Deployment of Earth Observation Applications on Testbed1.2

5th Datagrid Project Conference

WP6 Meeting Monday 2nd September

Work Package 9 – EO Applications

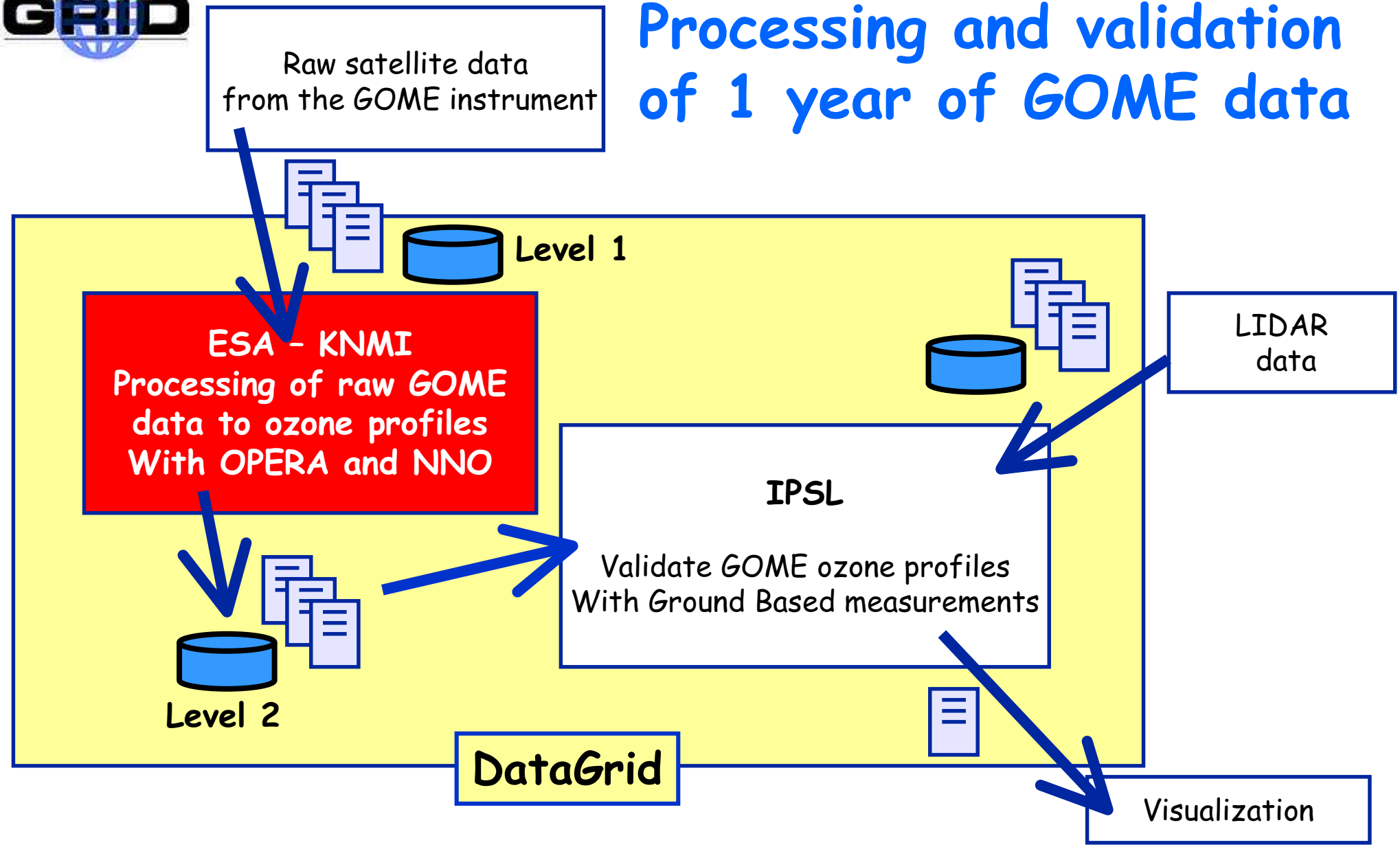
EO GOME Data Processing and Validation by KNMI, IPSL and ESA

- A collaborative effort involving 3 geographically distributed EO sites
- Leve1 data is raw satellite data which has to be analysed to retrieve actual physical quantities : Level2 data
- L2 data consists of measurements of OZONE, actual physical quantities for the ozone gas concentration at different pressure levels within a column of atmosphere at a given location (lat, lon) above the Earth's surface
- we use 2 different methods to process L1 GOME data to produce L2 data
- The L2 product is then validated by IPSL. L2 quantities are compared with quantities measured from ground-based (LIDAR) in coincident area and time. The result shows the difference between the two types of measurements. This is used to estimate the accuracy of the L2 product

