European Translational Information and Knowledge Management Services

eTRIKS Deliverable report

Grant agreement no. 115446

Deliverable 5.3: ROI Evaluation Report at Five Years

Due date of deliverable: 1st January 2017

Actual submission data: 14th September 2017

Dissemination Level: PU

Dissemination Level		
PU	Public	PU
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:	5.3
Deliverable title:	ROI Evaluation Report at Five Years
Deliverable version:	1.0
Due date of deliverable:	Month 51
Actual submission date:	Month 58
Leader:	Jay Bergeron, Scott Wagers
Editors:	David Henderson
Authors:	Scott Wagers, Neil Fitch
Reviewers:	Chris Marshall, Jay Bergeron
Participating beneficiaries:	U Luxembourg, Imperial College, BioSci Consulting
Work Package no.:	5
Work Package title:	Program Management
Work Package leader:	Jay Bergeron, Scott Wagers
Work Package participants:	C. Auffray, J. Bergeron, Y. Guo, D. Henderson, C. Marshall, M. Saqi, R. Schneider, S. Wagers, A Meiser, P Peeters
Estimated person-months for deliverable:	1
Nature:	Update on D5.2, which was associated with IMI Mid Term Review
Version:	1.0
Draft/Final:	1.0
No of pages (including cover):	13
Keywords:	eTRIKS, ROI

Contents

Executive Summary

This report is an update of D5.2, the ROI Evaluation Report at Two Years, presented at the eTRIKS mid-term review. The mid term review was conducted at the IMI offices in Brussels on 21-May-2015, with the formal written outcome presented to the Coordinator and eTRIKS Executive Committee (ExCom) on 6-June-2015. D5.2 included a number of points related to the mid term review that did not pertain only to the ROI. This report (D5.3) will focus on the ROI specific activities and achievements.

In the mid-term review the reviewers acknowledged that the eTRIKS project has been proceeding according to its *Description of Work* (DOW) and that both personnel and financial resources have been allocated in a manner that is consistent with the program's objectives. This opinion has continued during following periodic reports and conversations with IMI representatives, who have also attended eTRIKS cross consortium meetings.

Since D5.2, the development of an *open translational research data management platform (eTRIKS v1),* which was lauded as a substantial achievement in support of the research goals of eTRIKS' clients as well as the wider translational research community, has been substantiated by a number of iterations. The latest version of this tranSMART data warehouse system is eTRIKS v4. The operational synergies are arising from the comprehensive set of information services that eTRIKS provides to client projects both within, and beyond, the IMI, has been proven by the number of supported projects. At the mid term review, reviewers had cautioned that substantial challenges remained in meeting the eTRIKS' objective to support forty client projects. 13 projects were supported at that time. However, eTRIKS has been able to support 61 projects, a substantial achievement. As requested by the reviewers, additional effort was put into engaging and recruiting project clients.

Similarly, the eTRIKS *Public Server* was cited as a welcome extension to eTRIKS' planned deliverables and the *Standards Starter Pack* and *Guidelines for Secondary Reuse of Clinical Data* were commended as broadly applicable contributions to the EU scientific community. These outputs have been delivered and are in use within the community.

Other key indicators that support the success of the eTRIKS consortium have emerged since the mid-term review. These include the establishment of a standard CDA/MTA template for large scale consortia and effecting risk mitigation planning. The difficulties associated with concluding cross-collaboration Confidential Disclosure (CDA) and Material Transfer Agreements (MTA) were recognized to be serious impediments preventing eTRIKS from interacting with client projects. A legal group was established that processed issues and allowed support for projects to move forward through creating less legally demanding eTRIKS products, that did not require data to be hosted by eTRIKS. Further, a more rigorous project management approach was implemented and has assisted in reaching our goals.

To capitalise on the eTRIKS investment, and on the reviewers' recommendation, two new deliverables were agreed upon during the 2015 eTRIKS General Assembly meeting, to disseminate the eTRIKS experience and lessons learned.

1. **D5.7: End of 2015:** A deliverable to document the contracting difficulties experienced by the program.
2. **D5.8: End of Project:** A deliverable to document and disseminate expertise regarding the services created for eTRIKS.

