**European Translational Information and Knowledge Management Services**

# eTRIKS Deliverable report

## Grant agreement no. 115446

## Deliverable D1.5

SOPs of quality control processes and services

Due date of deliverable: Month 36

Actual submission date: Month 40

| 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: | D1.5 |
| Deliverable title: | SOPs of quality control processes and services |
| Deliverable version: | 1 |
| Due date of deliverable: | Month 36 |
| Actual submission date: | Month 40 |
| Leader: | Ghita Rahal |
| Editors: | |
| Authors: | Leslie-Alexandre DENIS, Jaroslav Martasek, Ghita Rahal |
| Reviewers: | David Henderson, Chris Marshall and Fabien Richard |
| Participating beneficiaries: | BioSci Consulting |
| Work Package no.: | 1 |
| Work Package title: | Platform Service Delivery |
| Work Package leaders: | Ghita Rahal |
| Work Package participants: | Ghita Rahal, Derek Marren, Pengfei Lui, Gino Marhcetti, Benjamin Guillon, Andreas Tielmann, David Henderson, Denny Verbeeck, Trevor Garrett, George Kremenek, Florian Guitton, Wei Gu, Chris Marshall, Peter Rice, Jaroslav MArtasek, Leslie Alexandre Denis |
| Estimated person-months for deliverable: | 2 |
| Nature: | Report |
| Version: | 1 |
| Draft/Final: | Final |
| No of pages (including cover): | 12 |
| Keywords: | Quality, Services, Procedures |
| | |

# I. Introduction

This document outlines the approach that CC-IN2P3 and WP1 are taking to ensure consistent delivery of the services comprising the eTRIKS platform. As the platform is constantly evolving, so are the processes and approaches to ensure the quality of the delivery. Hence this document, instead of going into specific implementation details, focuses on the high level description of the approach as a whole. The details can be obtained upon request from eTRIKS WP1 computing centre members.

As a part of ensuring the quality of deliverable, Quality of Service metrics have been defined. This document describes the basic ones, how they are collected and the evaluation processes.

# II. Executive Summary

The goal of the Standard Operating Procedures (SOPs) is to provide in a written form the documentation and the tools to allow the deployment, operation and maintenance of the eTRIKS platform. Likewise, to achieve the required level of Quality of Services (QoS) and performance, tools have been deployed, metrics have been set and the approach has been documented. This deliverable provides in Section III some definition on the terms that are used throughout this document. Section IV describes the management of the documentation, Section V the organization of the Configuration Management. Section VI focuses on the Change Management cycle and Section VII on the monitoring and metrics accumulated to trace the performances of the various components of the platform. Finally, Section VIII describes briefly the process followed when hosting a new project.

# III. Vocabulary

For the purposes of this document and documentation effort as whole, it is helpful to distinguish between the whole eTRIKS platform and a specific project (e.g. ABIRISK, OncoTrack):

**eTRIKS Platform** – common set of infrastructure, services and processes that are provided to projects e.g. database server, OpenStack, networking, authentication and authorisation service, quality management, support etc.

**Hosted Project** – a specific implementation of software build on top of the platform connected by a specific research goal. E.g. database instance, curation VMs

The following document categories are used throughout the platform:

**Installation and verification instructions (IVI):** A very specific set of instructions on how to install / set-up / configure a service. This is in a form of files that will be executed by a computer or a human. It contains verification instructions that confirm the success of the installation. IVIs ensure consistency in services configuration / behaviour (e.g. data curation VM set up).

**SOP (standard operating procedure):** Set of step-by-step instructions to achieve a predictable, standardized, desired result often within the context of a longer overall process. Very detailed, it does not leave much room for interpretation and ensures consistency of a business process (e.g. error reports handling, how IVIs are executed and what records are kept).

**Guidance:** a high-level description of a process. Allows for interpretation and right-sizing according to the context in which it is applied.

**Architecture / implementation description:** description that helps understanding the dependencies of the components of a system. It might refer more detailed descriptions.

