Zeincro. The available data should allow for the adequate testing of the study hypothesis. 2. Data operation: Manual test was conducted in order to verify that the outcome from the data cubes is close, In terms of numbers, to the actual data. 3. Data analysis: data analysis tools required for this scenario are available on the Linked2Safety platform
Risk Identification and alternative	Given the nature of data and the fact that the datasets do not have many common attributes since they originate from different studies with completely different CRF structure & design, it is possible that the results may not be as rich as foreseen.
Assessment Measures and Metrics to Expected Results	Number of eligible subjects successfully identified. Enrichment of subject counts from Zeincro with counts from CHUV.

A.3 Scenarios for Showcase 3

Scenario III.I Anti-depressants and Cardio Issues (CHUV)

Scenario Name	Associations between anti-depressant medicine (ADM) and cardio-vascular risk factors (CVRFs) and cardio-vascular diseases (CVDs)
Background and Expected Outcome	Background: Due to their adverse cardiac events, antidepressants are less frequently used in patients with CVD. Tricyclic anti-depressants (TCAs) may increase the risk of myocardial infarction, and contribute to hyperglycemia, and weight gain. Venlafaxine has also been associated with dose dependent increases in blood pressure. Mirtazapine has been reported to cause weight gain, increase body fat mass, and has been associated with effects on glycemia. Bupropion may increase systolic blood pressure and has been associate with effects on glycemia, and weight. Expected outcomes: given the evidence from previous studies, it is likely that the use of anti-depressants is associated with increased CVRFs and CVDs. The output of the data mining tools will highlight any associations in the data, between anti-depressant drug sub-structures and side effect groups such as hyperglycemia, hypertension and weight gain.
Data to be used	For this showcase, data from CHUV will be used. CHUV has data on myocardial infarction, BMI/weight, type II diabetes, hypertension as well as on the lifetime use of ADM. This showcase would also test whether any other CVRFs which have been less studied and on which few data are available (such as dyslipidemia, smoking or physical inactivity) or CVDs other than myocardial infarction are associated with the lifetime use of ADM. The corresponding data from subjects who have never used any psychotropic medicine will be used to constitute the norms for the control group. Any external data on the above mentioned molecular components and risk factors/diseases would be welcome. CING and ZEINCRO have data on type II diabetes and BMI/weight. Moreover, CING can provide information on hypertension and ZEINCRO on MDD (limited information, mostly from medical history).
"AS-IS" Methods and Tools used (without L2S platform)	Any researcher who would want to study these associations would need subjects with and without the use of ADM established over lifetime, CVRFs and CVDs. In order to obtain these subjects, researchers would need to contact research centres or clinical trial units that study these topics. Scientists studying this question would need to constitute variables with categorical outcomes (with and without the lifetime use of ADM, high versus low weight/BMI, with and without type II

	diabetes, etc.). First, statistical associations between all variables would need to be established using bivariate chi-square analyses. We could then conduct a series of logistic regression analyses with the use of ADM (yes/no) as the outcome variable and one of each of the CVRFs or CVDs as the independent variable, controlling for sex and age of subjects.
"AS-IS" Workflow (without L2S platform):	A clinical researcher designs the study, based on an extensive literature search, and defines the study hypothesis. He/she would need to identify sources of potential study subjects and make power calculations to determine the adequate sample size needed to prove his/her hypothesis. He/she would specify what the impact of this knowledge gain could be, and what the potential ramifications would be for public health policy (in this case, avoid the use of ADM and suggest alternative therapeutic means (perhaps psychotherapy, luminotherapy or electroconvulsive therapy for severely affected subjects) in order to impede weight gain, the onset of type II diabetes or hypertension, etc in depressed subjects). Usually, the researcher would need to put all this information into a study grant in order to obtain funding for the study from a governmental or private organization. Then, once the funding has been granted, the clinical researcher would request ethics approval from the competent Health Authorities to perform the study. Recruitment of subjects that fulfil inclusion criteria for the study and data collection would take place after obtaining their informed consent to participate. As this would be an observational and not an interventional study, the consent form would specify that study participation has no implications on treatment if the subjects are currently in treatment. Data analysis would then take place to assess the study hypothesis. The results of these analyses would be exposed in an article to be published in scientific journals. The researcher would determine what the limitations of the study were (e.g. self-reported measures for CVRFs or CVDs), and how improvements could be made in the future (e.g. use objective biological measures such as blood samples, actual measures of weight and height and medical records for documenting CVDs). These limitations and suggestions for improvement would figure in the discussion section of the publication, and would be applied where possible to any further studies on the topic.
Current issues /L2S Improvements	One of the biggest issues in this type of research is the difficulty to recruit enough subjects meeting inclusion criteria to conduct meaningful analysis (power analysis). Indeed, some subjects may drop out of the study, others may be

