EGI-Engage

First release of the new Accounting Portal deployed in production

D3.5

Date	25 May 2016
Activity	WP3
Lead Partner	CSIC / CESGA
Document Status	FINAL
Document Link	https://documents.egi.eu/document/2790

Abstract

This deliverable describes the first release of the EGI Accounting Portal during EGI-Engage including the developments made during the first year of the EGI-Engage project. The EGI Accounting Portal receives data from APEL and ultimately from sites participating in the EGI and WLCG infrastructures as well as from sites belonging to other Grid organisations that are collaborating with EGI. This is crossed with metadata from other sources to offer an integrated view of accounting data on the EGI Infrastructure. The production instance will be set up on the first half of May.

The EGI-Engage project is co-funded by the European Union (EU) Horizon 2020 program under Grant number 654142 http://go.egi.eu/eng

COPYRIGHT NOTICE

DELIVERY SLIP

	Name	***Partner/Activity***	***Date***
From:	Iván Díaz Álvarez	CSIC / WP3	28/03/2016
Moderated by:	Yannick Legrè	EGI.eu/WP1	24/05/2016
Reviewed by	Gergely Sipos Kostas Koumantaros	EGI.eu/WP6 GRNET/PMB	24/05/2016 24/05/2016
Approved by:	AMB and PMB		25/05/2016

DOCUMENT LOG

Issue	***Date***	***Comment***	***Author/Partner***
v1	03/28/2016	Document creation	I. Díaz / CSIC
v2	04/28/2016	Modifications after EGI-Engage External review	I. Díaz / CSIC
v3	24/05/2016	Final version implementing last reviewers' comments	I. Díaz / CSIC

TERMINOLOGY

A complete project glossary is provided at the following page: http://www.egi.eu/about/glossary/

EGI-Engage - The EGI-Engage project (Engaging the Research Community towards an Open Science Commons) started in March 2016, co-funded by the European Commission for 30 months, as a collaborative effort involving more than 70 institutions in over 30 countries. EGI-Engage aims to accelerate the implementation of the Open Science Commons by expanding the capabilities of a European backbone of federated services for compute, storage, data, communication, knowledge and expertise, complementing community-specific capabilities.

APEL - An accounting tool that collects accounting data from sites participating in the EGI and WLCG infrastructures as well as from sites belonging to other Grid organisations that are collaborating with EGI, including OSG, NorduGrid and INFN.

WLCG - The Worldwide LHC Computing Grid (WLCG) project is a global collaboration of more than 170 computing centres in 42 countries, linking up national and international grid infrastructures. The mission of the WLCG project is to provide global computing resources to store, distribute and

analyse the ~30 Petabytes (30 million Gigabytes) of data annually generated by the Large Hadron Collider (LHC) at CERN on the Franco-Swiss border.

UI - User Interface.

CESGA - Fundación Pública Galega Centro Tecnolóxico de Supercomputación de Galicia (CESGA) is the centre of computing, high performance communications systems, and advanced services of the Galician Scientific Community, the University academic system, and the National Scientific Research Council (CSIC).

CSIC - The Spanish National Research Council (CSIC) is the largest public institution dedicated to research in Spain and the third largest in Europe.

AJAX - (Asynchronous JavaScript and XML) is a set of web development techniques using many web technologies on the client-side to create asynchronous Web applications.

URL - A Uniform Resource Locator (URL), commonly informally termed a web address (which term is not defined identically)[1] is a reference to a web resource that specifies its location on a computer network and a mechanism for retrieving it

JSON - (JavaScript Object Notation) is an open-standard format that uses human-readable text to transmit data objects consisting of attribute–value pairs.

SSM - Secure Stomp Messenger (SSM) is the messaging system used by APEL to transmit messages. It is written in Python and uses the STOMP protocol.

VO - Virtual organisations (VOs) are groups of researchers with similar scientific interests and requirements, who are able to work collaboratively with other members and/or share resources (e.g. data, software, expertise, CPU, storage space), regardless of geographical location.

NGI - National Grid Initiatives, NGIs) are organisations set up by individual countries to manage the computing resources they provide to EGI.

RESTful - A RESTful API is an application program interface (API) that uses HTTP requests to GET, PUT, POST and DELETE data.

DN - (Distinguished Name) A composite string that uniquely identifies an entity, in this case an individual user.

DB - Database

GOCDB - The official repository for storing and presenting EGI topology and resources information.