**Record:** something that represents proof of existence and that can be used to recreate or prove state of existence. E.g. log entry showing user was granted / revoked access, person who approved of a document for publication.

# IV. Document and Project management

A consistent approach to **document management** has been established that allows for various types of documents to be reliably stored. A complete history of changes is recorded and metadata about the documents are captured.

**A Project Management tool** is used as the authoritative source of the documents unless specified otherwise (e.g. authentication logs and similar records are stored in their corresponding data stores). The different document categories, each one serving different purpose, are all recorded in this management tool but may refer also to other more specific management tools such as version control systems or technical software platform.

## a. Management tools

The Computing Centre uses a set of tools to manage and follow up the various undertaken activities. Some of them were chosen to be used within eTRIKS:

A flexible **project management**[1] **web application** is used to store documentation, draw Gantt charts, and provide issue tracking system and many other features that are shared by the platform contributors.

**Incident and change management tool**[2]: We use specific tracking tool to communicate and exchange tickets with all the members of the eTRIKS project. Again, this tool is the one deployed for the whole data centre and a specific queue was customized for eTRIKS use.

**Collaborative software** such as wikis, git[3], gitlab, are used to record the layers of the platform, allow versioning of the different changes and recovery in case of problems and issues.

Beside the tools themselves, regular meetings, presentations, trainings are organized to collect requests for change and share information on the evolution and the adequacy of the platform to the needs.

[1] www.redmine.org
[2] www.otrs.org ➔ https://usersupport.etriks.org/
[3] git-scm.com

### b. Dissemination and reference documentation on processes

Certain documents might be used for publication outside of the WP1 team. For this purpose https://portal.etriks.org and teamwork portal are utilised to publish either to wide public or just to the eTRIKS community respectively.

## V. Configuration management

Processes that manage the configuration of the platform and the one deployed for the hosted projects have been established. The implementation focused on maximum automation and traceability of the processes.

The technical documentation for the components of the eTRIKS platform and of each of the hosted projects is stored in the project management tool, organized in categories. Figure 1 shows the index of the documents, which map up to these categories described below:

- Architecture and Inventory

Listing of all the resources available is kept up-to-date, including internet addresses, virtual machines, security groups on the platform with their characteristics (number, size…)

- Conventions

A naming convention has been developed which standardizes the naming of components like Web addresses, virtual machines' names, and hosts' names. For example, the name of a Virtual Machine will reflect the project type, the service hosted in the VM, the application, the status and an ID number: otk-dev-curation01 will be the first (01) Oncotrack (otk) development (dev) virtual machine for curation.

- Deployment and Operation SOPs

Deployment procedures are indexed and available inside the project management tool's wiki and describe precisely how the main eTRIKS services are setup. One of them is for example "Deployment and configuration of the LDAP server".

Operation procedures are, in the same way, available in the project management tool's wiki and describe in detail how to proceed with general operations like migration of services, backup of data etc…

The configuration of each of the components is stored in the form of scripts in the GIT version control system in. Each configuration related to projects or core services is versioned from the starting point to the production level in order to fully track their lifecycle. Access to the source control system is authenticated via certificate to prevent data breach and the system is backed up to prevent data loss.

Each module (core service, gateway, web server) has its specifics and to deal with them, we use configuration management software, Ansible Playbook[4] in our case. This allows adapting all these configurations to their specific purpose and at the same time to be able to consistently reproduce

[4] http://www.ansible.com

them as many times as necessary. Playbooks have been developed with the same workflow as configurations, that way we can initialize every part of the platform at any time, regardless of the location.