	unavailable or refuse to participate, and some subjects may have missing information (unable to answer some questions because of recall problems, etc). In other cases, problems with data collectors are relatively common (difficulty in finding and training enough collaborators to collect data, trained collaborators leaving the study unit prematurely, or providing insufficient work). The potential gain from joining the Linked2Safety platform would be to combine own data with similar data from multiple institutions in order to increase the sample size without having to resort to interviewing more subjects or training additional collaborators. If enough data were available, some study hypotheses could even be tested without collecting any own data at all, using the data which is available on the Linked2Safety platform. Data mining will be used to generate further potential hypotheses (that might be missed by manual hypothesis creation), and which can be statistically tested in future work.
Objectives as an L2S showcase	There are many studies that have been completed on small groups of subjects, for instance because of study limitations as stated above. It would be beneficial to clinical researchers to be able to address their study questions by combining their data with similar data from other sources. Having enough power would enable researchers to respond to questions of high scientific relevance. In addition, using available data from other sources would have an obvious financial gain in that recruiting additional subjects and obtaining additional financial resources would not be needed to answer the study question. The generation of hypotheses, to be used for future statistical research into ADMs.
"TO-BE" Methods and Tools used with L2S platform	The data space will allow for the combination of variables (in specific data cubes) related to ADM and CVRFs and CVDs. These data cubes will be processed in the analysis space using the proposed tests (chi-square and logistic regression analyses). The examination of undocumented associations between the CVRFs/CVDs and ADMs would follow existing statistical rules (e.g. adjustment of the p-value) because these potential associations are not based on pre-existing hypotheses from the scientific literature. Data cubes containing the ADM, CVRF and CVD variables will be used for data mining, using decision tree, random forests and association rule mining algorithms.
"TO-BE" Workflow with L2S platform	Request the frequency count of subjects with available information on the relevant variables. Determine whether there are enough subjects from the

	available data sources to test the study hypothesis. Request a breakdown of the frequency table to show which institutions and where, have eligible subjects to test the study hypothesis. Determine whether these data sources are compatible (for e.g. determine possible sources of incompatibility or bias regarding the type of sample (e.g. general population sample versus clinical sample), subject selection (e.g. recruitment of particularly severe cases of depression in hospital settings versus inpatient settings, etc.), and methodology (e.g. the use of self-report versus biological measures). Conduct analyses for association testing (see above). Conduct analyses for data mining as follows: Before data mining is run: (Manually) Add rules considered to be as mandatory although they were not discovered by the data mining process (e.g. because the conditions of the rules never occur or because the conditions occur but not lead to any outcome) Once data mining has been run, manual selection of association rules (e.g. among a set of variables and a diabetes type II diagnosis) that have more than a given, user-defined, minimum "support" value and those that have more than a given, user-defined minimum "confidence" value. Filter rules obtained from data mining both automatically and by visual inspection. (Manually) Remove rules that do not apply to general population, but to the specific healthcare data analysed The expert user can then query the knowledge base for mined rules and experts' knowledge.
Feasibility Analysis	The feasibility analysis includes the following checks: 1. Data availability: these data are already available from CHUV. The available data should allow for the adequate testing of the study hypothesis and data mining analysis. 2. Data operation: the question regarding whether the data cube generator will produce the semantically enriched and linked data as required will be examined this year. 3. Data analysis: We have the data analysis / mining methods capable to discover knowledge and show benefits.
Risk Identification and alternative	Based on the feasibility analysis above, the outcomes from data analysis might not produce the expected results (ADM is not associated with any of the CVRFs or CVDs). In this case, it would be more difficult to explain and publish these negative results, and alternative mechanisms that would

	explain these associations would need to be tested. For example, the role of other important molecular fragments (e.g. anti-psychotic medication) on the risk of CVRFs or CVDs could be further examined. The data mining outputs, for example association rules, or decision trees, could assist with this.
Assessment Measures and Metrics to Expected Results	See above.
Any Other Comments	Potential data providers would need to provide a description of their recruitment methods and sample selection in order to establish the compatibility of their respective data sets. The study question that could be answered depends largely on the data that is made available on the Linked2Safety platform.

