Joint Security Policy Group

Grid Policy on the Handling of Logged Personal Information

Date:	**29 April 2007**
EDMS Reference:	https://edms.cern.ch/document/840299
Internal Version:	**V0.3**
Status:	**Draft**
Author:	**Joint Security Policy Group**

Document Log			
Issue	**Date**	**Author**	**Comment**
0.1	12 March 2007	David Kelsey	First draft based on earlier presentations to WLCG GDB
0.2	1 April 2007	David Kelsey	Following the March JSPG meeting, expand scope to include all logged operational information, e.g. accounting, monitoring and auditing. Basic ideas and structure
0.3	29 April 2007	David Kelsey	Draft for discussion by JSPG

1 Introduction

This document presents the policy on the handling of logged information in the Grid. Logged information is any operational data produced and stored by the Grid resources or services, by the execution of jobs or by the transfer or storage of data. This includes, for example, all accounting data, audit data, debug log files, monitoring data, system log files and security log files.

For most Grid sites there are laws which apply to data privacy and protection related to the storage of and access to data which is linked to a person's identity.

Each type of logged information in the Grid must be classified into one of four classes. Any type of logged information classified as "personal information" must have a document addressing the issues defined in this policy.

Appendix 1 addresses the handling of user-level accounting data.

2 Definitions

(or do we just refer to the top-level document?)

Grid

Services

Resource Admin

Site

VO

VO resource manager

Users

Etc…

3 Scope

This policy applies to all Grid logged information wherever it is produced and without regard to which service produced it. This therefore includes logged information generated by services run not only by sites but also by VOs or by jobs submitted by users.

This policy does not, however, apply to application data produced during the execution of Grid jobs or to application data produced, owned or used by VOs and/or users. Any issues related to private or personal information contained within application data must be addressed in a separate document prepared by the VO or user.

4 Classification of Logged Information

Each type of logged information must be assigned to one of the four classes presented below.

4.1 Private Information

Private information is any data which contains sensitive personal identifying information, e.g. nationality, passport number, date of birth, credit card number, home address, national identity number, personal telephone number etc.

In general the Grid does not create, collect, store or handle such data. This class is not considered further in this policy. In cases where private information is handled, e.g. during VO user registration, the policy on the handling of such data must be described in a separate document.

4.2 Personal Information

Personal information is any data which contains one or more of the name, e-mail address or X.509 certificate Distinguished Name of a user, but does not contain any private information.

4.3 Non-public Information

Non-public information is any data which does not contain private or personal information, but which, for reasons of confidentiality or other security concerns, needs access to be restricted to within a VO or within a site or to specified individuals. While the handling of this class of data requires special care, it is not considered further in this policy.

4.4 Public Information

Public information is any data not included in any of the above classes for which controlled access is not required. This information may therefore be considered as “world readable” and is therefore not considered further in this policy.

5 Policy requirements for logged personal information

Each type of logged personal information must be described in an appendix to this document or a separate document and must address the following issues:

a. The purpose and reasons for the collection and storage of the information

b. What is stored and where

c. How is user consent obtained

d. Ownership of and access rights to the information

e. The period of retention of the information

f. How the information is published

g. How the information is protected

h. Is the information transferred across international borders and any special considerations related to this

i. How do users access, contest and correct incorrect information

j. Which people need to sign an agreement related to the handling of the information

The user-level accounting data for Grid jobs is classified as logged personal information and is addressed in Appendix 1.

Appendix 1: Policy on user-level accounting for Grid jobs

Each job executed on a Grid resource produces an accounting record. The schema for this accounting record is an international standard and is discussed and defined by the Usage Record Working Group (UR-WG) in OGF. This schema includes the X.509 certificate Distinguished Name of the submitting user and this information is therefore classified as personal information.

1.1 The purpose and reasons for the collection, 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.

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 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.

1.2 What is stored and where

Each site must collect and store an accounting record for each job executed at their site. These records are stored locally at the site. (how much detail?)

Each site is responsible for sending its accounting records, with user DN encrypted, to a central data base at the GOC on a regular basis, e.g. daily.

The central accounting database located at the GOC contains all the individual job records from each of the sites submitting such records.

The GOC is responsible for the generation of aggregated statistics from the data and the publication on its portal.

1.3 How is user consent obtained

User consent for the collection and handling of accounting data is obtained during their first registration or subsequent renewal with their VO. During registration users must accept the conditions of the Grid Acceptable Use Policy. This AUP has a clause on logged information.

1.4 Ownership of and access rights to the information

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

The GOC central database containing individual job accounting records and aggregated data is owned by the Grid management.

A number of GOC staff are authorised to have access to the individual job records. All other persons have no access.

Access to aggregated data at the VO level is public information and requires no access control. Access to VO group/role aggregated data 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 can not easily be connected to a user name. Access to this data is only for members of the VO.

Access to a portal that allows the decoding of the encrypted name into a person's name will be restricted to individuals in the VO appointed to be VO resource managers.

1.5 The period of retention

The Sites are responsible for deleting the local accounting records after one year.

The GOC is responsible for deleting the individual accounting records in the central database after one year and for preserving the aggregated data for as long as this is required by the VO or Grid management.

1.6 How the information is published

The GOC is responsible for the publication of the accounting data on its web portal.

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

1.7 How the information is protected

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

The GOC is responsible for the generation of the aggregated statistics from the data and the publication on its portal.

The GOC is responsible for the maintenance, protection and curation of the central accounting database. Appropriate access control mechanisms must be used to prevent unauthorised access.

1.8 Is the information transferred across international borders

The individual job accounting records are transferred between the sites and the central database at the GOC. Many of these transfers cross international borders. The user name is encrypted before it is sent across the network. From just one accounting record 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.

Transfer to non-EU countries?

1.9 How do users access, contest and correct incorrect information

A user has the right to access her/his own accounting records and the same mechanism must be implemented as for the VO Resource Manager to access the aggregated user information. The GridSite front-end can make this information available based again on the user's credentials. Not only must the individual user to be able to see which data is stored it must also be possible to change that data if she/he can justify/prove that the stored data is wrong. In that case she/he must contact the corresponding VO Resource Manager and the authorized GOC managers and with such a request. In case of agreement the data in the database must be corrected. In case no agreement can be reached, the VO Resource Manager decides.

1.10 Which people need to sign an agreement related to the handling of the information

The VO manager must sign a copy of this document to confirm the appointment of an authorised VO resource manager. The VO resource manager must sign the same 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.

It is strictly forbidden for VO resource managers to expose any user's name to any un-authorised person. The personal information must be handled with care and only be used to resolve problems.

The GOC and or Grid Security Operations staff who have access to the central accounting database must also sign to confirm that they authorized GOC persons and that they understand and accept the requirements of this policy.

It is strictly forbidden for authorised GOC staff to expose any user's name to any un-authorised person. The personal information must be handled with care and only be used to resolve problems.