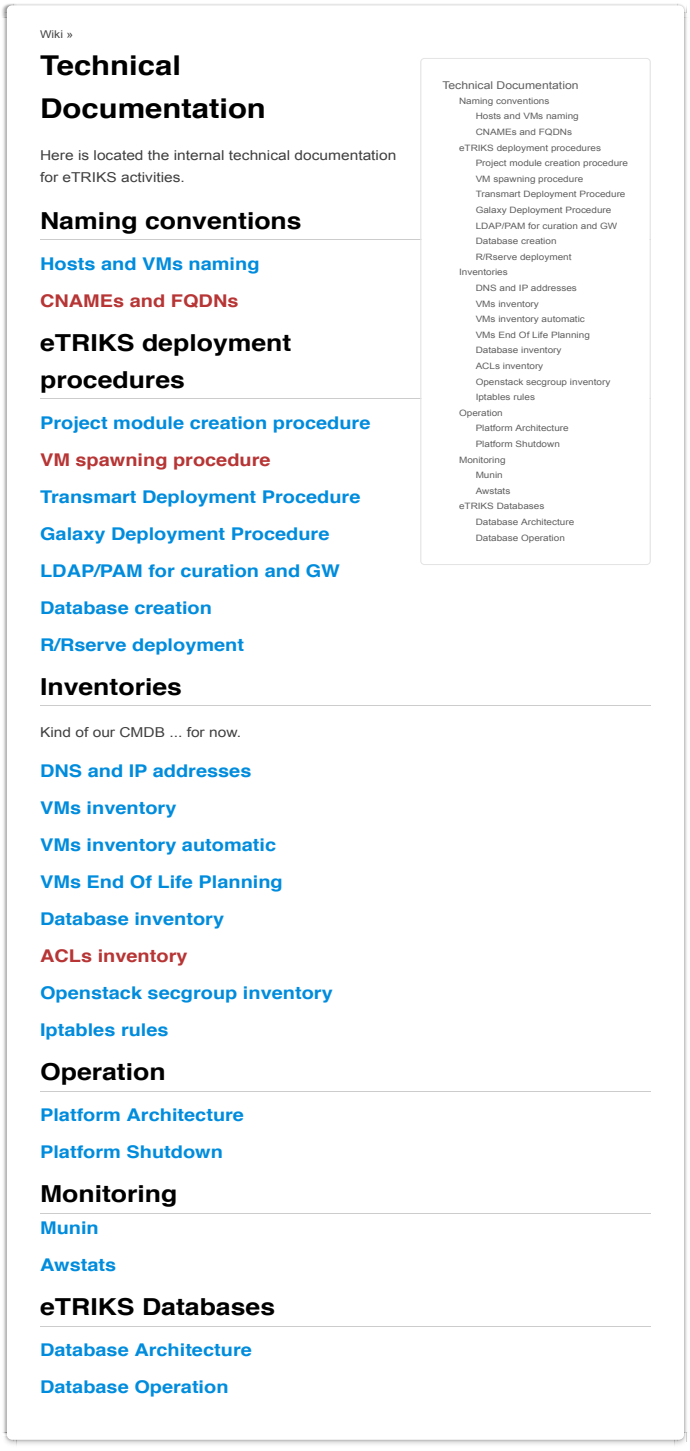


**Figure 1: Index of the wiki page describing configuration management and more.**

Ansible is used for any deployment of the services included in a virtual machine and communicates through a secure protocol (i.e. SSH) with the platform. The virtual machine is built through an

image recorded inside the Openstack repository. Execution of a configuration change is authenticated, traced and the change is recorded into the project management application.

The process is being built and once achieved, will allow for full audit of what configuration of software has been used at any point in time.

# VI. Incident and change management

## a. Under external trigger

All incidents and requests reported by users are tracked as tickets in a helpdesk software tool[5] that handles service management for the project. The helpdesk treats the tickets reported by the user almost in real time during business days and they are escalated according to the complexity and the specificity of the incident or request. A detailed description of the escalation mechanism and the service level agreement can be found in pages 11/12 of the eTRIKS deliverable D1.7[6] and in the diagram of figure 2.