XML - Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. It is defined by the W3C's XML 1.0 Specification and by several other related specifications, all of which are free open standards.

API - (Application Programming Interface) is a set of routines, protocols, and tools for building software and applications.

TierX - The Worldwide LHC Computing Grid (WLCG) is composed of four levels, or "Tiers", called 0, 1, 2 and 3. Each Tier is made up of several computer centres and provides a specific set of services. Between them the tiers process, store and analyse all the data from the Large Hadron Collider (LHC).

OSG - (Open Science Group) - OSG provides common service and support for resource providers and scientific institutions using a distributed fabric of high throughput computational services. The OSG does not own resources but provides software and services to users and resource providers alike to enable the opportunistic usage and sharing of resources. The OSG is jointly funded by the Department of Energy and the National Science Foundation of the United States.

VOMS - (Virtual Organization Membership Service) is a system for managing authorization data within multi-institutional collaborations. VOMS provides a database of user roles and capabilities and a set of tools for accessing and manipulating the database and using the database contents to generate Grid credentials for users when needed.

MoU - A memorandum of understanding (MoU) describes a bilateral or multilateral agreement between two or more parties. It expresses a convergence of will between the parties, indicating an intended common line of action.

REBUS - (REsource, Balance and USage) is a WLCG service that centralizes the topology information, resource pledges, and installed capacities.

SpecInt - A computer benchmark specification for CPU integer processing power

CSV - (Comma-Separated Values) A file format that stores tabular data (numbers and text) in plain text. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. The use of the comma as a field separator is the source of its name.

GPGPU - (General-Purpose computing on Graphics Processing Units) is the use of a graphics processing unit (GPU), which typically handles computation only for computer graphics, to perform computation in applications traditionally handled by the central processing unit (CPU).

Contents

Executive summary

This deliverable describes the first EGI-Engage release of the EGI Accounting Portal including the developments made during the first year of the project. The EGI Accounting Portal receives data from APEL and ultimately from sites participating in the EGI and WLCG infrastructures as well as from sites belonging to other Grid organisations that are collaborating with EGI. This is crossed with metadata from other sources to offer an integrated view of accounting data on the EGI Infrastructure.

The developments covered here are rewritten UI and look & feel, AJAX functionality, new graphing technologies, new URL structure, JSON capabilities and a new restricted view substituting older privileged views.

The dependencies are in two groups, the main data gathered from APEL using the Secure Stomp Messenger (SSM) and Metadata gathered from several sources that is used to categorize and make sense of the accounting data stored on the main Database crossing it with metadata. This release is the first with a new implementation based on the Django Python framework, and the Dojo and Bootstrap Javascript frameworks that both update the interface to support the latest web technologies and exposes new functionality and flexibility to the end user. The next features to be implemented are the rest of the accounting reports, the Storage view, new geographical graphing capabilities and the ability to compare data from two separate periods.

1 Introduction

The following table provides a summary of the tool covered in this release.

Tool name	*Accounting Portal*
Tool url	*https://accounting-devel-next.egi.cesga.es/*
Tool wiki page	*https://wiki.egi.eu/wiki/Accounting_Portal*
Description	EGI Core Service – The Accounting Portal provides data accounting views for users, VO Managers, NGI operations and the general public.
Value proposition	This new release of the Accounting Portal uses a new architecture and user interface that updates it to use new Web technologies and exposes more flexibility to the end user.
Customer of the tool	EGI, WLCG, Site and VO Admins, Infrastructure Users, others.
User of the service	Any stakeholder interested on the accounting of jobs in the EGI infrastructure.
User Documentation	https://documents.egi.eu/document/2789
Technical Documentation	https://documents.egi.eu/document/2545
Product team	*CESGA / CSIC*
License	*Apache License, Version 2.0*
Source code	*https://github.com/cesga-egi/accounting*

2 Service architecture

The service architecture provides an overview of the key (logical) service components and their dependencies to help better understand the structure and logical as well as technical setup of the main data gathered from APEL using SSM and Metadata gathered from several sources that is used to categorize and make sense of the accounting data

High-Level Service architecture

The Accounting Portal is a web application which has as its primary function to provide users, like VO managers, Site Admins, non-privileged users and other stakeholders, with customized accounting reports, containing tables and graphs, as web pages. It also offers RESTful web services to allow external entities to gather accounting data.

