European Translational Information and Knowledge Management Services

eTRIKS Deliverable report

Grant agreement no. 115446

Deliverable 3.8

Translational standards report for eTRIKS development - The Standards Roadmap

Due date of deliverable: Month 25

Actual submission date: Month 31

Dissemination Level		
PU	Public	XX
PP	Restricted to other programme participants (including Commission Services)	
RE	Restricted to a group specified by the consortium (including Commission Services)	
CO	Confidential, only for members of the consortium (including Commission Services)	

DELIVERABLE INFORMATION

Project	
Project acronym:	eTRIKS
Project full title:	European Translational Information and Knowledge Management Services
Grant agreement no.:	**115446**
Document	
Deliverable number:	3.8
Deliverable title:	Translational standards report for eTRIKS development - The Standards Roadmap
Deliverable version:	1
Due date of deliverable:	M25
Actual submission date:	M31
Leader:	
Editors:	
Authors:	Paul Houston, Dorina Bratfalean, Philippe Rocca-Serra
Reviewers:	Chris Marshall, Peter Rice
Participating beneficiaries:	
Work Package no.:	3
Work Package title:	Standards Research and Coordination
Work Package leader:	Michael Braxenthaler, Paul Houston
Work Package participants:	
Estimated person-months for deliverable:	1 PM
Nature:	Document
Version:	1
Draft/Final:	Draft
No of pages (including cover):	13
Keywords:	Data standards

Purpose

As set out in the Description of Work, WP3 aims to identify, develop and adopt data standards and controlled terminologies for translational researchers using the eTRIKS Platform and in the wider community. The consistent application of standards to a scientific discipline is a prerequisite for meaningful exchange of information and combining of data from multiple sources.

In this document, we review the current state of the translational research standards landscape and describe a roadmap for the identification, adoption and dissemination of standards for translational research data entities that are not clinical in origin.

Intended audience

The readership of this document is assumed to be familiar with eTRIKS and its overall aims, including being aware of the work completed to date with respect to the tranSMART for eTRIKS software.

Contents

Overview of clinical and preclinical studies in translational medicine

Translational research refers to two distinct domains: T1 research, the "bench-to-bedside" enterprise of translating knowledge from the basic sciences into the development of new treatments; and T2 research, translating the findings from clinical trials into everyday practice[1]

Translational standards should focus on the patient and allow for closer and rigorous control of patient data whilst creating a data ecosystem that supports with appropriate controlled vocabulary and biomarker selection algorithms. As an example the CDISC and CFAST therapeutic area standards include biomarkers designed to reduce the risks and to improve safety and efficacy of drug development programmes.

'Translational' may also refer to making scientific inferences from 'animal models' to 'clinical' or 'in man' studies. While exploratory clinical studies are designed to demonstrate that the biomarkers are associated with therapeutic area (e.g. asthma) in humans, it is important to show that the biomarkers can also detect disease states in animals. Therefore, ''reverse translation'' could be performed by testing the biomarkers in preclinical studies.

Work Package 3 aims to show the importance of translational standards, which assure that the data is consistent, appropriate, accurate and processed correctly, ensuring data integrity and increased interoperability between data sets of similar disease groups and to allow for knowledge gains between preclinical and clinical studies.

1. Woolf SH. The Meaning of Translational Research and Why It Matters. JAMA 2008;299;211-213

A. Clinical and non-clinical research

The objectives of WP3 is to guide and support more streamlined research processes from protocols design through to analysis by recommending the implementation of appropriate standards.

The CDISC modules are able to facilitate consistent and efficient data aggregation to glean greater knowledge from cumulative data assets contributed by patients for the advancement of research.

The CDISC standards and the terminology recommended by Work Package 3 to support eTRIKS projects is provided in D3.4.

Two panels of research in translational medicine is presented below in figure 1.