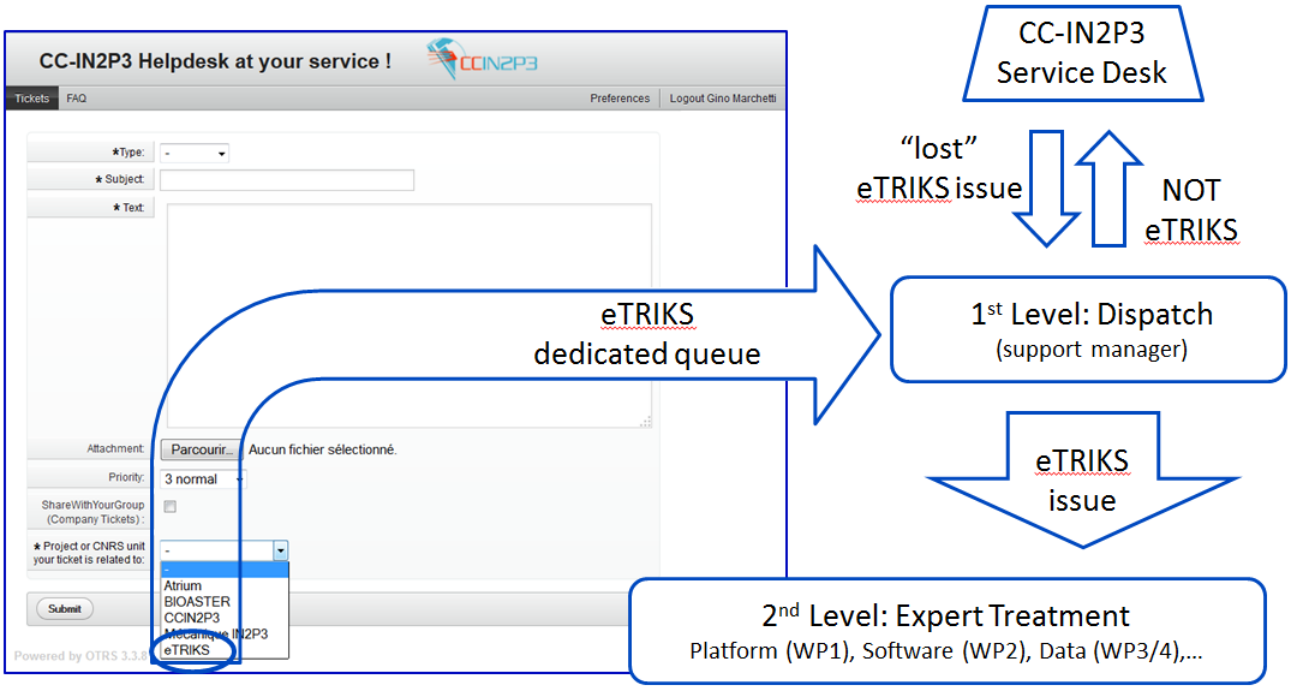


**Figure 2**: Request-dispatching from the general CC-IN2P3 helpdesk system to the specific eTRIKS internal ticketing treatment queue.

Weekly reports on the status of the tickets are made at the "Operation Update TeleConference" meeting to ensure their follow-up. The Operation Group gathers a team of people from most of the Work Packages.

[5] https://usersupport.etriks.org/
[6] https://ws.onehub.com/workspaces/284851/files/245214925

Should a ticket require a change to be implemented, the decision would be taken either by the concerned expert or, if more complex, the members of Operations Group would propose a line of treatment of the incident, validated by the concerned Work Package.

Training and documents[7] are provided to both the agents treating the tickets and the newcomers in order to get them informed on the processes put in place in eTRIKS for incident and change management.

Monthly TCs are organized between the Work Package 1 members and the users of the projects that have joined eTRIKS. On these meetings issues or general requests that require technical exchanges are discussed between the hosted project representatives and eTRIKS..

All the documents and the specific meetings described above compose the SOP on how the incident and change management is handled.