This chapter details the basic architecture of the Portal, which consists on:

1. A backend, which aggregates both data and metadata in a MySQL database, using the APEL SSM messaging system[1] to interact with the Accounting Repository and several scripts, which periodically gather the data, and metadata described below.
2. A Model represented by database schemas both external and internal which define database tables for several types of accounting (grid, cloud, storage, multicore, user statistics etc.) and metadata (topology, geographical data, site status, nodes, VO users and admins, site admins etc.) and a series of parametrized queries,
3. A set of views that expose the data to the user. These views contain a form to set the parameters and metric of the report, a number of tables showing the data parametrized by two selectable dimensions and filtered by several parameters, a line graph showing the table data, and pie charts showing the percentage distribution on each dimension. It is planned that this part of the portal will evolve with interactive graphs, responsive in real time, reactive and only exposing advanced controls on user demand.

A graphical representation of these components is depicted on Fig. 1.

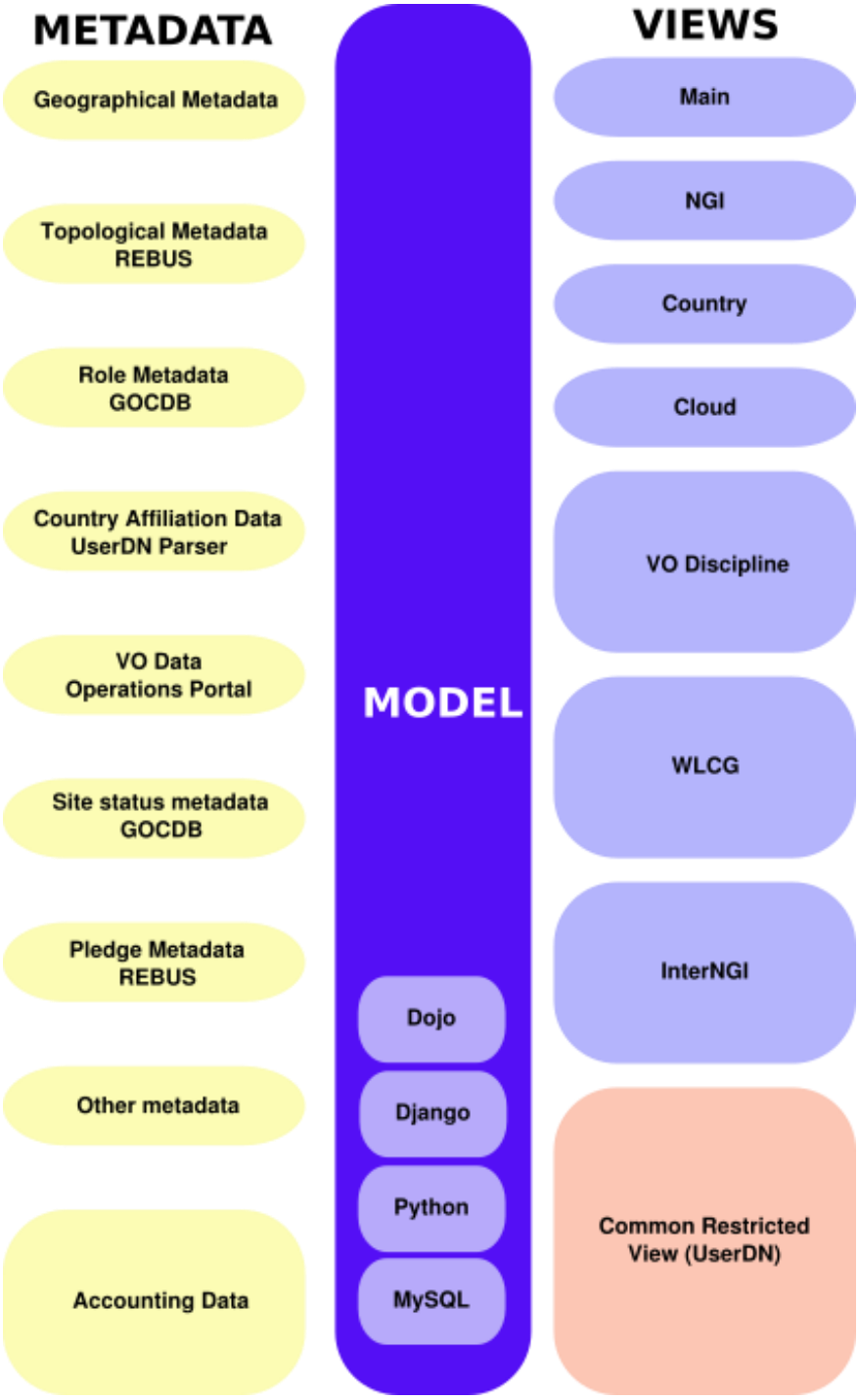

Figure 1 – Accounting Portal Architecture

APEL SSM

The Accounting Portal has to refresh its database periodically with data from the Accounting Repository to assure that information published is up-to-date. Secure Stomp Messenger (SSM), a queue messaging system based on ActiveMQ, is used for synchronization purpose and also for the communication between sites and the Accounting Repository. The SSM system is composed by:

1. A SSM loader for each accounting source (multicore, cloud, storage, etc.). This daemon waits for messages arriving on a queue and authenticates them with a DN and certificate. If a message is deemed valid, it is saved to a spool directory for further processing.
2. A DB loader, this daemon monitors the spool directory and, if there are messages, these are imported in the DB in order. This import at present does not delete the previous data in the tables; it only overwrites it. Manual intervention is needed for stale data.

The accounting data is sent several times per day by APEL in chunks of 1000 registers. This obviates the need for the portal to do pull requests.

Metadata Gathering

Metadata is a category of data that complements the raw accounting data and allows the portal to organize, categorize and impart new meaning to it. This metadata includes:

1. **Geographical Metadata**: Country and NGI affiliation of sites. Generally, this follows current borders, but there are important exceptions. This is gathered from GOCDB using its XML-based API.