Scenario III.II: Asthma: ADEs and Formoterol (ZEINCRO)

Scenario Name	Associations between adverse events and Formoterol in patients with asthma.
Background and Expected Outcome	Background: Clinical interventional study (double blind): Equivalence study between formoterol (Drug A) administered via Device A versus innovative formoterol (Drug B) administered via device B in patients with mild to moderate persistent asthma. Asthma is a chronic inflammatory disorder of the airways in which many different types of cells and various cellular components are involved. The chronic inflammation causes an increased hyper-responsiveness of the airways, which leads to recurrent episodes of wheezing, breathlessness, chest tightness, and coughing, particularly at night or in the early morning hours. Regarding the pharmacological treatment of asthma, inhaled corticosteroids (CS) are the basis of maintenance therapy while the beta2-agonists are the long-term preferred additional therapy. The main clinical advantages of administering the medication directly into the lungs are associated with safety and efficacy: the side effects associated with the systemic circulation are zero, while high concentrations of the active substance can be directly attributed to the action points. The scope of this study is primarily the evaluation of the therapeutic equivalence of a new generic formoterol dry powder 12mg (drug A) compared the prototype drug formoterol (drug B). A secondary scope is to evaluate the safety and tolerability of a new formulation of formoterol generic 12mg and to evaluate the long-term bronchodilator formoterol formulation of a new dry powder of 12mg. Expected Outcomes: to show that Drug A is not inferior to Drug B. The output of the data mining tools will highlight any associations in the data, between Formoterol, innovative Formoterol drug sub-structures and groups of adverse drug reactions in patients with asthma.
Data to be used	For this scenario, Zeincro data on drug use in patients with asthma in relation to presence of adverse drug reactions will be investigated. This showcase cannot be enriched from CING or CHUV data. CING does not have data to contribute and CHUV cannot contribute to this showcase as it is not in line with their consent (only CVD/CVRF and Psychiatric disorders).
"AS-IS"	The method used without the Linked2Safety platform is that

Methods and Tools used (without L2S platform)	the investigator has to recruit subjects with asthma that are in line with the inclusion criteria of the study. In order for the investigator to recruit these subjects he would have to contact collaborating research centres and/or hospitals that treat or deal with asthmatic patients. Monitoring of these subjects must be performed throughout the study duration.
"AS-IS" Workflow (without L2S platform):	1. An investigator formulates a hypothesis for a study 2. Creation of the study protocol, statistical plan and all relevant safety documentation (Clinical safety reports, SAE forms etc) 3. Requesting approval from the Health Authorities and Ethics Committee to perform the proposed study 4. (After receiving approval) The investigator contacts physicians, hospitals/sites and fellow researchers in order to identifies sites that can provide the desired subjects 5. Invitation of subjects 6. Recruitment of subjects and retrieval of the inform consent forms 7. The investigator has to proceed with the randomization process (usually by an external vendor) since the study is double blind. 8. The study is being conducted and the subjects are being monitored throughout the study duration 9. Statistical plan is carried out again usually by external vendors 10. Review of outcomes 11. Submission of all documentation to Health Authorities
Current issues /L2S Improvements	Recruitment of subjects is a process that usually can take a lot of time and effort especially if the investigator wants to conduct a clinical trial without external vendors or a sponsor. With Linked2Safety an investigator can have access to a large pool of subjects fulfilling the desired inclusion criteria without having to contact fellow researchers or sites. Additionally the potential gain from joining the Linked2Safety platform would be to combine own data with similar data from multiple institutions in order to increase the sample size without having to resort to interviewing more subjects or training additional collaborators. If enough data were available, some study hypotheses could even be tested