Minutes of all meetings are held in the eTRIKS TeamWorks area and are accessible to all eTRIKS project members.

### b. Under internal trigger

Overall monitoring of the platform is performed by various probes that are able to send notifications to the concerned experts. The notification is escalated following a procedure, from a standard recipe that can be followed up by the first person notified, to the escalation to the experts on the relevant subsystems.

Tools like Nagios[8], widely used in data centres, are used for infrastructure monitoring. Any alert or notification triggers an email message with its level of criticality to a pool of people.

Munin[9] is the resource monitoring open source tool that analyses the resource usage trends and displays the monitored performance through a web interface. All the Virtual Machines that host eTRIKS resources are monitored this way. System administrators can track any degradation of performance and control the effect of the adjustments.

Log files of all the events are kept on the machines that host the application or process source. Overall changes in the state of the platform are recorded into project management web applications like Redmine[10] that is used as a configuration management tool (see Configuration Management Section V).

As much as possible, our choice of tools favours those that are standard and open-source so that they can be deployed even in smaller data centres without incurring any additional costs.

[7] Guides and presentation:
https://ws.onehub.com/workspaces/284851/folders/185622596
[8] https://www.nagios.org/
[9] http://munin-monitoring.org
[10] www.redmine.org

## VII. QoS metrics: definition and monitoring

Both the platform and hosted projects are constantly monitored and a set of metrics are defined to be continuously collected. In the case of a breach of a limit for a critical metric or metrics, an incident process (see Section VI. Incident and change management) is triggered. At the same time, selected metrics are published online on the project dashboard available to the project owner of the respective project.

There are two main levels of monitoring in place: IT systems level and application level.

On the IT systems level it is the infrastructure, hardware and operating system that are monitored. Any degradation in quality on this infrastructure level will affect the metrics in the project level above.

The following areas are monitored: virtual machines status including operating system status (is the machine up and responsive, what is the CPU / memory load), networking status (can all devices be reached and in what time) and storage status (what is the load on storage subsystem, capacity status and data growth projections).


**Figure 3 List of the metrics monitored for a given Virtual Machine**

The application level covers more specific functions of the systems – e.g. number of requests, their success, service availability, response time, etc. These application level metrics are collected for both the eTRIKS platform services (e.g. database, authorisation service) and for each hosted project services (e.g. tranSMART requests).

For each server, graphs are provided to monitor the system and the application. One can see in figures 3 and 4 below the set of parameters for which evolution over time is provided and an illustration with the memory usage of a VM.