2. **Topological Metadata**: Sites are presented in trees, there are Country and NGI trees that correspond to geographical classifications, but there are also trees based on topological classifications like Tier1 and Tier2 sites, OSG sites and uncategorised sites. Inside Tier2 sites, the federation they belong to is also important and can trigger special code in some cases. Gathered from several sources, including OSG and WLCG databases.
3. **Role Metadata**: VO members and managers, and the site admins records. This metadata controls the access to restricted views. Information is gathered from GOCDB and individual VOMS servers constructing a list of individual VOMSes and querying them with the VOMS API.
4. **Country affiliation data**: Each user record contains a user identifier that has his/her user name, institution and sometimes country. Scripts in the backend map each user with a country based on the institution which issues their certificate. This data is used in anonymized statistics per country on: how much resources from other countries are used by given country and the distribution of its resources used by other countries.
5. **VO Data**: To make possible VO selection in the user interface, the portal stores lists of VOs. They are also used to filter incorrect VO names, provide access to VO managers, and arrange

accounting by VO discipline (such as “High Energy Physics”, “Biomedicine”, “Earth Sciences”, etc.). Information is gathered from the Operations portal using its XML based APIs.

6. **Site status metadata**: Sites must be filtered to exclude those that are not in production (due to being closed or being in test mode). There must be also metadata to aggregate the accounting history of sites whose name has been changed. There are requirements to extend this functionality to NGIs. Information is gathered from GOCDB using its XML tables and internal tables compiled as part of EGI PROC 15[2].
7. **Pledge metadata**: The WLCG reports have to contain only those sites where MoUs or other pledges between VOs and sites are honoured, so the validity date and pledged hours are needed. Information is gathered from WLCG using the REBUS service.
8. **Other metadata:** There are also other metadata like local privileges, SpecInt calculations, publication status, VO activities and more. Some of these metadata is calculated internally using other types of metadata and published for other EGI operational tools, like VO activity data and Site UserDN publishing

3 Release notes

Requirements covered in the release

This release is the first with a new implementation based on the Django Python framework, and the Dojo and Bootstrap JavaScript frameworks that both update the interface to support the latest web technologies and exposes new functionality and flexibility to the end user.

