Direct User Support and updates on EGEE Use Cases documentation

Marco Bencivenni INFN-CNAF

EGEE09 Conference – Barcelona 22/09/2009

www.eu-egee.org

e-infrastructure

EGEE and gLite are registered trademarks

Outline

- **Direct User Support (DUS) introduction**
  - Purpose and organization
- **EGEE Use Cases Documentation**
  - Format and contents
- **New Use Cases**
  - Some examples

DUS Mandate

- Our purpose is to consolidate and co-ordinate support activities centered on specific user requests in EGEE concerning generic and application related documentation.
- The DUS group responsibilities are related to the direct support of user requests on a daily basis.
- Implementation through the establishment of
  - Separate support unit within the GGUS system
  - Focus on documentation development, use cases and documentation improvement.

Who are we?

- **Diverse team with ~15 people from**
  - CSIC, UPV (Spain)
  - CNRS (France)
  - INFN (Italy)
  - VUB (Belgium)
  - GRNET and Partners (Greece, Bulgaria, Romania, Israel)

- **Co-ordinated by Christos Filippidis and Christos Markou (NCSR Demokritos, Athens, Greece)**

Documentation Type

- **Three types of documentation**
  - The **User Guide** provides basic information on how to use the grid services
  - **Use Cases** are mini-tutorials on how to perform common tasks
  - More detailed information about gLite services can be found in the **individual service manuals**

Use Cases documentation

Two main activities:

- Reviewed the entire set of Use Cases (produced by the User Information Group during EGEE II): updated, corrected and added some new use cases
- Converted the entire documentation to DocBook format

- **You can find the new Use Cases documentation here: http://www.italiangrid.org/usecases/Index.html (momentary URL)**
  - the definitive URL probably will be http://technical.eu-egee.org/index.php?id=363

- **You can find all the files we used to obtain the documentation (xml, html, pdf, xsl etc.) here: svn+ssh://svn.cern.ch/reps/DUS-NA4/trunk/UIG**

Why Docbook

- **DocBook is a semantic markup language for technical documentation**
- **It is *multi-mode rendering*: can be published in a variety of formats: HTML, XHTML, PDF etc. without requiring users to make any changes to the source**
- **Tools for the conversion:**
  - **html2dbk**: xhtml → xml
  - **dblatex**: xml → pdf/ps/dvi/rtf
  - **xsltproc**: xml → html

Documentation outline

- **The sequence of the chapters:**
  - Chapters **from 1 to 10** go through all the basics for setting up the user to be ready to use the grid
  - Chapters **from 11 to 18** are for advanced users and feature detailed, in-depth information on more complex concepts and operations
  - Chapters **from 19 to 21** describe some particular Life Sciences community use cases focusing on Data Management

- **A visual formatting is used to render this document: command, output, technical term, code etc.**

Chapters reviewed

- **Chapter removed: Low Latency (SDJ) Scheduling (not used)**
- **Chapters joined: Simple Job Cycle + Retrieving Job History + Job-compatible Resources**
- **Chapters modified: not important modifications, but all the chapters have been updated. In particular old links, new commands, variables and other details**