The reviewers were uncertain concerning the prospects of the eTRIKS Data Science Network and requested that the sustainability models be more fully detailed. These have been drawn up and the eTRIKS network is ready to operate.

With regard to forward looking activities, the reviewers enthusiastically endorsed developing the eTRIKS Harmonization Service (EHS) and the accompanying Meta Data Repository (MDR). Both services have been created and are in use by select projects.

Inputs and Outputs from Related Deliverables

Inputs

As an evaluation of the total investment of the IMI into eTRIKS', all project deliverables are, essentially, inputs. However, the periodic report deliverables 1 to 5, submitted in Sept 2013, Oct 2014, Oct 2015, Oct 2016 and Oct 2017 are direct inputs to this ROI evaluation.

This deliverable was put forward under the guidance of WP5 with input from all work package group leaders.

Outputs

The following specific outputs arise from this deliverable. There are no related or dependent deliverables.

1. Operation of the eTRIKS Data Science Network, built on the proven value of the eTRIKS outputs.

Description of Work Achieved

Summary of Accomplishments

eTRIKS offers a comprehensive open data management platform and accompanying best practices to public private partnerships, consortia and collaborative research groups that engage in translational research. The system is designed to facilitate management and analysis of exploratory clinical and multi-OMICs data. eTRIKS has engaged sixty-one client programs at the time of this writing, providing software development, hosting, data curation and data import services. Additionally, eTRIKS has trained over 100 investigators in the use of the eTRIKS platform and detailed CDISC-based data standards to facilitate the reuse and integration of translational datasets. eTRIKS has assembled and promotes guidelines to foster the ethical use of human data for exploratory research and maintains an open access portal containing, at the time of this writing, close to 200 curated public domain translational studies. Although eTRIKS has been challenged in closing Confidential Disclosure (CDA) and Material Transfer (MTA) Agreements, the collaboration has responded by establishing a mandate allowing the coordinator (Astra Zeneca [AZ]) to authorize confidentiality agreements with client projects and developing an MTA template to facilitate multi-party negotiations.

The eTRIKS platform is based on the popular open source *tranSMART* translational research data warehouse. Modified by eTRIKS to incorporate the open source PostgreSQL relational data base, this fully open source tranSMART platform, initially released as eTRIKS v1.0/tranSMART 1.1 and currently revised and published as eTRIKS v4.0/tranSMART 16.2, has become a prominent translational research system for both academic institutions and commercial entities alike.

To augment the eTRIKS platform, mature operational practices have been developed for application release, hosting, data curation/loading and client training/support. Business development/outreach practices have been established to promote eTRIKS to prospective clients. The result is a comprehensive suite of products and services that enable investigators to curate and process complicated and variable clinical and molecular data elements in a manner that is consistent, comprehensive, economical and open to inspection.

Motivated to sustain these best practices beyond the completion of the collaboration, eTRIKS conceived and launched the *eTRIKS Data Sciences Network (eSDN)* which assembles cooperatives of eTRIKS-enabled service providers to respond to the translational information needs of large collaborative research projects on a fee for service basis. From training to hosting to program management, the eSDN provides both comprehensive and targeted information services tailored to fit the needs of translational research programs. By month sixty, the eTRIKS network had signed data curation and hosting contracts with two end term projects utilising the company ITTM (*Information Technology for Translational Medicine S.A.*), a commercial spinoff arising from eTRIKS and the U. of Luxembourg, as a provider of eTRIKS services.

Program Structure

eTRIKS is overseen by Astra Zeneca (Coordinator) and Imperial College London (Managing Entity) with the assistance of an Executive Committee (ExCom) having representation from Biosci Consulting, Imperial College, U. of Luxembourg, EISBM/CNRS, Astra Zeneca, Bayer and Johnson and Johnson to month 48 and thereafter Pfizer to month 60). The ExCom is accountable for ensuring that the project executes per the vision of the partners as put forward in the eTRIKS *Full Project Proposal* and *Description of Work (DOW)*. The ExCom is supported by a Program Management Office (Biosci Consulting) and a Scientific Coordinator (Pfizer).

