

EUROPEAN MIDDLEWARE INITIATIVE

MJRA1.6 - FILE CATALOGUE COMMON FRONT-ENDS

EC MILESTONE: MS31

Document identifier:	EMI-MJRA16-1277535- File_Catalogue_Common_Frontends-v1.0.doc
Date:	31/07/2011
Activity:	JRA1
Lead Partner:	DESY
Document status:	Final
Document link:	http://cdsweb.cern.ch/record/1277535

Copyright notice:

Copyright (c) Members of the EMI Collaboration. 2010-2011.

See <http://www.eu-emi.eu/about/Partners/> for details on the copyright holders.

EMI ("European Middleware Initiative") is a project partially funded by the European Commission. For more information on the project, its partners and contributors please see <http://www.eu-emi.eu>.

This document is released under the Open Access license. You are permitted to copy and distribute verbatim copies of this document containing this copyright notice, but modifying this document is not allowed. You are permitted to copy this document in whole or in part into other documents if you attach the following reference to the copied elements: "Copyright (C) 2010-2011. Members of the EMI Collaboration. <http://www.eu-emi.eu>".

The information contained in this document represents the views of EMI as of the date they are published. EMI does not guarantee that any information contained herein is error-free, or up to date.

EMI MAKES NO WARRANTIES, EXPRESS, IMPLIED, OR STATUTORY, BY PUBLISHING THIS DOCUMENT.

Delivery Slip

	Name	Partner / Activity	Date	Signature
From	Patrick Fuhrmann	DESY	22/07/2011	
Approved by	PEB		02/09/2011	

Document Log

Issue	Date	Comment	Author / Partner
1	22/7/2011	Initial Version	Patrick Fuhrmann/DESY
2	29/7/2011	Corrections after PO and TD review	Patrick Fuhrmann/DESY
3	02/09/2011	v1.0 PEB approved version	PEB

Document Change Record

Issue	Item	Reason for Change
1		

References

R1	CMS Dataset Bookkeeping Service http://cmsdbs.cern.ch/
R2	CMS Trivial File Catalogue http://indico.cern.ch/getFile.py/access?contribId=55&sessionId=11&resId=0&materialId=slides&confId=5490
R3	EMI LFC-UNICORE integration DJRA1.2.2 Data Area Work Plan and Status Report (M12) Section 3.6 http://cdsweb.cern.ch/record/1277616/files/EMI-DJRA1.2.2-1277616-Data_Area_Work_Plan_M12-v1.0.pdf
R4	EMI Catalogue synchronization DJRA 1.2.2 Data Area Work Plan and Status Report (M12) Section 3.5
R5	gLite LFC https://svnweb.cern.ch/trac/lcgdm/wiki/Lfc/User/Tutorial
R6	EMI WebDAV frontend evaluation for LFC and SE's DJRA 1.2.2 Data Area Work Plan and Status Report (M12) Section 4.14

MILESTONE REPORT

This milestone was intended to indicate the availability of a standard interface to access file catalogues, which should simplify the integration with user frameworks, portals and third party components.

Background

The data models of the WLCG experiments assume a worldwide distribution of their datasets to ensure overall availability and accessibility. Consequently, mechanisms are used to make the various locations of the data available to file transfer and job submission services in order to transfer data between sites and to direct job submissions to sites where most of the required data is stored. One possibility to keep track of data locations is the use of catalogues, which is the preferred mechanism of most of the WLCG experiments. Over time, the different experiments have built different implementations of such data location services into their frameworks. The goal of this milestone was to allow access to those catalogues by defining and implementing a common interface in order to simplify the integration with user frameworks, portals and third party components.

Results

When evaluating the possibility of providing a common access protocol for the various experiment-specific catalogues, it was concluded the effort would be of no significant benefit to the stakeholders. Only the experiment computing frameworks have direct access to the catalogues and there is no intention to share those catalogue implementations amongst experiments.

For example, this is how CMS and ATLAS are handling the location of their data:

- The CMS experiment is using a central catalogue, the CMS Data Bookkeeping Service (DBS) [R1], which maps logical file names to locations. The LFN to physical filename mapping is done through the CMS Trivial File Catalogue (TFC) [R2], which is part of the CMS software distribution.
- In ATLAS, the Data Management system (Don Quijote) DQ2 is mapping datasets to sites, while the mapping to the final physical location is done through a set of gLite LFCs. ATLAS is currently in the process of centralizing those LFCs.

The EMI software portfolio only contains one file catalogue component, the gLite LFC [R5]. An open C-API for LFC exists and within the EMI-Data activity a Java-API has been adopted, see LFC for UNICORE, DJRA1.2.2 Data Area Work Plan and Status Report Section 3.6 [R3]. The only use case for common access to the proprietary experiment catalogues is the EMI-Data “Catalogue Synchronization” activity, DJRA1.2.2 Section 3.5 [R4]. But instead of forcing the experiments to provide a common standard for accessing their catalogues, a less intrusive approach has been chosen. The “Catalogue Synchronization” framework provides call-outs instead, which are connected to the experiment catalogues by means of adapters, using their specific API’s or mechanisms.

Another interesting approach, in a similar direction, is a new EMI-Data activity for year two [R6]. EMI-Data will evaluate the possibility of providing a WebDAV interface to the gLite LFC and the referred storage elements. That would give access to the entire name space of an LFC through a standard protocol and would make use of the distributed nature of the data.

Summary

EMI-Data concluded the effort to standardize access to the WLCG catalogues could not be justified by its potential benefits. Instead, other mechanisms are used to overcome problems where components need to have access to catalogues other than LFC. For LFC, EMI-Data ensures all EMI components have access to the database and evaluates the possibility of a standard access to the LFC and the referred storage elements.