New chapters

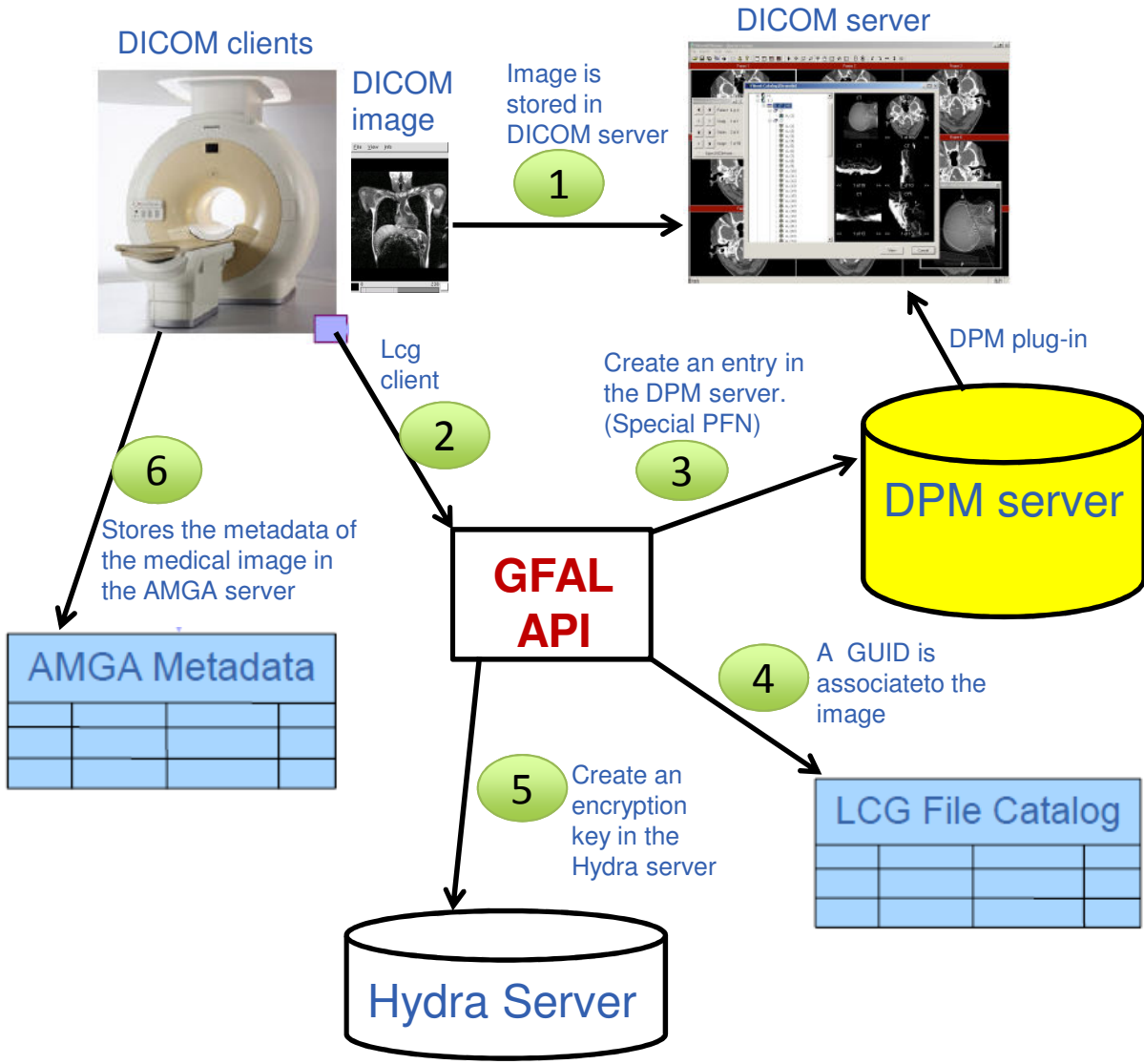
- **Three new chapters to cover specialized tools developed by the Life Science community to satisfy specific requirements**
  - These tools have been developed having in mind specific requirements from the LS community although they could be good examples of interface with gLite middleware for a broad range of scientific applications (see next slides)
  - The tool developers gave us all the documents that we reviewed and adapted focusing on how communities exploit Data Management services

Medical Data Manager

- **The Medical Data Manager (MDM) is an interface between DICOM (Digital Image and COmmunication in Medicine) compliant storage and the gLite middleware**
- **It aims at:**
  - providing access to medical data sources for computing without interfering with clinical practice
  - ensuring transparency so that accessing medical data does not require any specific user intervention
  - ensuring a high data protection level to preserve patients privacy
- **EGEE middleware enables the federation of many DICOM servers geographically distributed and provides a unified view of the data archived**