All eTRIKS participants share equal influence with regard to changes to the program via General Assembly (GA) meetings that are held at the annual meeting and otherwise as required. For example, amendments to the *Description of Work* require a General Assembly majority vote.

eTRIKS is comprised of seven work packages (WP) aligned with the key efforts of the consortium. Work packages actively collaborate to promote the development of comprehensive solutions via recurring cross-WP meetings. The *"Delivery Package"* formalizes cross-WP working relationships for sizable deployments to clients, including multi-faceted support for select IMI projects such as U-BioPRED, ABI-RISK and OncoTrack. Each WP is led by representatives from Pharma and the Beneficiaries. As a group, the WP leaders meet by phone monthly and face to face thrice annually to ensure coordination of effort and satisfactory progress towards goals.

eTRIKS also participates in a wider translational research informatics community that includes the tranSMART Foundation(a U.S. non-profit responsible for coordinating tranSMART development and the larger tranSMART community) and the CTMM TRaiT consortium (a Netherlands-focused public private partnership, similar to eTRIKS, that promotes translational research information best practices and also uses tranSMART as a base software system). eTRIKS' academic leader and scientific coordinator have served in leadership roles with respect to the tranSMART Foundation. Moreover, eTRIKS and CTMM TRaiT personnel interact closely, meeting regularly to promote cooperative interaction.

- Note that the tranSMART Foundation reformed under new ByLaws and structure as the tranSMART/I2B2 Foundation on or about May 2017 consolidating the I2B2 and tranSMART

community efforts under a single non-profit coordinating entity. eTRIKS' academic leader acts as CTO of the tranSMART/I2B2 Foundation.

eTRIKS Project Challenges

Timely prosecution of CDA and MTA agreements

Closing CDA and MTA agreements has been highly problematic given the need for individual review and authorization by each Participant within both eTRIKS and the client project. The iterative rounds of refinement have led to turn-around-times of many months for CDAs and months/years for MTAs. In order to expedite CDAs, eTRIKS partners signed a mandate allowing the coordinator (AZ) to authorize CDAs on behalf of all eTRIKS partners provided that the CDA conforms to a pre-authorized template. MTAs have proven especially difficult given the variability in legal philosophy across partners and national borders. eTRIKS has developed an MTA template for use by new projects. However, as of month 30, agreements between specific data handlers (organizations *and* individuals) were put in place to promote data transfer. The difficulties associated with MTAs have prompted eTRIKS to consider using alternative IMI *Data Processing* agreements or to simply abandon cross-consortium MTA agreements altogether. During the final two years of the collaboration MTA's between eTRIKS and ABI-RISK as well as eTRIKS and Oncotrack were executed based on the templates developed by eTRIKS serving as case examples as to how such agreements can be pursued by future public private partnerships.

Multi-Instance Implementation

Related to the prosecution of CDAs and MTAs, project-specific data use constraints have forced eTRIKS to create a distinct software instance corresponding to each supported project. A multitenant environment in which security and access would be maintained by eTRIKS was considered an ideal outcome at the start of the project. The multitenant instance was to serve as a publically accessible source for IMI data following project completion. The public server certainly has become such an open platform for public domain data although the vision of a single physical translational environment was not achievable due to perceived risk with regard to data protection. eTRIKS is planning for post project data sustainability via UoL who also host an ELIXIR node.

Scaling Best Practices

The number of projects that eTRIKS has been able to support was limited by the number of eTRIKS staff members who can apply informatics best practices on behalf of clients. These practices include client engagement, data curation/import, systems training, coordination of legal agreements, custom software/analytics development and consulting. These services are provided comprehensively to certain clients, such as U-BIOPRED. However, it became clear that providing such comprehensive support to all potential clients would severely limit the number of projects for whom eTRIKS could provide services. Recognizing that many prospective eTRIKS clients have very specific information management requirements and that informatics expertise was often available among the client's staff, eTRIKS developed a targeted engagement model designed to provide select data management services tailored to address the priority needs of individual clients. Moreover, eTRIKS enables client staff in the performance of these best practices allowing the clients themselves to sustain the best practices with their existing staff members. The targeted delivery model makes more efficient use of the limited resources available and has allowed eTRIKS to support 63 client projects.

