

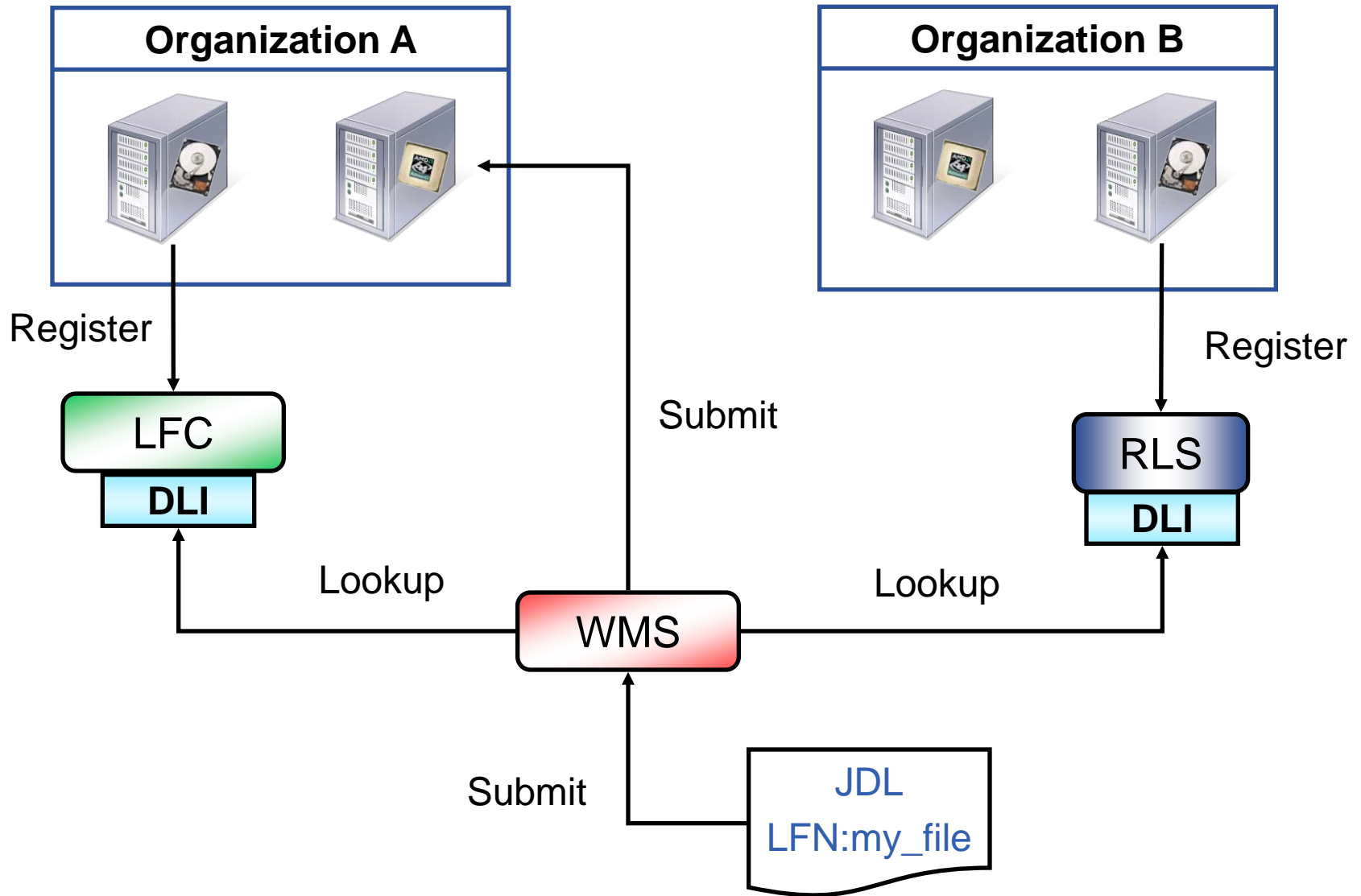
The Data Location Interface

Laurence Field
OGF 25

- **Data is required as input to the computation activity**
- **Small files can be transferred with the job**
 - Described using the InputSandbox and OutputSandbox JDL attributes
 - The values of these attributes are the filenames to be transported
- **However, data resources are typically distributed geographically**
 - Jobs are steered to a “close” computational resource
 - Or vice versa
- **A description of the data and its location is required for a job**
 - In EGEE this is achieved using the InputData JDL attribute
- **Specific files are described using a Storage URL (SURL)**
 - Provides information about the location of a specific file.
 - `srm://storage.host/my_namespace/my_file.txt`
- **SURLs are typically resolved into Transport URLs (TURLs)**
 - The information required to retrieve a physical file
 - `gsiftp://diskserver.host:2811/internal_path/internal_file_name`

- **There may be many replicas of a particular file**
 - The locations of these replicas are recorded in file catalogs.
- **A file is registered in a catalogue**
 - And assigned a Global Unique Identifier (GUID). T
- **Logical File Name (LFN) are usually assigned to a GUID**
 - The resolution from LFNs and GUID to SURLS/TURLs
 - Achieved by querying the catalogue
 - Described using the ReplicaCatalog attribute in the JDL
- **A Logical Data Set (LDS) is a set of physical files**
 - That need to be available as an atomic unit of data
 - The resolution from LDSs to SURLS/TURLs
 - Achieved by querying the catalog.
 - Custom dataset catalogs may provided by a VO
 - Which define their own query mechanisms and interfaces

- **The Data Location Interface (DLI)**
 - Is used where catalogs may present different interfaces.
- **It is a SOAP based request-response protocol**
 - Between the client and a Data Catalogue.
- **Generic Query Strings may need to be used.**
 - Which are resolved by the catalog
 - Returning the locations of the requested datasets
- **The DLI implementation provides an plugin API**
 - To simplify interfacing to different catalogues



- **Document**

- <http://cmsdoc.cern.ch/cms/grid/docs/DataLocationInterface.pdf>

- **WSDL**

- <http://glite.cvs.cern.ch:8180/cgi-bin/glite.cgi/LCG-DM/dli/DataLocationInterface.wsdl>

- **Paper**

- <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.60.7414&rep=rep1&type=pdf>