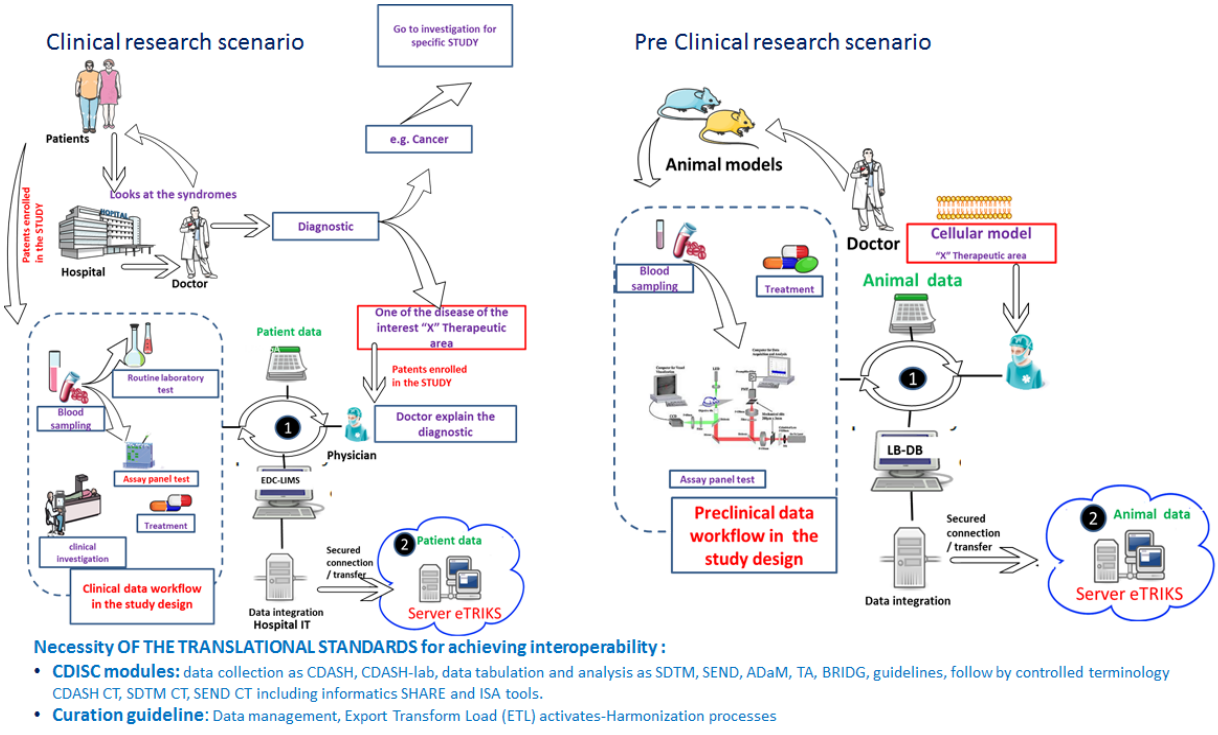


Figure 1. CLINICAL & NON CLINICAL RESEARCH SCENARIO

B. Overview of implementing standards in translational medical research on eTRIKS project

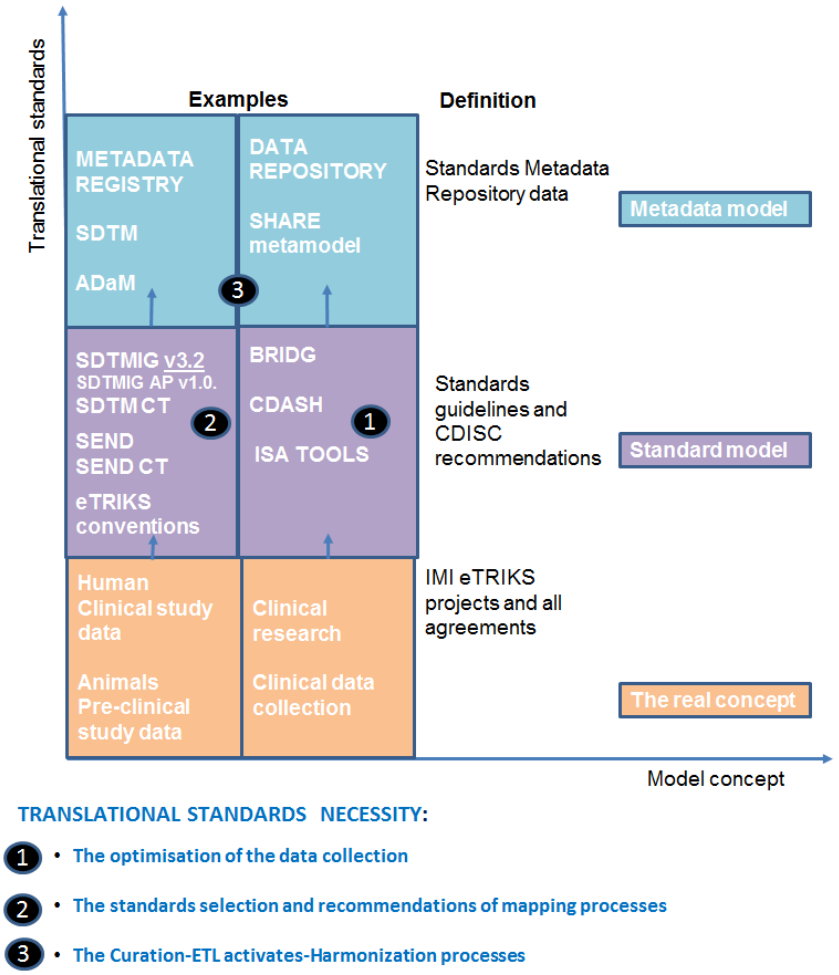
CDISC standards enable pooling of diverse sources of clinical data to address research questions and look for subtle signals in large data sets. The CDISC Mission is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare.

The genomic standards are widely disparate and are not integrated like the CDISC standards. The ISA Tab templates and ISA metadata tracking tools help to manage an increasingly diverse set of life science, environmental and biomedical experiments that employ one or several combinations of technologies.

Workpackage 3 are working towards implementing a Meta Data Registry that will allow for an increasingly seamless application of standards from the clinical (CDISC models) and the genomic (ISA and CDISC models) domains allowing for interoperability and comparability of data.

ETRIKS standards model concept

The diagram presented in the Figure 2 illustrates the process of standard selection where WP3 can guide and support more streamlined standards selection processes from protocol models through to complete study lifecycle metadata models.



http://www.cdisc.org/

Figura 2. Translational Standards for the eTRIKS PLATFORM

C. Overview of statistical data exploration in translational medicine

Translational standards should confirm if the data follows the previously defined statistical requirements. The curator-tabulated data follows the standards recommendation (e.g SDTM for clinical studies, SEND for preclinical and LAB model and ISA-Tab for OMICS). Based on the data model the statisticians should be able to more easily handle the data allowing for the creation of new subsets of data for the exploratory statistical analysis phase.

The majority of studies supported by the eTRIKS project are exploratory such as: observational and/or interventional studies. The nature of the protocol is exploratory and there are many unknowns. Therefore, some of the predefined analyses described below, in Figure 3,

will likely be modified to some degree in the light of the data. Also, further analyses may be performed as questions raised by the data present themselves.

Statistical exploratory methods in translational medicine

The exploratory methods, on the basis of various criteria will be used to score the biomarker candidates. The datasets needed in exploratory methods will be populated with the existence of published data in preclinical and clinical settings. The tabulated data from clinical or preclinical research will include the clinical or preclinical parameters related to the feasibility of appropriate sampling or of large-scale biomarker measurement, with the respect of intellectual property status.

The exploratory data analysis summarizes the characteristics of datasets usually by visualisation methods. These statistical exploratory methods encourage statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiment. This raises the question of how one would confirm the utility of biomarkers that may have been initially discovered and evaluated only in preclinical animal models but that have not been evaluated in relevant human populations.

The panel of exploratory methods in translational medicine is presented below in the figure 3.

