EGI-InSPIRE

GRID POLICY ON THE HANDLING OF USER-LEVEL JOB ACCOUNTING

Document identifier	EGI-SPG-Accounting-V2_0
Document Link	https://documents.egi.eu/document/85
Last Modified	**19/03/2013**
Version	**2.0**
Policy Group Acronym	**SPG**
Policy Group Name	**Security Policy Group**
Contact Person	**David Kelsey/STFC, UK**
Document Status	**Approved**
Approved by	**EGI.eu Executive Board**
Approved Date	**10/09/2010**

Policy Statement

This document presents the minimum requirements and policy framework for the handling of user-level accounting data created, stored, transmitted, processed and analysed as a result of the execution of jobs on the Grid.

I. COPYRIGHT NOTICE

II. DELIVERY SLIP

	Body	Date
Approved by	EGI.eu Executive Board	10/09/2010

III. DOCUMENT LOG

Issue	Date	Comment	Author/Organization
1.0	14/07/2010	Imported from JSPG policy document with the same title. No changes to wording were made. See https://edms.cern.ch/document/855382 (V1.0, dated 6 Aug 2009) for the old JSPG document.	David Kelsey/STFC
2.0	19/03/2013	Just one minor change to increase the period of retention of data (section 7) containing personal identifying information from one year to 18 months. This change was approved by both the OMB and the UCB.	David Kelsey/STFC
3.0			
4.0			

IV. PROJECT SUMMARY

To support science and innovation, a lasting operational model for e-Science is needed − both for coordinating the infrastructure and for delivering integrated services that cross national borders.

The EGI-InSPIRE project will support the transition from a project-based system to a sustainable pan-European e-Infrastructure, by supporting 'grids' of high-performance computing (HPC) and high-throughput computing (HTC) resources. EGI-InSPIRE will also be ideally placed to integrate new Distributed Computing Infrastructures (DCIs) such as clouds, supercomputing networks and desktop grids, to benefit the user communities within the European Research Area.

EGI-InSPIRE will collect user requirements and provide support for the current and potential new user communities, for example the ESFRI projects. Support will also be given to the current heavy users of the infrastructure, such as high energy physics, computational chemistry and life sciences, as they move their critical services and tools from a centralised support model to one driven by their own individual communities.

The objectives of the project are:

1. The continued operation and expansion of today's production infrastructure by transitioning to a governance model and operational infrastructure that can be increasingly sustained outside of specific project funding.
2. The continued support of researchers within Europe and their international collaborators that are using the current production infrastructure.
3. The support for current heavy users of the infrastructure in earth science, astronomy and astrophysics, fusion, computational chemistry and materials science technology, life sciences and high energy physics as they move to sustainable support models for their own communities.
4. Interfaces that expand access to new user communities including new potential heavy users of the infrastructure from the ESFRI projects.
5. Mechanisms to integrate existing infrastructure providers in Europe and around the world into the production infrastructure, so as to provide transparent access to all authorised users.
6. Establish processes and procedures to allow the integration of new DCI technologies (e.g. clouds, volunteer desktop grids) and heterogeneous resources (e.g. HTC and HPC) into a seamless production infrastructure as they mature and demonstrate value to the EGI community.

The EGI community is a federation of independent national and community resource providers, whose resources support specific research communities and international collaborators both within Europe and worldwide. EGI.eu, coordinator of EGI-InSPIRE, brings together partner institutions established within the community to provide a set of essential human and technical services that enable secure integrated access to distributed resources on behalf of the community.

The production infrastructure supports Virtual Research Communities − structured international user communities − that are grouped into specific research domains. VRCs are formally represented within EGI at both a technical and strategic level.

TABLE OF CONTENTS

1 GRID POLICY ON THE HANDLING OF USER-LEVEL JOB ACCOUNTING DATA: INTRODUCTION

This document presents the minimum requirements and policy framework for the handling of user-level accounting data created, stored, transmitted, processed and analysed as a result of the execution of jobs on the Grid.

Each job executed on a Grid resource produces an accounting record. The schema for this accounting record is based on GFD.98 (*http://www.ogf.org/documents/GFD.98.pdf*) as defined by the Usage Record Working Group (UR-WG) in OGF. This schema includes the X.509 certificate Subject Distinguished Name of the submitting user and this information is therefore classified as personal information and has to be properly handled to meet the legal requirements related to data protection.

2 SCOPE

This document addresses the handling of accounting data resulting from the execution of jobs on the Grid. It does not cover any other forms of accounting or monitoring data.

The document is aimed at EU-based Grids and more specifically at:

- Site Managers to allow them to share user-level job accounting with the Grid, for the purposes described below.
- VO Resource Managers and the Grid Operations personnel who have access to the user-level accounting from more than one site.

3 THE PURPOSE AND REASONS FOR THE COLLECTION AND STORAGE OF THE INFORMATION

Accounting data is required:

- For the VOs to find out how much of their resource allocation has been used in total and by which group or role within the VO. This allows the VO to monitor, plan and control the use of their resource allocation.
- For the sites to find out how the resources they provide to the Grid are being used and by whom. This allows them to (re-)assign their resources properly and plan purchases in a timely fashion.
- By the Grid management and/or VOs to find out if any pledged resources have indeed been provided and properly used by the VOs. This allows for better monitoring, control and planning.
- For VOs, Grid Management and Sites to report on usage to their funding bodies.
- For operational and scientific analysis only anonymised and aggregated accounting data will be used.