- **Accounting Portal navigation tree** - https://rt.egi.eu/rt/Ticket/Display.html?id=10156 – This is included as the new VO Discipline navigation tree that allows the user to see the three-level discipline hierarchy, view all sub disciplines and VOs for any discipline in any level and jump seamlessly between them.
- **Report generator in Portal -** https://rt.egi.eu/rt/Ticket/Display.html?id=9733. This will be covered in the next revision with an special report type.
- **Front page** - https://rt.egi.eu/rt/Ticket/Display.html?id=8823 - Implemented, it uses natural 1, 3, 6, 12 months, since shorter intervals expose artifacts of the summarization process.
- **Export data function** - https://rt.egi.eu/rt/Ticket/Display.html?id=10159 - This is supported as JSON for the time being. XML and CSV will be included later.
- **Unofficial VO list in the Portal** - https://rt.egi.eu/rt/Ticket/Display.html?id=9704 - There is a curated list of EGI VOs - disciplines now. There is still need of updating this list from official sources.
- **EGI Scientific Discipline Classification in the accounting portal**- https://rt.egi.eu/rt/Ticket/Display.html?id=9583 - As with #10156 above.
- **Simplify access to some basic functionality. Avoid the use of complex forms for common statistics and get accounting information for some common queries** - https://rt.egi.eu/rt/Ticket/Display.html?id=9075 – Implemented with an accordion hiding advanced form options by default.
- **RESTFUL interface to XML output (or JSON, csv) -** https://rt.egi.eu/rt/Ticket/Display.html?id=9052 - We currently use JSON.
- **Accept Storage Records by messaging**. https://rt.egi.eu/rt/Ticket/Display.html?id=9051 Done.
- **Improve graphs visualization (New)** https://rt.egi.eu/rt/Ticket/Display.html?id=8827 Done.
- **Portal to accept all data by messaging** - https://rt.egi.eu/rt/Ticket/Display.html?id=9048 We currently only use SSM, a messaging service.
- **Restricted Views are Combined-** Restricted views were split into the VO Manager and Site Manager views. - https://rt.egi.eu/rt/Ticket/Display.html?id=11076
- **Resize of the page** – The page and its elements correctly resize on Chrome, Firefox and Internet Explorer. https://rt.egi.eu/rt/Ticket/Display.html?id=11042
- **Increase the size of the table showing the accounting data –** The table shows at least 50 elements, configurable to 100 or 200 https://rt.egi.eu/rt/Ticket/Display.html?id=11040

- **Pagination –** New pagination in the table, it shows the results in pages if the list is too long, including controls to move between pages and control the number of elements shown by page. The controls are correctly styles https://rt.egi.eu/rt/Ticket/Display.html?id=11039
- **Breadcrumb –** The position in the portal is shown in a breadcrumb component in the top right. https://rt.egi.eu/rt/Ticket/Display.html?id=11038
- **Graph area in the accounting portal –** The zooming and panning functionality of the graph only works after clicking it, avoiding user inconvenience. https://rt.egi.eu/rt/Ticket/Display.html?id=11031
- **Remember viewing options –** The portal keeps the state of the options when using a link that is not in the menu area. https://rt.egi.eu/rt/Ticket/Display.html?id=10982
- **Display SubmitHost –** SubmitHost is now a valid Row or Column variable. https://rt.egi.eu/rt/Ticket/Display.html?id=10978
- **Fewer significant digits in portal views –** Numbers with decimals are rounded in the views. https://rt.egi.eu/rt/Ticket/Display.html?id=10977
- **Permanent URL for Portal view –** A permanent URL for the current view and options is displayed. https://rt.egi.eu/rt/Ticket/Display.html?id=10976
- **Units of Measure –** The portal shows the units for those metrics in which they are relevant in the titles and axis of graphs. https://rt.egi.eu/rt/Ticket/Display.html?id=9662

4 Feedback on satisfaction

NGIs, EGI.eu operations, WLCG and appointed reviewers were involved in testing the new Accounting Portal. They provided several comments and requirements after the evaluation of a testing instance of the portal (see the Accounting Portal requirement queue[1]).

All these requirements have been satisfied in the final release of the portal or scheduled in a next version that will be released in June 2016. A further evaluation was done on this new release and the final feedback was positive.

[1] https://rt.egi.eu/rt/Dashboards/5538/AccPortal-Requirements

5 Future plans

The next features to be implemented:

- **More Accounting reports –** Specialised report pages will be done, in some cases utilizing the geographical capabilities below.
- **Storage view –** A view will be added for the storage accounting records, with appropriate metrics and automatic unit scaling.
- **New geographical graphing capabilities –** The graphing engine based in Dojo allows to represent numerical data using a geographical vectorial map. Maps will be provided for European countries and the World that will allow a new way to visualize the accounting data.
- **Comparing data from two separate periods –** The normal accounting view already has a start and end date selector that defaults to the last year. This will include a second time period that will serve as comparison base to compare any metric and grouping and derive trends.
- **GPGPU Accounting –** Accelerated computing records will be supported shortly after they are made available from APEL.