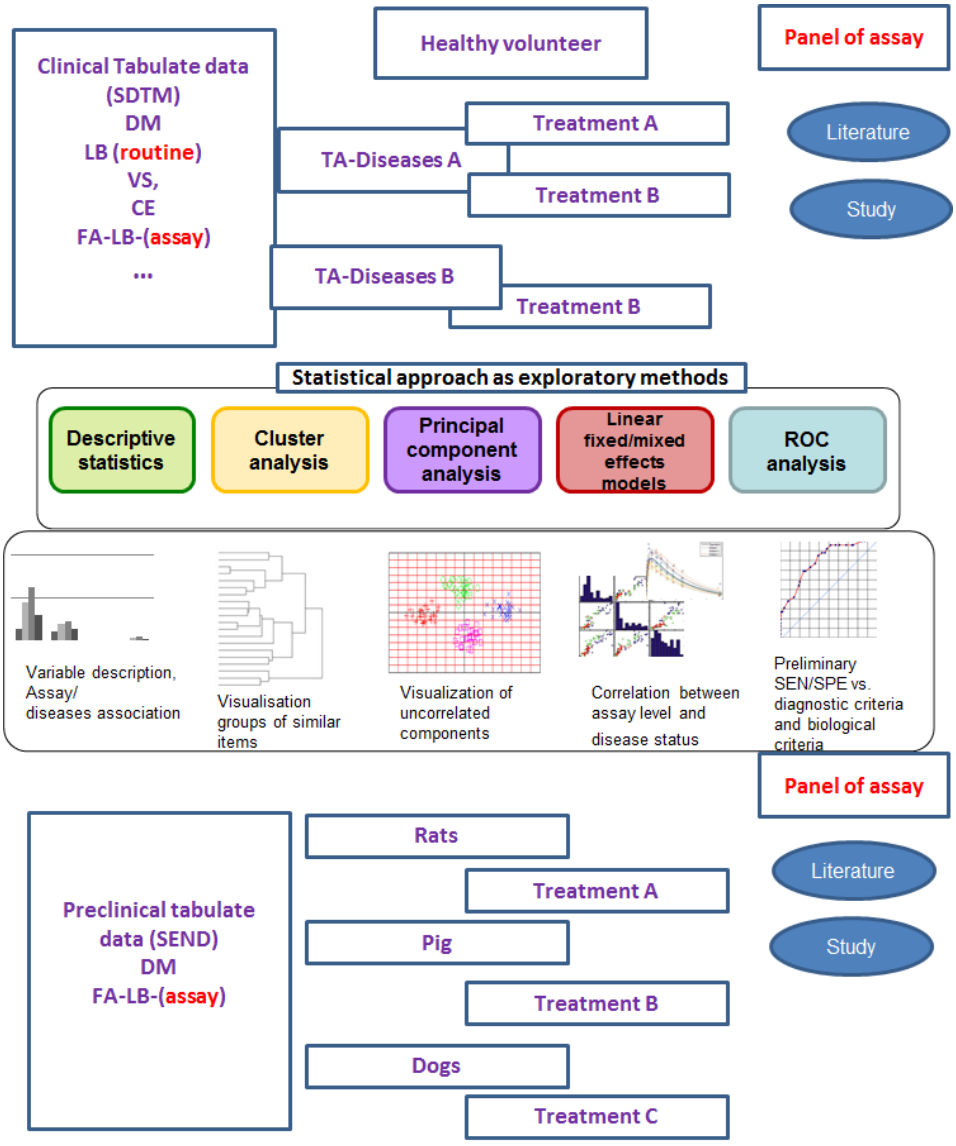


Figure 3. STATISTICAL EXPLORATORY METHODS IN TRANSLATIONAL MEDICINE

D. WP3 - The translational standard effort in knowledge management:

The 'starter pack' (http://goo.gl/iUAylz) fully describes existing standards and the potential benefits of applying them properly so real knowledge management gains can be achieved. In summary work package three will be working to:

- Further develop the value proposition for all stakeholders involved in the eTRIKS process
- Define the responsibilities set out from the start of a project, also agreeing with other eTRIKS work packages have agreed and integrated processes.
- Examine current standard implementations in IMI projects and apply CDISC-eTRIKS standards
- ETL data transformation and harmonization processes platform

The Standards Roadmap

The work package 3 activities have been divided between five workstreams. (see considerations and issues below).