User-level accounting is required:

- For the VOs to understand and control how many and which individuals within the VO, group or role are using resources.
- For VOs, Grid Management and Sites to report on anonymised and aggregated statistics to their funding bodies.
- For Grid Operations during operational troubleshooting and debugging.
- For Grid Security Operations in forensic analysis of security incidents.

All other uses of the accounting data are forbidden.

4 ACCOUNTING DATA STORAGE

Each site collects and stores an accounting record for each job executed at their site. These records are stored locally at the site according to national data privacy laws.

Each site is responsible for sending its accounting records on a regular basis, e.g. daily, with at least user DNs encrypted in transport, to a central data base defined by the Grid. This database is located at an Accounting Data Centre (ADC). The location of the ADC needs to be chosen carefully according to data privacy laws.

The ADC securely stores all the individual job records from each of the sites submitting such records.

The ADC generates and securely stores aggregated statistics from the data. Levels of aggregation are defined by the Grid, e.g. per Site, per VO, per Month, per User, etc.

There may be more than one ADC in the Grid, e.g. one per country. Accounting records and aggregated data may be transferred between ADCs within a Grid or between ADCs belonging to different Grids, providing both Grids have adopted this policy. User DNs must always be encrypted in transport between ADCs. Whenever this policy refers to "the ADC" this should be interpreted to refer to all ADCs under the control of the Grid.

The specification of which accounting data needs to be transferred to which ADC, and the various access control requirements, is subject to agreement between the Grid and the VO.

5 INFORMING THE USER

The user is informed about the collection and handling of accounting data during their first registration or subsequent renewal with their VO. During registration users must accept the conditions of the Grid Acceptable Use Policy (AUP). This AUP has a clause on the use of logged information.

6 CONTROL OF AND ACCESS RIGHTS TO THE INFORMATION

The local accounting record for a job is controlled by the site at which the job is executed. The submitting user's DN may be unencrypted in this information and access is restricted to the local resource administrators or other authorised persons.

Copies of individual job accounting records and aggregated data in the ADC central database are controlled by the Grid.

ADC staff, according to their role or job responsibilities, may be authorised to have access to the individual job records. All other persons have no access to the database.

Access to aggregated data at the VO level may be public information if the VO agrees but otherwise appropriate access control will be required.

Access to VO group/role aggregated data, if required by the VO, is restricted to members of that VO.

The aggregated data of a user must be properly protected. All user data in the database is anonymous in the sense that the user data cannot be connected to a user name. Access to this anonymised data, if requested by the VO, must be restricted to members of that VO.

Access to a portal that allows the decoding of the anonymised name into a person's DN is restricted to individuals in the VO appointed to be VO Resource Managers.

Appropriately authorised individuals of the Grid Operations and Grid Security Operations teams have read access to the user-level accounting data for the purposes of operational troubleshooting, debugging and/or security incident response.

7 THE PERIOD OF RETENTION

The Sites are responsible for deleting the local accounting records according to local personal data retention policy. This needs to be long enough to ensure that all records have been successfully transferred to the ADC database.

The ADC is responsible for deleting the copies of the individual accounting records in the central database, or for removing or anonymising personal identifying information, e.g. the CommonName or e-mail components of subject DNs, from these records, at the latest 18 months after receipt of the data in the ADC. Personal identifying information, e.g. the CommonName component, contained in aggregated data must be treated in the same way.

8 PUBLICATION OF THE INFORMATION

The ADC publishes accounting data on its web portal. Appropriate access control for all published data must be agreed between the Grid and the VO.

The ADC provides aggregated data on CPU usage per Group and Role as defined in the Virtual Organization Management Service.

The ADC publishes user-level accounting data to the authorised VO Resource Managers, if requested by the VO.

9 PROTECTION OF THE INFORMATION

The Site managers and resource administrators are responsible for the secure storage of the local accounting data. Appropriate access control mechanisms must be used to prevent unauthorised access.

The ADC must implement appropriate technical and organisational measures to protect the accounting database and the accounting web portal.

10 TRANSFER OF THE INFORMATION ACROSS INTERNATIONAL BORDERS

The individual job accounting records are transferred between the sites and the central database at the ADC. Many of these transfers cross international borders. Personal identifying information must be encrypted before it is sent across the network, so it is not possible to derive the identity of the user. Multiple records from jobs from the same user contain different cipher text for the DN.

Individual job accounting records and/or aggregated data may be transferred between ADCs, if this is required by the Grid(s). Many of these transfers will cross international borders. Personal identifying information must be encrypted before it is sent across the network, so it is not possible to derive the identity of the user.

The ADC database must be located in a country that has adequate protection of personal data as defined by the EU Directive 95/46/EC (*http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML*), e.g. the European Union, The European Economic Area and some other countries. There must be no transfer of this data to countries outside of this area.

11 ACCESS BY THE USER TO THEIR OWN INFORMATION

A user has the right to access her/his own accounting records and a similar mechanism must be implemented as for the VO Resource Manager to access the aggregated user information. It must also be possible to correct that data if she/he can justify/prove that the stored data is wrong. In that case she/he must contact the corresponding Site Manager who is responsible for notifying any third party to whom the data has been sent, including the ADC. In case of agreement the data in the database must be corrected.

12 SIGNATURE OF AN AGREEMENT RELATED TO THE HANDLING OF THE INFORMATION

Any person having access to the user-level details from more than one site must sign a copy of this document to confirm that she/he understands what can and can not be done with the user related information from the database.

Information that they gain from such access must not be disclosed to anyone who does not have legitimate access to such data.