Resource Allocation

eTRIKS v7 will be developed by ICL in the coming period. In parallel UoL will host the public server, continuing to offer eTRIKS products to clients.

Strategy

With an emphasis on research and operations, the following strategic elements will be eTRIKS priorities during the remainder of the project (Period 6).

1) Significantly improve the sustainability of eTRIKS by development of a consolidated version 5 of the eTRIKS/tranSMART platform with the following advanced features compared to eTRIKS4
   a) Improved performance by building on tranSMART v17.1
   b) Support of longitudinal data
   c) New, more usable graphical interface
   d) Extension to the eTRIKS platform to incorporate the eTRIKS Harmonization and Integration platform (eHS) and the eTRIKS Analytical Environment (eAE) with tranSMART 17.1 to make a uniform eTRIKS system
   e) Enhancement of built-in tranSMART SmartR workflows.
2) a) Promote eTRIKS/tranSMART v5.0 as a platform that fosters collaboration between data scientists and translational researchers through specific use cases that utilize the different components developed as part of eTRIKS. Examples are the U-BIOPRED and PRECISESADS projects, for which there are rich data sets much of which are yet to be analysed.
   b) Broaden adoption of the eTRIKS Labs.

Period 6 will be limited in scope, and concentrate on the work of WP2, with WP5 and WP6 supporting the core development activities with management, dissemination and sustainability actions. All existing tasks have been completed by end of period 5 and the objectives above relate to new tasks.

Operational Excellence

New projects will be able to be supported by the eDSN, and the eTRIKS tools will be available for use. The same rigour seen in the project period will be applied in the eDSN newtwork, with the focus on delivering quality service that expands the ability of translational research projects to deliver their goals.

Sustainability

Development of the eDSN will continue with a focus on growing customers and delivering fee based informatics services by suppliers committed to eTRIKS best practices. BioSci Consulting is leading the business development activities as detailed in the Deliverable report D5.5. The eDSN is planned to continue long after the eTRIKS project ends. eTRIKS will continue to work with the tranSMART/I2B2 Foundation with respect to development and delivery of new eTRIKS/tranSMART software where there is the promise of mutual benefit.

Return on investment

This report (D5.3) updates the original ROI calculations provided in D5.2.

Return on investment (ROI) is a financial concept defined mathematically as:

$$ROI = (Profit - Cost)/Cost$$

When dealing purely with the sale of a product the calculation of ROI is straightforward. When the impact of an investment is more long term or more broadly distributed it becomes more difficult to calculate ROI. When you do not know the Profit you have to make estimation. It becomes even more

difficult when the return is more in terms of a strategic or social benefit. There is precedent for calculating a social ROI [1]. A social ROI uses the same formula with a 'monetised' social benefit in place of Profit.

The main point of calculating an ROI is to provide a quantification that will help in decision-making. Of course the more estimations one makes the less reliable the quantitation will be. However, physicists make estimations routinely, using order of magnitude, or Fermi estimates. The principle is that you estimate the order of magnitude of the number you are uncertain about. It is easier to make an estimate based on orders of magnitude. To increase the accuracy you break down the estimate into a series of estimates that affect that input to the estimate you are trying to obtain and put in real numbers when they are available. By combining a series of estimates you have a better chance of arriving at a more accurate estimate. These types of estimates are being applied to help make business decisions[2].

Many of the benefits from the project are difficult to quantify as the impact of the benefit is wide and data is not available for all the various benefits and can only be estimated. For the most part, however, it is possible to have estimates that are better than an order of magnitude. The main point of this calculation is to determine if the eTRIKS project has had a positive ROI and whether it will continue to have a positive ROI.

Table 1 summarizes the estimates used to monetize the benefits from eTRIKS at 5 years. The left hand column includes the assumptions that determine the inputs. The calculated costs are highlighted in grey.