	without collecting any own data at all, using the data which is available on the Linked2Safety platform. Data mining will be used to generate further potential hypotheses (that might be missed by manual hypothesis creation), and which can be statistically tested in future work.
Objectives as an L2S showcase	This showcase demonstrates the time and effort that is required in order for a clinical trial to be conducted by an investigator. Through this showcase it has been noted that Linked2Safety can make the recruitment procedure a lot easier and cost and time effective through the ability of locating subjects through the Linked2Safety platform. The generation of hypotheses, to be used for future statistical research into Formoterol drugs.
"TO-BE" Methods and Tools used with L2S platform	Data Cubes generated in closed room with components from the Interoperable EHR data space will be created and processed in the analysis space trough algorithms (such as chi-square) Data cubes containing the Formoterol and adverse event variables will be used for data mining, using decision tree, random forests and association rule mining algorithms.
"TO-BE" Workflow with L2S platform	Request the frequency count of subjects with available information on the relevant variables. Determine whether there are enough subjects from the available data sources to test the study hypothesis. Request a breakdown of the frequency table to show which institutions and where, have eligible subjects to test the study hypothesis. Determine whether these data sources are compatible (for e.g. determine possible sources of incompatibility or bias regarding the type of sample (e.g. general population sample versus clinical sample), subject selection (e.g. recruitment of particularly severe cases of depression in hospital settings versus inpatient settings, etc.), and methodology (e.g. the use of self-report versus biological measures). Conduct mining techniques and association testing to determine if Formoterol is associated with which adverse event in patients with asthma and which common molecular fragments might be responsible. Conduct this analysis as follows: Before data mining is run: (Manually) Add rules considered to be as mandatory although they were not discovered by the data mining process (e.g. because the conditions of the rules

	never occur or because the conditions occur but not lead to any outcome) Once data mining has been run, manual selection of association rules (e.g. among a set of variables and a diabetes type II diagnosis) that have more than a given, user-defined, minimum "support" value and those that have more than a given, user-defined minimum "confidence" value. Filter rules obtained from data mining both automatically and by visual inspection. (Manually) Remove rules that do not apply to general population, but to the specific healthcare data analysed The expert user can then query the knowledge base for mined rules and experts' knowledge.
Feasibility Analysis	The feasibility analysis includes the following checks: 1. Data availability: Data are already available from Zeincro. The available data should allow for the adequate testing of the study hypothesis and data mining analysis. 2. Data operation: Manual test was conducted in order to verify that the outcome from the data cubes is close, In terms of numbers, to the actual data. If data is available, check whether the data cube generator will produce the semantically enriched and linked data as required. 3. Data analysis: data analysis / mining tools required for this scenario are available on the Linked2Safety platform
Risk Identification and alternative	Given the nature of data and the fact that the datasets do not have many common attributes since they originate from different studies with completely different CRF structure & design, it is possible that the results are not as rich as foreseen.
Assessment Measures and Metrics to Expected Results	Number of eligible subjects successfully identified.

Scenario III.III Depression: Anti-depressant Medicine and Glycaemic Effects (CING)

Scenario Name	Associations between anti-depressant medicine (ADMs) and glycaemic effects in patients with depression
Background and Expected Outcome	Background: It is known from the literature that there exists a positive relationship between a history of depression and insulin resistance syndrome (MetS), but the degree to which ADMs are involved in this association is not well established. Some ADMs were evidenced to have glycaemic effects. For example, tricyclic antidepressants (TCAs) were shown to contribute to hyperglycaemia whereas other ADMs such as Selective serotonin reuptake inhibitors (SSRIs), were not associated with glycaemic effects. Expected Outcomes: It is hypothesized that chemoinformatics analysis in patients with depression that develop glyceamic effects, will cluster together those drugs that do cause this side effect (including TCAs) and identify possible common molecular fragments that could be causing the glyceamic effects. The output of the data mining tools will highlight any associations in the data, between anti-depressant drug sub-structures and glycaemic effects.
Data to be used	For this showcase, data from CHUV will be used. CHUV has data on glyceamic side effects and drug use, including anti-depressant drug use. CHUV data could be enriched with ZEINCRO data if enough ZEINCRO subjects, with available information on blood glucose, also reported ADM use.
"AS-IS" Methods and Tools used (without L2S platform)	Any researcher who would want to study these associations would need subjects with depression. Drug use information should also be available for these subjects. Subjects with depression will be divided into groups depending on the ADM prescribed to them. All groups will be followed for a period of time to determine adverse glyceamic effects. For this scenario, blood glucose measurements will need to be obtained at various time points following recruitment. Whether glyceamic side effects are more common in patients on a particular ADM can then be determined using statistical methods such as chi square test. In addition, a series of logistic regression analyses could be performed with the use of a particular ADM (yes/no) as the exposure variable and development of glyceamic effects as the outcome variable, controlling for sex and age of subjects. The ADMs shown to be causing glyceamic side effects can then be chemoinformatically analyzed to determine molecular