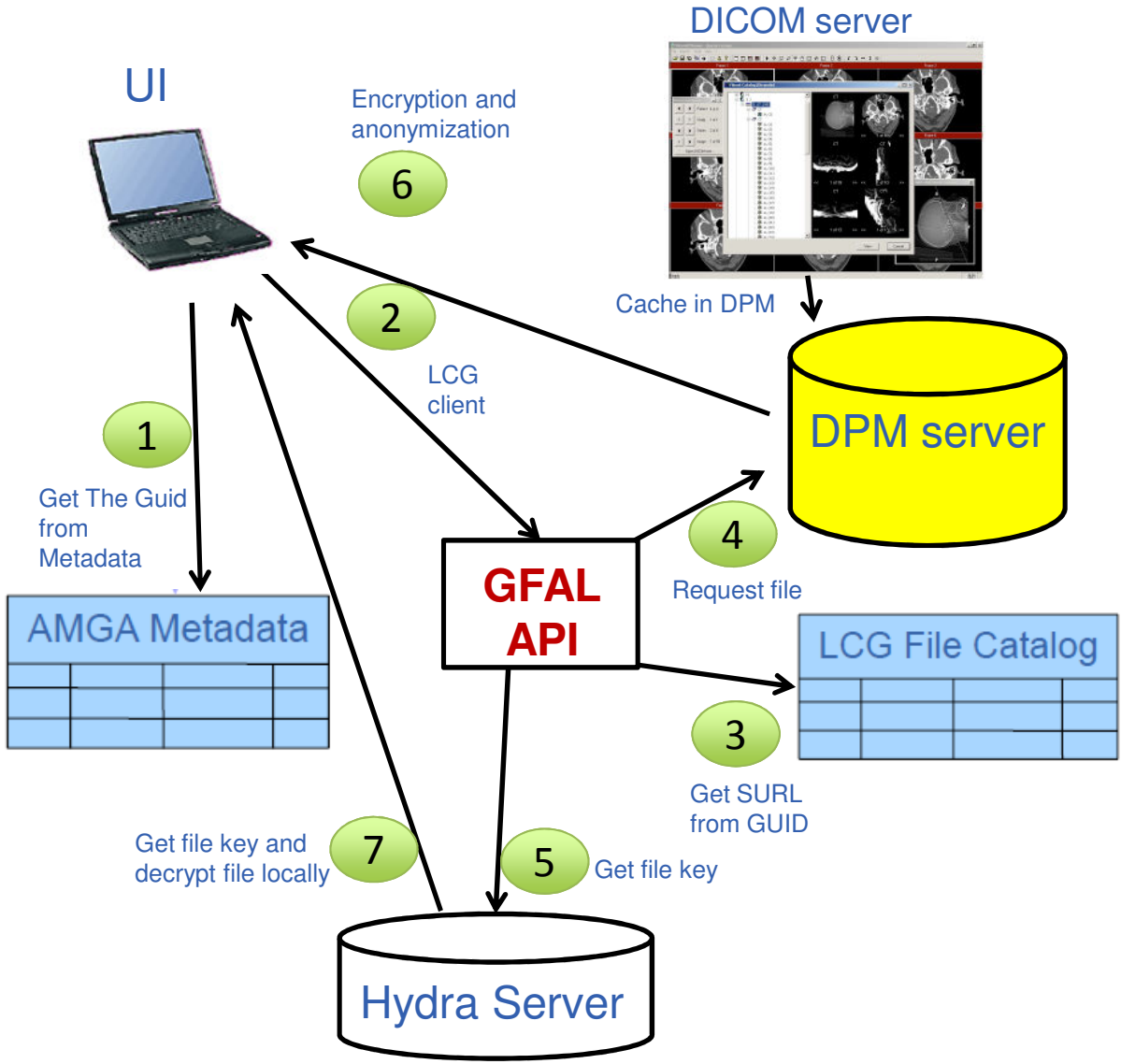
MDM Use Case 1

- **Recording a DICOM image on the grid**
- Retrieving a DICOM slice
- Retrieving a 3D image
- Removing an image
- How to give permissions to another user