Table 2 summarizes the return on investment calculation made at the ½ way point, the projected project return at the end of project from D5.2 and the revised calculation at the 5 year point.

[1] Millar & Hall (2012) Social Return on Investment (SROI) and Performance Measurement. In Public Management Review, DOI:10.1080/14719037.2012.698857, p.4

[2] http://lesswrong.com/lw/h5e/fermi_estimates/

Table 1

Variable	Value	Basis for estimation
Deploy a translational research knowledge platform	768.000	This is based upon the estimate that it would require 4 people working for 2 years to deploy and customise a translational research knowledge management system for a consortium. The person month cost was estimated at 8.000 euros, a reasonable estimate for a higher valued bioinformatician and developer. The original budget for knowledge management for U-BIOPRED was of a similar magnitude.
Resource for ETL	96.000	For ETL the calculation is that the typical project required 1 person working full time for a year at 8.000 euros/month
Number of projects supported at 5 years by eTRIKS	61	This is the number of projects supported in some form or another by eTRIKS.
Number of projects supported with full ETL support	4	These are the projects where eTRIKS was primarily responsible for ETL
Cost of a statistician to perform exploratory analyses	2.000	This is based upon an estimate of 1 week of effort to obtain data and perform statistical analyses to answer simple statistical exploratory analyses
No. of exploratory queries per project	10	This is a FERMI based estimation using the rationale that 100 exploratory queries would be too many, so the next order of magnitude lower, 10, was chosen
Cost of software development	1,080.000	This is based upon 3 FTEs working full time for 5 years at an FTE cost of 6.500/month
Reduction in resource realised when standards are used	50%	Estimate based upon the fact that even if standards are used there will be some ETL effort, but that will be reduced significantly because of readily available mappings and a reduced need to spend effort in gaining an understanding of data types
Reduction in effort to understand data re-use principles and regulation	8.000	Assumed to be 2 months of effort to understand and set up processes for data re-use compliance. A reduction of 50% of that effort would be realised if the code for the re-use of medical ethics is used. Person month cost 8.000 euros
Number of projects benefitting from the code of medical data re-use	100	The code of re-use has been downloaded a number of times. It is not possible to know for certain if it is being used, however training has been delivered to over 100 organisations. Again based on FERMI estimation principles the number 100 was chosen
Number of projects expected to benefit from standards at 5 years	272	This is based upon a 20% up take rate for 1.000 new non IMI projects over the past 2.5 years and uptake by the 63 additional project supported by eTRIKS
Planned number of projects supported by the end of eTRIKS	40	This is the goal number of projects eTRIKS aims to support by the end of the project
Software development	2.160.000	This is calculated as a doubling of the development estimate

Table 2

1/2 point return summary		Projected 5 year return summary		Actual 5 year return summary	
Curation supported projects	3.456.000	Standards benefit	10.800.000	Standards benefit	16.200.000
Non curation supported projects	9.504.000	Medical Re-use guidelines benefit	800.000	Medical Re-use guidelines benefit	1.200.000
Exploratory queries	300.000	Support of 40 projects	31.104.000	Support of 61 projects	46.656.000
Software development	540.000	Software development	1.080.000	Software development	1.620.000
Medical Reuse guidelines	80.000				
Total return at 1/2 way point	13.880.000	Return at end of project	43.784.000	Return at end of project	65.676.000
Total spend at 1/2 way point	7.831.287	eTRIKS total budget	20.890.728	eTRIKS total budget*	20.890.728
Differential	6.048.713	Differential	22.893.272	Differential	44.785.272
ROI	77%	ROI	110%	ROI	214%

*Total budget is used for reference, including Period 6 activities

Benefit to Pharma

Assessing the benefit of eTRIKS with repect to the Pharma partners is critical. The following chart maps twenty public private partnerships supported by eTRIKS relative to the corresponding participating eTRIKS pharma partners. The top six highlighted projects were provided comprehensive support by eTRIKS. Pfizer, Roche, GSK were associated with at least half of the projects while Sanofi and Janssen participated in eight projects each. Although Bayer was only associated with four of projects and Lilly with five, each company was participated in three projects that benefitted from comprehensive support from eTRIKS. By contrast, Lundbeck (0), Lilly (5 total, 2-comprehensive) appear to have received the least advantage from eTRIKS. This chart helps to explain the relative contribution from different partners and demonstrates that Pharma partners received benefit, in some cases substantial benefit, in a manner that is unlikely to be explicitly reflected in annualized company/departmental ROI.