	fragments that could be causing the adverse events. In order to locate and recruit these individuals, researchers would have to contact collaborating research centres and/or hospitals that treat or deal with mental disorders such as depression. Researchers would also have to ensure that drug use data for the same subjects is also available so as to be able to determine drug exposure, separate the groups, and determine sample sizes. If the sample sizes of available subjects from cooperating research centres and hospitals are not enough, then researchers would have to contact more research/hospital centres and/or recruit more individuals from the general population something which would be very time consuming
"AS-IS" Workflow (without L2S platform):	1. A clinical researcher formulates a hypothesis and designs the study 2. The clinical researcher requests approval from the Health Authorities to perform the proposed research/study 3. (After receiving approval) The clinical researcher contacts via e-mail or telephone collaborating physicians, research centres, institutions and hospitals to identify sources of subjects diagnosed with depression with available information on drug use. 4. After identification of the centres that treat/recruit subjects who fit the selection criteria, the clinical researcher will determine whether the available sample sizes – considering the expected participation rate in such a study – are enough to give the study sufficient power so as to make the investigation worthwhile. 5. Given that the sample sizes are enough, the clinical researcher will then request the collaborating hospitals, institutions, and research centres to identify those subjects with available information that are eligible, and invite them to participate in the study. 6. The clinical researcher will recruit those subjects that have responded to the invitation and given their informed consent to participate. 7. These subjects will be divided into "prescribed ADM" groups and followed for a particular time period. Blood glucose and other glycaemic index measurements will be collected at various time points, depending on the study design. 8. Statistical analyses (logistic regression) will be performed to determine the relationship between drug use and blood glucose as well as other glycaemic index levels. The analyses will need to be performed independently for each ADM group and at each time point since adverse reactions

	are drug specific and may take some time to develop. 9. The ADMs shown to be causing glyceamic side effects will be chemoinformatically analyzed to determine molecular fragments that could be causing the adverse events 10. If the number of subjects available through the collaborating institutions is not enough to achieve sufficient power, the researcher may need to abandon the study or recruit more individuals, fulfilling the inclusion criteria, from the general population or from other institutions
Current issues /L2S Improvements	One of the biggest issues in this type of research is the difficulty to recruit enough subjects meeting inclusion criteria to conduct meaningful analysis (power analysis). Indeed, some subjects may drop out of the study, others may be unavailable or refuse to participate, and some subjects may have missing information (unable to answer some questions because of recall problems, etc). In other cases, problems with data collectors are relatively common (difficulty in finding and training enough collaborators to collect data, trained collaborators leaving the study unit prematurely, or providing insufficient work). The potential gain from joining the Linked2Safety platform would be to combine own data with similar data from multiple institutions in order to increase the sample size without having to resort to interviewing more subjects or training additional collaborators. If enough data were available, some study hypotheses could even be tested without collecting any own data at all, using the data which is available on the Linked2Safety platform. In addition, Linked2Safety enables "backward-tracing" of causes of adverse side effects, so that grouping of subjects based on ADM prescribed and separate analyses for each group are not necessary. This increases the sample size and thus power of investigations considerably. More specifically, Linked2Safety enables scientists to query and mine the information on adverse side effects and drug use to identify groups of chemical compounds/drugs causing a particular adverse event. Following the formation of these natural clusters, substructure mining techniques can be applied to identify chemical fragments contained in the molecules of each cluster followed by the overlay of the side-effect profile for the identification of correlations between certain substructures and side-effects.
Objectives as a L2S showcase	There are a lot of studies that have been completed on small groups of subjects, for instance because of study limitations as stated above. It would be beneficial to clinical researchers to be able to address their study questions by combining their data with similar data from other sources. Having enough