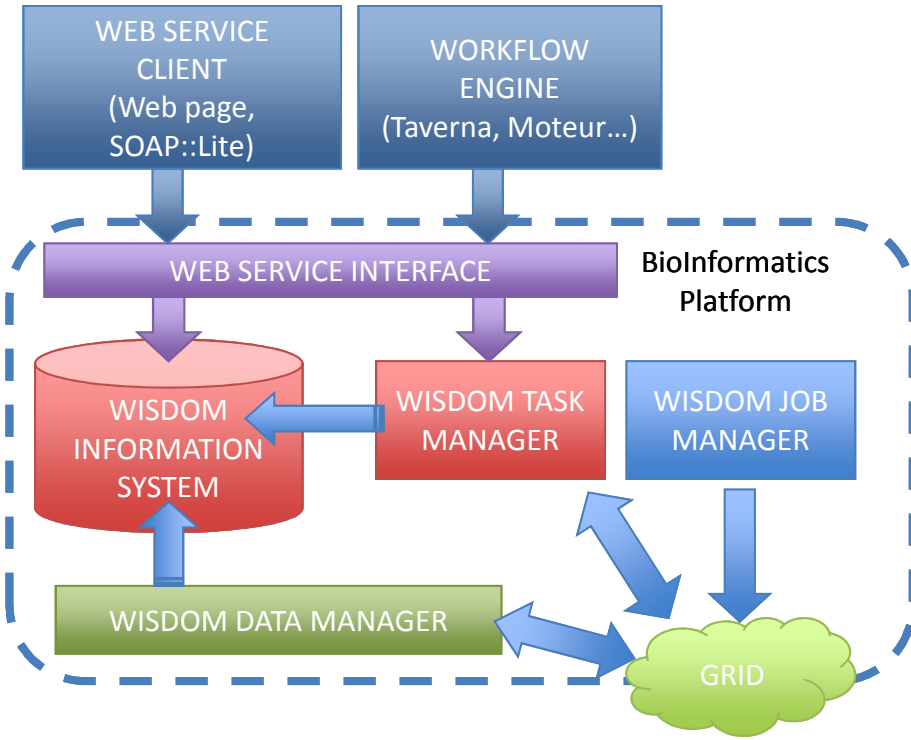
MDM Use Case 2

- Recording a DICOM image on the grid
- **Retrieving a DICOM slice**
- Retrieving a 3D image
- Removing an image
- How to give permissions to another user

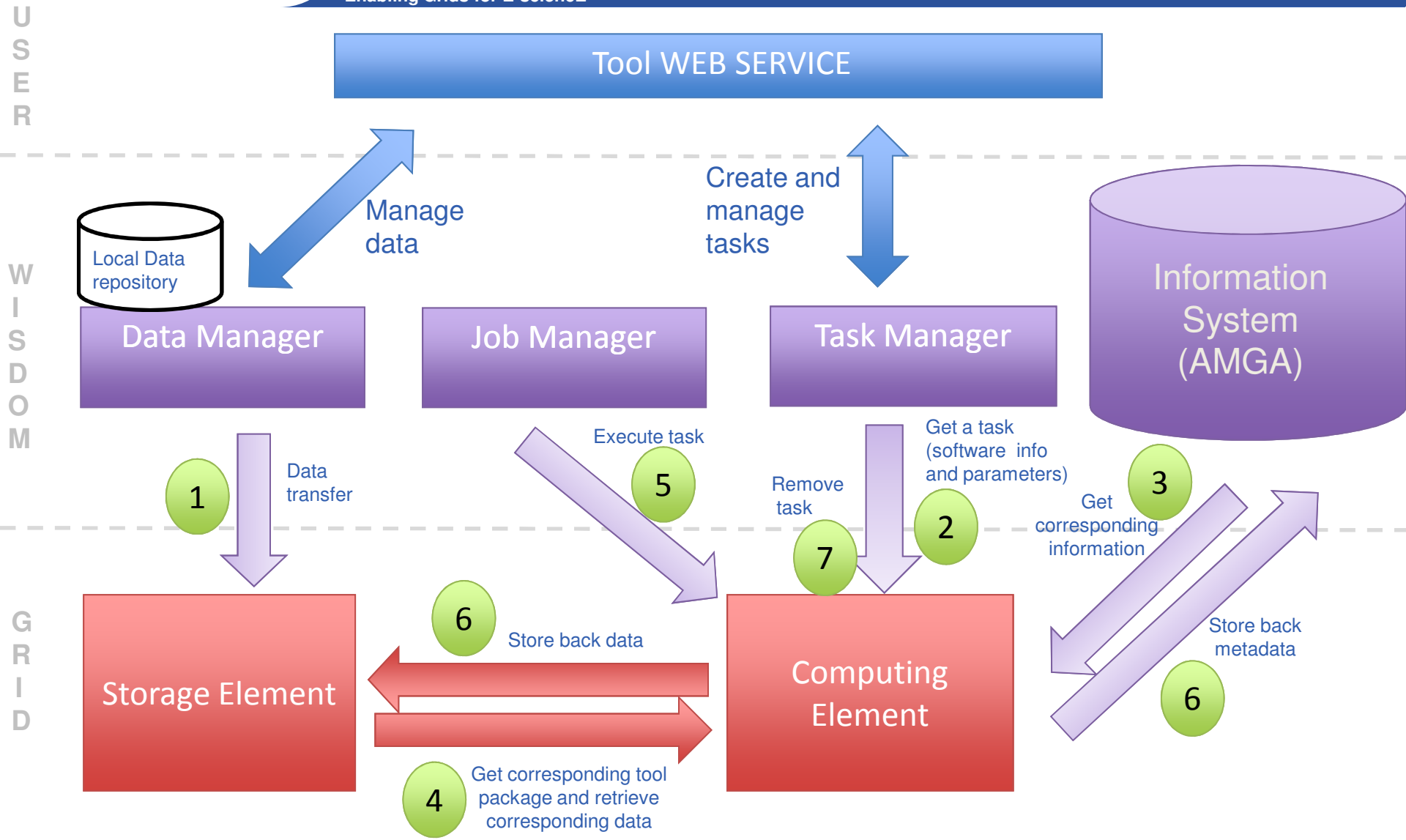


Wisdom Data Manager

- **Is part of the Wisdom Production Enviroment**
- **Can be used and integrated with basically any middleware**
- **Provides an interface with the grid Data Management**
  - deploy automatically files on multiple infrastructures
  - replicate them on those infrastructures
  - resolve the physical locations for each of these infrastructures
- **Users have only to specify:**
  - a file repository they want to deploy on the grid
  - how many replicas of the repository
  - on which grid infrastructure