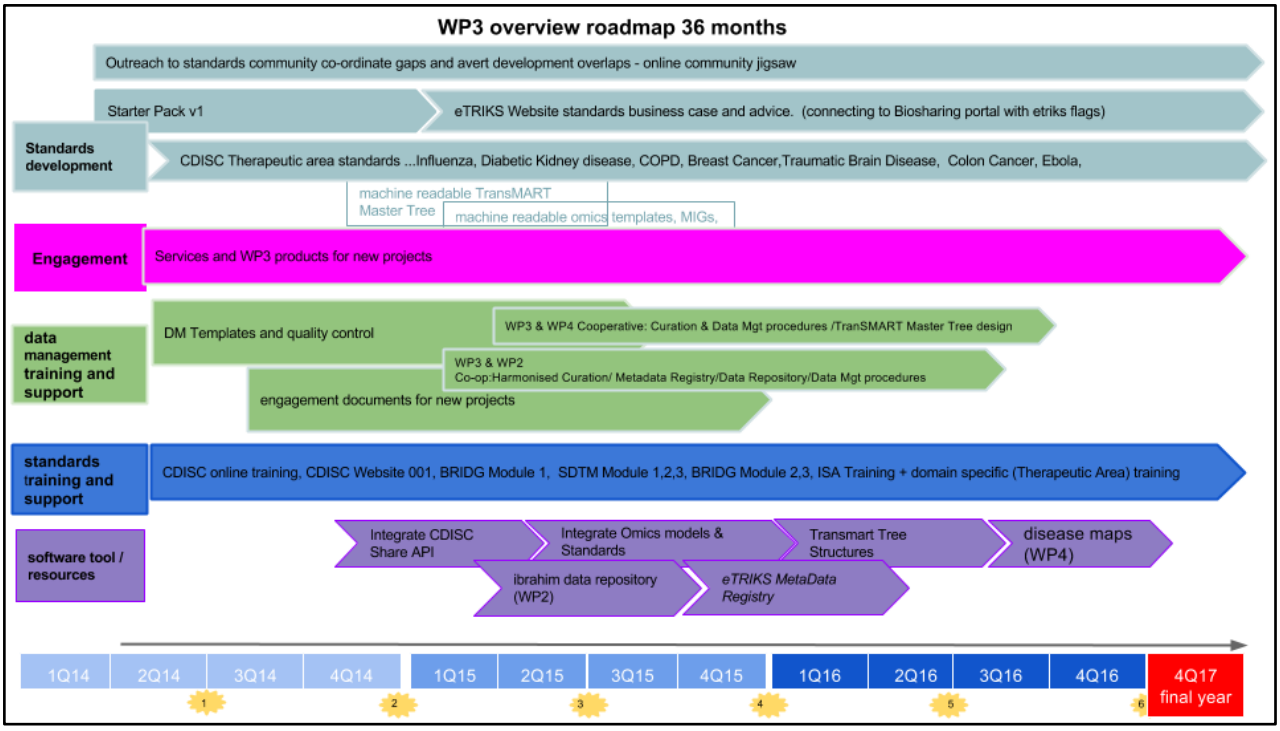


Figure 4. Draft WP3 ROADMAP

eTRIKS WP3 members are acutely aware of the importance to demonstrate impact of WP3 work on eTRIKS supported projects. The Work Package 3 strategy, where possible, will

employ mechanics such as a Meta Data Registry (MDR), which will build metadata and variable profiles for future projects to provide a level of automation to increase harmonisation of data. Such mechanics combined with a comprehensive set of written working procedures that have been agreed between work packages 3 and 4, will create a cohesive and efficient standards and data management system within the eTRIKs infrastructure. The Work Package 3 streams of activity together with the proposals below focus on improving the quality and homogeneity of data so the curation bottlenecks will diminish and related data sets will become more comparable.