	power would enable researchers to respond to questions of high scientific relevance. In addition, using available data from other sources would have an obvious financial gain in that recruiting additional subjects and obtaining additional financial resources would not be needed to answer the study question. The generation of hypotheses, to be used for future statistical research into ADMs and glycaemic effects.
"TO-BE" Methods and Tools used with L2S platform	The data space will allow for the combination of variables (in specific data cubes) related to ADM drug use and glyceamic side effects. These data cubes will be processed in the analysis space using the proposed tests and mining techniques to derive clusters of drugs associated with a particular adverse event. Then the chemical structures of drugs in a cluster will be investigated to identify common molecular fragments that might be causing the adverse events. Data cubes containing the ADM and glycaemic effect variables will be used for data mining, using decision tree, random forests and association rule mining algorithms.
"TO-BE" Workflow with L2S platform	Request the frequency count of subjects with available information on the relevant variables. Determine whether there are enough subjects from the available data sources to test the study hypothesis. Request a breakdown of the frequency table to show which institutions and where, have eligible subjects to test the study hypothesis. Determine whether these data sources are compatible (for e.g. determine possible sources of incompatibility or bias regarding the type of sample (e.g. general population sample versus clinical sample), subject selection (e.g. recruitment of particularly severe cases of depression in hospital settings versus inpatient settings, etc.), and methodology (e.g. the use of self-report versus biological measures). Conduct mining techniques and association testing to determine if and which ADMs are associated with glyceamic side effects in patients with depression and which common molecular fragments might be responsible. Conduct this as follows: Before data mining is run: (Manually) Add rules considered to be as mandatory although they were not discovered by the data mining process (e.g. because the conditions of the rules never occur or because the conditions occur but not lead to any outcome) Once data mining has been run, manual selection of

	association rules (e.g. among a set of variables and a diabetes type II diagnosis) that have more than a given, user-defined, minimum "support" value and those that have more than a given, user-defined minimum "confidence" value. Filter rules obtained from data mining both automatically and by visual inspection. (Manually) Remove rules that do not apply to general population, but to the specific healthcare data analysed The expert user can then query the knowledge base for mined rules and experts' knowledge.
Feasibility Analysis	The feasibility analysis includes the following checks: 1. Data availability: these data are already available from CHUV. The available data should allow for the adequate testing of the study hypothesis and data mining analysis. 2. Data operation: the question regarding whether the data cube generator will produce the semantically enriched and linked data as required will be examined this year. 3. Data analysis: We have the data analysis / mining methods capable to discover knowledge and show benefits.
Risk Identification and alternative	Based on the feasibility analysis above, the outcomes from data analysis might not produce the expected results (no ADMs associated with glyceamic side effects). In this case, it would be more difficult to explain and publish these negative results, and alternative mechanisms that would explain these associations would need to be tested. The data mining outputs, for example association rules, or decision trees, could assist with this.
Assessment Measures and Metrics to Expected Results	Number of eligible subjects successfully identified Enrichment of subject counts from CHUV with counts from ZEINCRO if applicable Successful performance of data mining and clustering of drugs in biologically meaningful and literature supported ways (for example clustering of TCA drugs as a group that causes glyceamic adverse events, confirming reports that have already been published)

Scenario III.IV: Hypercholestrerolemia: ADEs and Simvastatin (ZEINCRO)

Scenario Name	Associations between adverse events and simvastatin in patients with hypercholestrerolemia.
Background and Expected Outcome	A multicenter clinical observational study lasting 48 weeks to investigate the compliance of patients with hypercholesterolemia which are treated with simvastatin 40 mg. The output of the data mining tools will highlight any associations in the data, between Simvastatin drug sub-structures and groups of specific adverse side effect categories in patients with hypercholesterolemia.
Data to be used	For this scenario, Zeincro data on drug use in patients with hypercholesterolemia in relation to the compliance of the patients and the presence of adverse drug reactions will be investigated. This showcase could be enriched by CHUV data. CHUV has information on level of cholesterol and intake of cardiovascular system related drugs (yes/no self-reported). In addition they have a few subjects who took Simvastatin but they could extend their search to other similar treatment.). CHUV could contribute providing enough participants take the drug. CING does not have relevant data.
"AS-IS" Methods and Tools used (without L2S platform)	The method used without the Linked2Safety platform is that the investigator has to recruit subjects with hypercholesterolemia that are in line to the inclusion criteria of the study. In order for the investigator to recruit these subjects he would have to contact collaborating research centres and/or hospitals that treat or deal with patients with hypercholesterolemia. Monitoring of these subjects must be performed throughout the study duration.
"AS-IS" Workflow (without L2S platform):	An investigator formulates a hypothesis for a study Creation of the study protocol, statistical plan and all relevant safety documentation (Clinical safety reports, SAE forms etc) Requesting approval from the Health Authorities and Ethics Committee to perform the proposed study (After receiving approval) The investigator contacts physicians, hospitals/sites and fellow researchers in order to