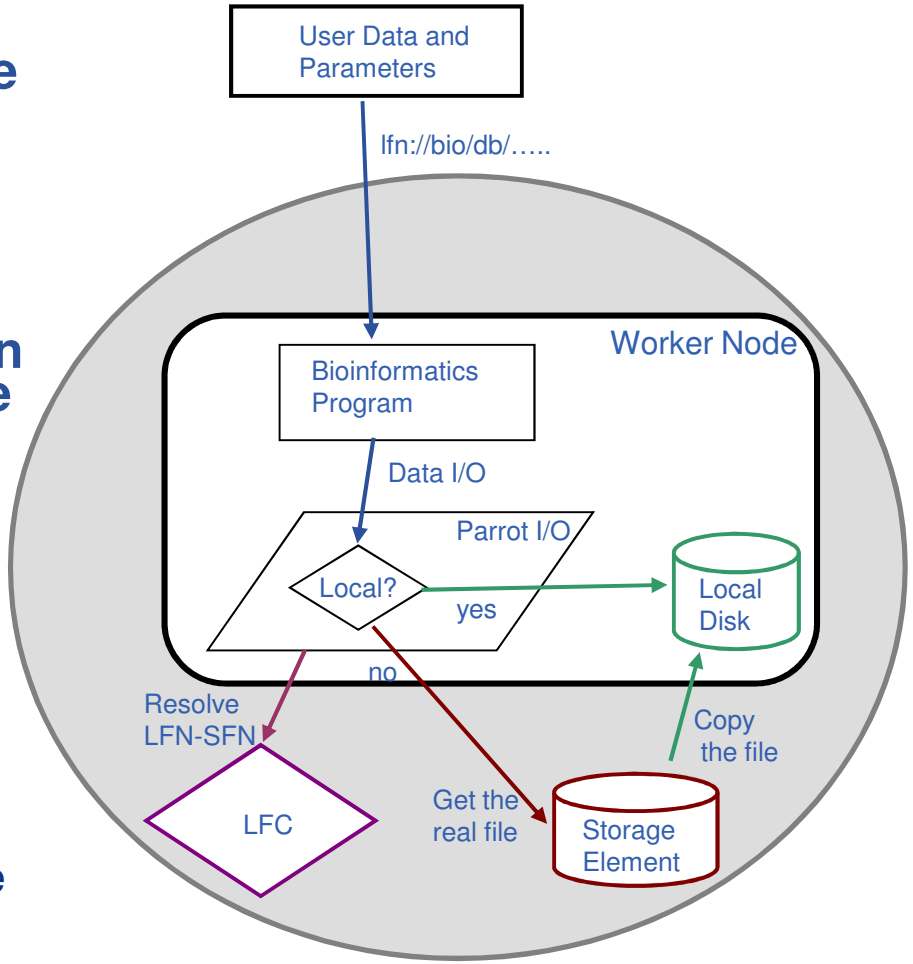


Wisdom Workflow



Bioinformatics Data Manager

- **Grid computing as a solution to distribute and integrate large bioinformatics databases and to make them usable by bioinformatics programs**
- **A customized version of Parrot has been used to connect applications to the EGEE Data Management system in order to support a logical name space**
- **Two ways to access to biological databases:**
  - remote I/O
  - file replication on local WN
- **For large files the remote I/O mode is preferred because:**
  - no matter about free local space
  - too late to check of the free space when the job is on the WN



Future work

- Any new use case to be added?
- Improvement of the process and tools used to convert the xml to other formats
- We will provide instructions about the process and tools used for convertion of xml files to other formats on a web page

Acknowledgements

- Vangelis Floros
- Cal Loomis
- Christos Markou
- Christos Filippidis
- Johan Montagnat
- Romain Texier (Medical Data Management)
- Jean Salzeman (Wisdom)
- Christophe Blanchet (Bioinformatics)
- Frank Harris

Questions