The risks of not implementing such MDR strategies would mean that the standards implementation would be degraded and in turn the ability to make effective analysis or cross comparison of the datasets in eTRIKS would be compromised. It is for the very reason of 'wiggle room' in the implementation of standards that many companies, including CDISC, have implemented MDRs so that the interpretation and implementation of standards is consistent and the resultant data is of the highest quality achievable, fit for purpose and fit to compare with like data or related medical data sets.

Work Package 3 acknowledges that the development resources of eTRIKS are limited and that this roadmap and its proposed developments will have to be weighed in value against other eTRIKS developments that will compete for the resource budget. The proposals below therefore try to factor in cost effective collaboration, or adoption of existing technology options. The options are by no means exhaustive, any alternatives or adaptations are welcomed. Upon the agreement to move forward with any of the proposals below, or any agreed variation of, Work Package 3 could then carry out more detailed analysis and requirements gathering alongside any other eTRIKS work packages that need to be involved.

Considerations and issues:

- Without automated standards mechanics:
    - there will be a high impairment in the compliance and adoption of standards
    - less enforceable data harmonization across projects and lower data quality
    - no sustainable and scalable support of curation for projects

    - no mean of building, storing and re-using the eTRIKS curation know-how for curating data of new studies
    - weak mean of "negotiating" eTRIKS feedback with standards organizations.

- The lack of access to project data still represents a major impediment to WP3 efforts.

- The enhancements required in TranSMART to meet the needs for effective implementation and management of standards need to be included and prioritised in the development worklist for WP2 and WP4. This issue was recognised and discussed at the Barcelona Annual Meeting and a new working group including WP2, WP3 and WP4 has been formed to coordinate the development of eTRIKS across the needs of all Work Packages (called the eTRIKS Harmonisation System, eHS). http://www.etriks.org/blog/etriks-2nd-annual-meeting/

- WP3 resources needs to be prioritising across the parallel developments, Starter Pack, MDR, Project Engagement, Outreach etc. Resource from Roche is being redirected to WP3 to help alleviate shortage of resource here.
- Quality control automation has still to be looked at as part of a coordinated approach to improving data quality. This has been made a priority in the focus of the eHS.

- U-BIOPRED data mapping exercise to CDISC has been extensively carried out by CDISC experts (Dorina and Lauren). However, as Data access is being restricted, it presents WP3 with difficulties. The metadata may be released for further exploration and analysis.

- A strategy of preparing data standardisation processes has been developed by an expressive model, and follows the next steps:
    - future standards development would be driven by analysing the contents of the eTRIKS metadata registry (eMDR) which would be informed by existing projects meta data templates.
    - annotated datasets by specification rules
    - validation process of data standardization

- To provide an analysis of the main elements of the data management policy including reference to existing suitable standards. If these do not exist, an outline on how and what metadata will be created. The data management procedure (data flow, data management plan, data transfer procedure, data annotated) documents are being drafted and need to be finalized before being reviewed by the eTRIKS consortium.

- Need to finalise the development and/or implementation with WP4 of the extract, transform and load (ETL) activities based on use case. We would expect to implement and test the first scenario for one use case by 2015.

- Establish the general outline of eTRIKS policy for data management. The described policy should reflect the current state of consortium agreements regarding data curation and/or management and be consistent with those referring to exploitation and protection of results

- Establish the harmonisation processes across the eTRIKS studies and /or projects with WP4 to get homogenous data within tranSMART database.

- Establish the procedure and develop an CDISC standard model to display am TranSMART Master Tree in collaboration with WP4.

- A risk mitigation plan is required to take into account the possibility that eTRIKS cannot deliver a functional metadata registry. The current roadmap intends such tool to be available at the end of 2015 assuming Pharma partners can release entirely or partially existing components as discussed during the annual eTRIKS meeting. WP3 assumes it will be able to rely on those components, to build on top of them and deliver this critical infrastructure component. However, there is currently no provision, nor software development resources allocated for developing a metadata registry from the ground up, should the release in the open source domain of existing solution do not materialize. WP3 is aware of similar effort being developed by the Hyve relying on Protege tool under the EMIF project. WP3 will therefore maintain close contact and follow progress of various efforts in the field.