	Sanofi	Roche	AstraZeneca	Janssen	Bayer	Merck	Lilly	Lundbeck	Pfizer	GSK	
UBioPred		x	x	x						x	4
ABIRISK	x				x	x			x	x	5
Oncotrack		x	x		x	x	x		x		6
PreciseADS	x				x		x				3
Predict-TB	x			x						x	3
Approach						x				x	2
COPD MAP			x						x	x	3
SAFE-T	x	x	x		x		x		x	x	7
BioVacSafe	x									x	2
Matura		x		x					x		3
RA-MAP		x	x	x					x	x	5
Aetionomy	x										1
Quic-Concept	x	x	x			x	x		x	x	7
BioAster	x										1
EMIF		x		x		x			x	x	5
ND4BB	x		x	x						x	4
Masterplans		x									1
PSORT				x					x	x	3
EU-AIMS		x		x			x		x		4
Genomics England		x	x							x	3
Comprehensive	3	2	2	2	3	3	2	0	2	4	
Total	**9**	**10**	8	8	4	5	5	0	**10**	**13**	

Conclusions

The eTRIKS project has delivered a positive return on investment of an estimated 214% at the 5-year mark of the project. The project activities will further continue and be reinforced by the period 6 activities, which will be delivered for no additional cost. This highlights one of the most important value propositions of working in a collaborative project – it is more cost efficient. This is not always the outcome. It has been highlighted that there is often a perception that collaborative consortium based projects are less productive than expected.

The positive ROI is driven largely by the fact that eTRIKS is providing an open source platform to projects that would otherwise assume very costly efforts to develop and deploy their own custom systems. The ROI outcome emphasizes the efficiency gain that is realized from working in a standardized manner with a common set of tools. This type of efficiency gain does require consistent effort from not only those deploying the system, but also from those using the system. Conforming to standards takes effort. As this analysis demonstrates it is a worthwhile effort. However, what is not reflected in this analysis is the value gains from combining and re-using datasets which efforts to transform and load data such as eTRIKS make possible.

The past 2.5 years have seen the project significantly exceed the ambitious targets set at the mid-term review, particularly in supporting 63 projects vs the desired goal of 40. The re-enforcement of the core deliverables has also ensured that quality support and quality products have been achieved. We expect the ROI to continue to improve as eTRIKS initiatives the eTRIKS Network and sustainability activities, built on the value of the core deliverables.

Content of Deliverable

The following elements comprise this deliverable:

1. ROI Evaluation Report at Two Years (D5.2). The ROI evaluation report will serve as the primary artifact associated with this deliverable.
2. eTRIKS interim review consolidated report.
3. ROI Evaluation Report at Five Years (this document)

List of Abbreviations

AZ:	Astra Zeneca
CC:	Computing Center of the Institute of Nuclear and Particle Physics, CNRS
CDA:	Confidential Disclosure Agreement
CDISC:	Clinical Data Interchange Standards Consortium
CNRS:	Centre National de la Recherche Scientifique
DOW:	eTRIKS Description of Work
EHS:	eTRIKS Harmonization Service
EISBM:	European Institute for Systems Biology and Medicine, CNRS
eTRIKS:	European Translational Research and Information Knowledge management Services
ExCom:	eTRIKS Executive Committee
GA:	General Assembly
IMI:	Innovative Medicines Initiative
ITTM:	Information Technology for Translational Medicine
KM:	Knowledge Management
MDR:	Metadata Repository
MoU:	Memorandum of Understanding
MTA:	Materials Transfer Agreement
tranSMART:	Open source/license translational data warehouse
WP:	Work Package