	identifies sites that can provide the desired subjects Invitation of subjects Recruitment of subjects and retrieval of the inform consent forms The investigator has to proceed with the randomization process (usually by an external vendor) since the study is double blind. The study is being conducted and the subjects are being monitored throughout the study duration Statistical plan is carried out again usually by external vendors Review of outcomes Submission of all documentation to Health Authorities
Current issues /L2S Improvements	Recruitment of subjects is a process that usually can take a lot of time, cost and effort especially if the investigator wants to conduct a clinical trial without external vendors or a sponsor. With Linked2Safety an investigator can have access to a large pool of subjects fulfilling the desired inclusion criteria without having to contact fellow researchers or sites. Additionally the potential gain from joining the Linked2Safety platform would be to combine own data with similar data from multiple institutions in order to increase the sample size without having to resort to interviewing more subjects or training additional collaborators. If enough data were available, some study hypotheses could even be tested without collecting any own data at all, using the data which is available on the Linked2Safety platform. Data mining will be used to generate further potential hypotheses (that might be missed by manual hypothesis creation), and which can be statistically tested in future work.
Objectives as an L2S showcase	This showcase demonstrates the time and effort that is required in order for a clinical trial to be conducted by an investigator. Through this showcase it has been noted that Linked2Safety can make the recruitment procedure a lot easier and cost and time effective through the ability of locating subjects through the Linked2Safety platform. The generation of hypotheses, to be used for future statistical

	research into Simvestatin.
"TO-BE" Methods and Tools used with L2S platform	The data provided by CHUV will be semantically linked with the ones from Zeincro by the Linked2Safety Linked Medical Data Space. Data Cubes generated in closed room with components from the Interoperable EHR data space will be created and processed in the analysis space trough algorithms (such as chi-square). Data cubes containing the Simvastatin and adverse event variables will be used for data mining, using decision tree, random forests and association rule mining algorithms.
"TO-BE" Workflow with L2S platform	Request the frequency count of subjects with available information on the relevant variables. Determine whether there are enough subjects from the available data sources to test the study hypothesis. Request a breakdown of the frequency table to show which institutions and where, have eligible subjects to test the study hypothesis. Determine whether these data sources are compatible (for e.g. determine possible sources of incompatibility or bias regarding the type of sample (e.g. general population sample versus clinical sample), subject selection (e.g. recruitment of particularly severe cases of depression in hospital settings versus inpatient settings, etc.), and methodology (e.g. the use of self-report versus biological measures). Conduct mining techniques and association testing to determine if Simvastatin is associated with which adverse event in patients with hypercholestrerolemia and which common molecular fragments might be responsible. Conduct this analysis as follows: Before data mining is run: (Manually) Add rules considered to be as mandatory although they were not discovered by the data mining process (e.g. because the conditions of the rules never occur or because the conditions occur but not lead to any outcome) Once data mining has been run, manual selection of association rules (e.g. among a set of variables and a diabetes type II diagnosis) that have more than a given, user-defined, minimum "support" value and those that have

	more than a given, user-defined minimum "confidence" value. Filter rules obtained from data mining both automatically and by visual inspection. (Manually) Remove rules that do not apply to general population, but to the specific healthcare data analysed The expert user can then query the knowledge base for mined rules and experts' knowledge.
Feasibility Analysis	The feasibility analysis includes the following checks: Data availability: Data are already available from Zeincro. The available data should allow for the adequate testing of the study hypothesis and data mining analysis. Data operation: Manual test was conducted in order to verify that the outcome from the data cubes is close, In terms of numbers, to the actual data. Data analysis: data analysis / mining tools required for this scenario are available on the Linked2Safety platform
Risk Identification and alternative	Given the nature of data and the fact that the datasets do not have many common attributes since they originate from different studies with completely different CRF structure & design, it is possible that the results are not as rich as foreseen.
Assessment Measures and Metrics to Expected Results	Number of eligible subjects successfully identified. Enrichment of subject counts from Zeincro with counts from CHUV.