

# Unlucky for Some

## The Thirteen Core Use Cases of HEP Metadata

*AHM 2005 Proposal*

Tim Barrass

Peter Kunszt, Gavin McCance, Krzysztof Nienartowicz, Ricardo Rocha

Randolph Herber, Adam Lyon, Wyatt Merrit

Mõrag Burgon-Lyon, Tony Doyle, Steven Hanlon, Paul Millar, Richard St. Denis

Solveig Albrand, Jerome Fulachier

James Werner

Carmine Cioffi, Stefan Stonjek

Stephen Burke, Owen Synge

*Bristol*

*CERN EGEE project*

*Fermilab*

*Glasgow*

*LPSC, Grenoble*

*Manchester*

*Oxford*

*RAL*

## Introduction

Use cases are presented that can be considered fundamental to a Grid-like computing system for a HEP experiment and that imply requirements on the metadata catalogues used. High-energy physics requirements documents have been surveyed and core use cases harvested.

Only physicist-level operations are considered; that is, only the requirements of the end user. Furthermore, the term 'metadata' is taken to refer high-level information about jobs, physics events and collections of data (or 'datasets'). In particular, it does not include file and replica information. It is assumed that some lower-level catalogue will exist to perform this function, as laid out in the ARDA document [1].

Information from several experiments has been used to compile this document:

- from the LHC experiments, the HEPCAL I [2] and II [3] documents;
- the CDF use case document, CDF5858 [4];
- "Catalog Services for ATLAS" from the ADA group in ATLAS [5]; and
- "BaBar Analysis Grid Application – Use Case and Requirements Document" [6].

## Use Cases

The identified common use cases are described. The list is not intended to be complete but to provide an overlapping set. It is a proposed common ground for all experiments to meet on and use as a basis for collaboration.

The use cases have been divided into three categories:

- Dataset Handling – operations related to the creation and access of data and metadata.
- Analysis – operations dealing specifically with the needs of physics analysis.
- Job Handling – operations for submitting, querying and controlling jobs submitted to a computing resource.

The use cases are described from the user's standpoint and in terms of metadata implications. Table 1 summarises the identified core use cases and their relationships to the source material.

		Core Use Case	Source Use Cases			
			HEPCAL	CDF	ATLAS	BaBar
Dataset Handling	1.1	Specify a new Dataset	dsreg	4.2	UC1.1, UC2.1, UC2.4, UC3.1, UC4.1, UC4.3, UC4.4	
	1.2	Read Metadata for Datasets	dsmdacc	4.1	UC1.5, UC5.3	U1.5
	1.3	Update Metadata for a Dataset	dsmdupd	Not Allowed	UC5.3	
	1.4	Resolve Physical Data	dsaccess, remdsacc	4.6, 4.7, 4.8	UC5.3	U1.6
	1.5	Download a Dataset to Local Disk	dsdownload	4.5, 4.9		
Analysis	2.1	Run a Physics Simulation Program	simjob	4.4, 4.10, 4.11		
	2.2	Select a Subset of a Dataset	analysis1	4.7	UC4.2, UC4.4	U1.5
	2.3	Run an Algorithm over an Input Dataset	dstrans, analysis1	4.6, 4.9	UC1.4, UC1.5, UC2.2, UC2.3, UC2.4, UC5.1	U1
Job Handling	3.1	Submit a Job to the Grid	jobsub	4.3		U1.9
	3.2	Retrieve/Access Output of a Job	joboutaccess	4.9		U1.14
	3.3	Estimate the Resource cost of Running a Job	jobreest	MW6		
	3.4	Monitor the Progress of a Job	jobcont, jobmon	MW9		U2
	3.5	Repeat a Previous Job	See HEPCAL II	4.6		U4, U5

Table 1: The Thirteen Core Use Cases for HEP Metadata.

## Conclusions

All available HEP use case documents have been surveyed and the most fundamental operations have been identified and described. A general set of HEP metadata operations has been defined. These form a common ground for developers to work from.

## References

- 1: P. Buncic et al., [Architectural Roadmap Towards Distributed Analysis](http://lcg.web.cern.ch/lcg/PEB/arda/public_docs/ARDA_report_final.pdf), 2003  
[http://lcg.web.cern.ch/lcg/PEB/arda/public\\_docs/ARDA\\_report\\_final.pdf](http://lcg.web.cern.ch/lcg/PEB/arda/public_docs/ARDA_report_final.pdf)
- 2: F. Carminati et al, [Common Use Cases for a HEP Common Application Layer - HEPCAL](http://project-lcg-gag.web.cern.ch/project-lcg-gag/LCG_GAG_Docs/HEPCAL-prime.pdf), 2002  
[http://project-lcg-gag.web.cern.ch/project-lcg-gag/LCG\\_GAG\\_Docs/HEPCAL-prime.pdf](http://project-lcg-gag.web.cern.ch/project-lcg-gag/LCG_GAG_Docs/HEPCAL-prime.pdf)
- 3: D. Barberis et al, [Common Use Cases for a HEP Common Application Layer for Analysis - HEPCAL II](http://project-lcg-gag.web.cern.ch/project-lcg-gag/LCG_GAG_Docs/HEPCAL-II.pdf), 2003  
[http://project-lcg-gag.web.cern.ch/project-lcg-gag/LCG\\_GAG\\_Docs/HEPCAL-II.pdf](http://project-lcg-gag.web.cern.ch/project-lcg-gag/LCG_GAG_Docs/HEPCAL-II.pdf)
- 4: B. Todd Huffman et al, [The CDF/D0 UK GridPP Project](http://www.gridpp.ac.uk/datamanagement/metadata/SubGroups/UseCases/docs/cdf5858.ps.gz), 2002  
<http://www.gridpp.ac.uk/datamanagement/metadata/SubGroups/UseCases/docs/cdf5858.ps.gz>
- 5: David Adams, [Catalog Services for ATLAS](http://www.usatlas.bnl.gov/ADA/docs/ada_catalogs.pdf), 2004  
[http://www.usatlas.bnl.gov/ADA/docs/ada\\_catalogs.pdf](http://www.usatlas.bnl.gov/ADA/docs/ada_catalogs.pdf)
- 6: [BaBar Analysis Grid application - Use Case and Requirements Document](http://www.slac.stanford.edu/BFROOT/www/Computing/Offline/BaBarGrid/usecases.pdf), 2001  
<http://www.slac.stanford.edu/BFROOT/www/Computing/Offline/BaBarGrid/usecases.pdf>