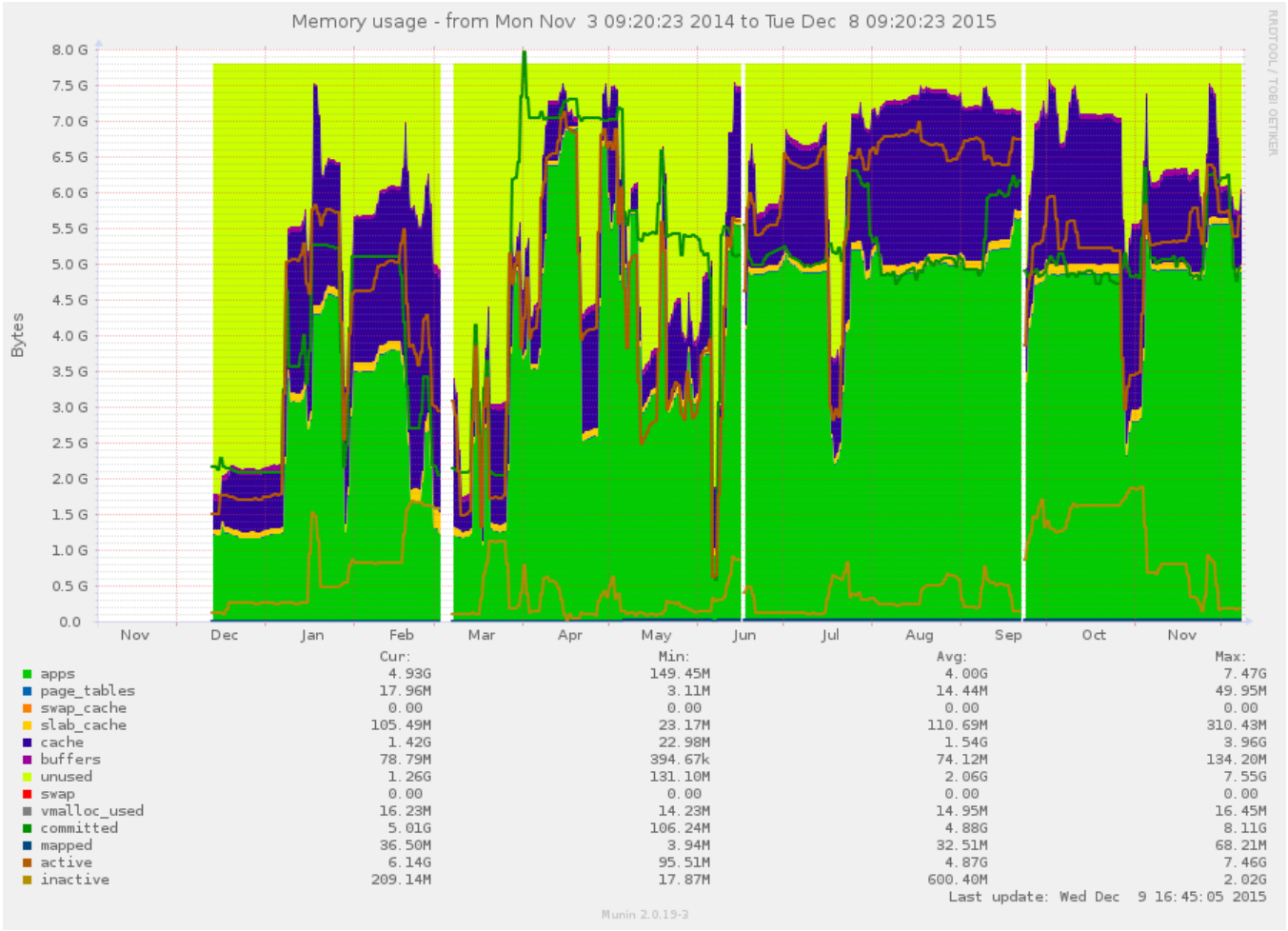


**Figure 4: Graph of Memory Usage for a given Virtual Machine**

The quality of service provided by the platform to the projects is evaluated according to these basic metrics:

- Availability – the percentage of a time period when the system was available.
- Reliability – the percentage of the requests that were completed successfully. Response time – average response time for accessing tranSMART web site application.

The value may vary from project to project based on the data size and query complexity.

These metrics are measured on case-by-case basis from the metric stored above. They can be reported on a more automatic basis if this is requested by the project.

As security is crucial element in dealing with patient data, there are metrics collected that allow for detection and prevention of security breaches. The security model was described in Deliverable 1.1 (eTRIKS KM service based on eTRIKS platform) and includes collecting metrics on unsuccessful authentication and authorisation attempts.

## VIII. Project Onboarding

Training, guidance and SOPs ensure that all necessary steps are followed when a new project is being on boarded. Face-to-face and teleconference meetings are organized to evaluate the requirements such as the capacity needs, the levels of security and access and the roles for each of the members of the project. Specific trainings are organized for the project owner and other members to ensure that each one is aware of their role and duty and to train each one on the tools appropriate to their role. All the information about the members: the group they belong to, the roles and the privileges are stored on the LDAP (Lightweight Directory Access Protocol) server and the project owner has access and validation privileges to it.

For each hosted project, a contractual documents such as an MTA (Material Transfer Agreement) or a DPA (Data Processing Agreement) describe the processes and are signed by both parties.