Executables and Data Files

- The KNMI executable processes L1 data and produces L2 data. It is written in C++, it is called OPERA
- The ESA executable processes L1 data and produces L2 data. It is written in IDL, it is called NNO
- The IPSL executable compares the L2 data with LIDAR data. It is written in FORTRAN, C and IDL.
- Typical input/output file sizes:
  - Level 1 data, a 15 Mb file contains the measurements taken during a full orbit of the satellite/sensor
  - Level 2 data (produced by L1-L2 processing), a 10-12 kb file containing the results of the L1 data analysis (as described above)
  - Lidar data, a 60 Mb data file containing one month of Local ground-based measurements around Haute-Provence observatory
  - Validation result, a 10 kb file

Processing and validation of 1 year of GOME data



EO Use Case File Numbers

1 Year of GOME data

Data	Number of files to be stored and replicated	Size
Level 1	4,724	15 Mb
Level 2 NNO (ESA)	9,448,000	10 kb
Level 2 Opera (KNMI)	9,448,000	12 kb
Validation Lidar (IPSL)	12	2.5 Mb
Total:	**18,900,736**	**267 Gbyte**

GOME has a data set of 5 years

GOME is relatively large in both size and number of files

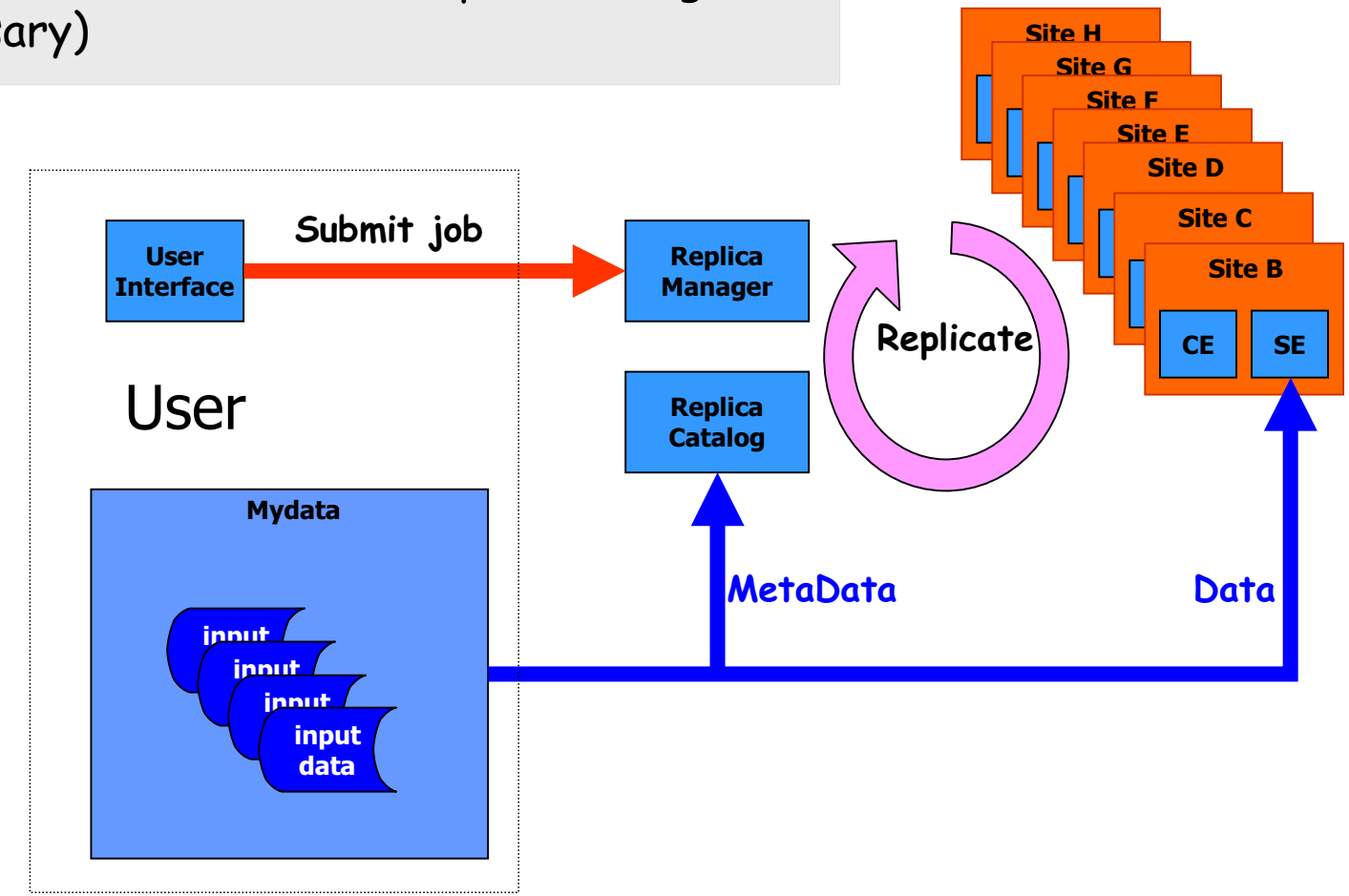
Processing Steps

1. Transfer Level1 (raw) data to the Grid Storage Element
2. Register Level1 data with the Replica Manager
3. Submit jobs to process the Level1 data, produce Level2 data products
   - Jobs running on the CEs locate Leve1 data by using the BrokerInfoAPI
4. Repeat step 1-3 for level 2 products
   1. Transfer Level2 data products to the Storage Element
   2. Register Level2 data products with the Replica Manager
   3. Submit jobs to the Grid to validate Level2 data products
5. Retrieve validation results and visualize on the User Interface

GOME data processing and validation steps

Step 1: Transfer Level1 data to the Grid Storage Element

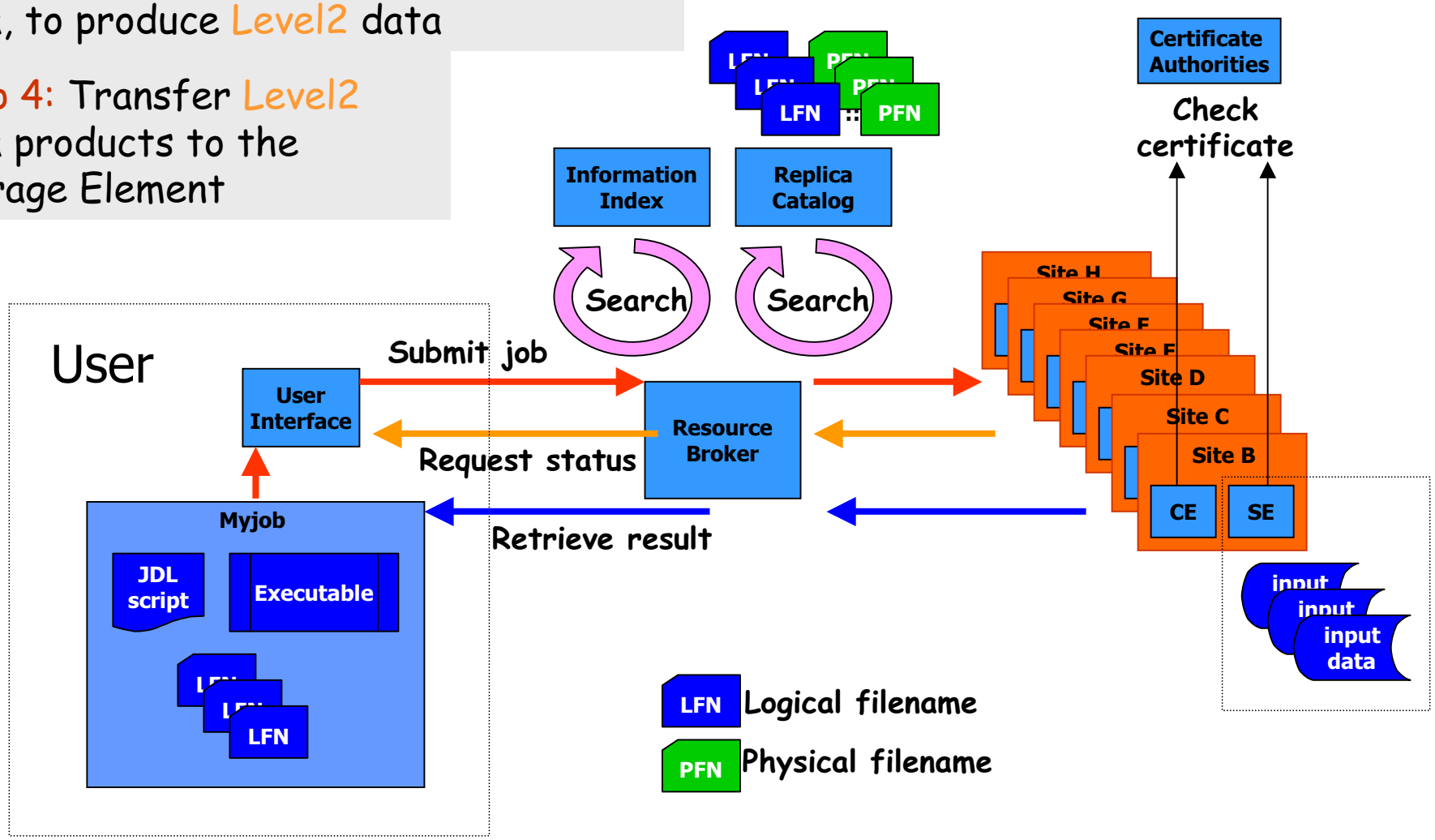
Step 2: Register Level1 data with the Replica Manager (replicate if necessary)



GOME data processing and validation steps

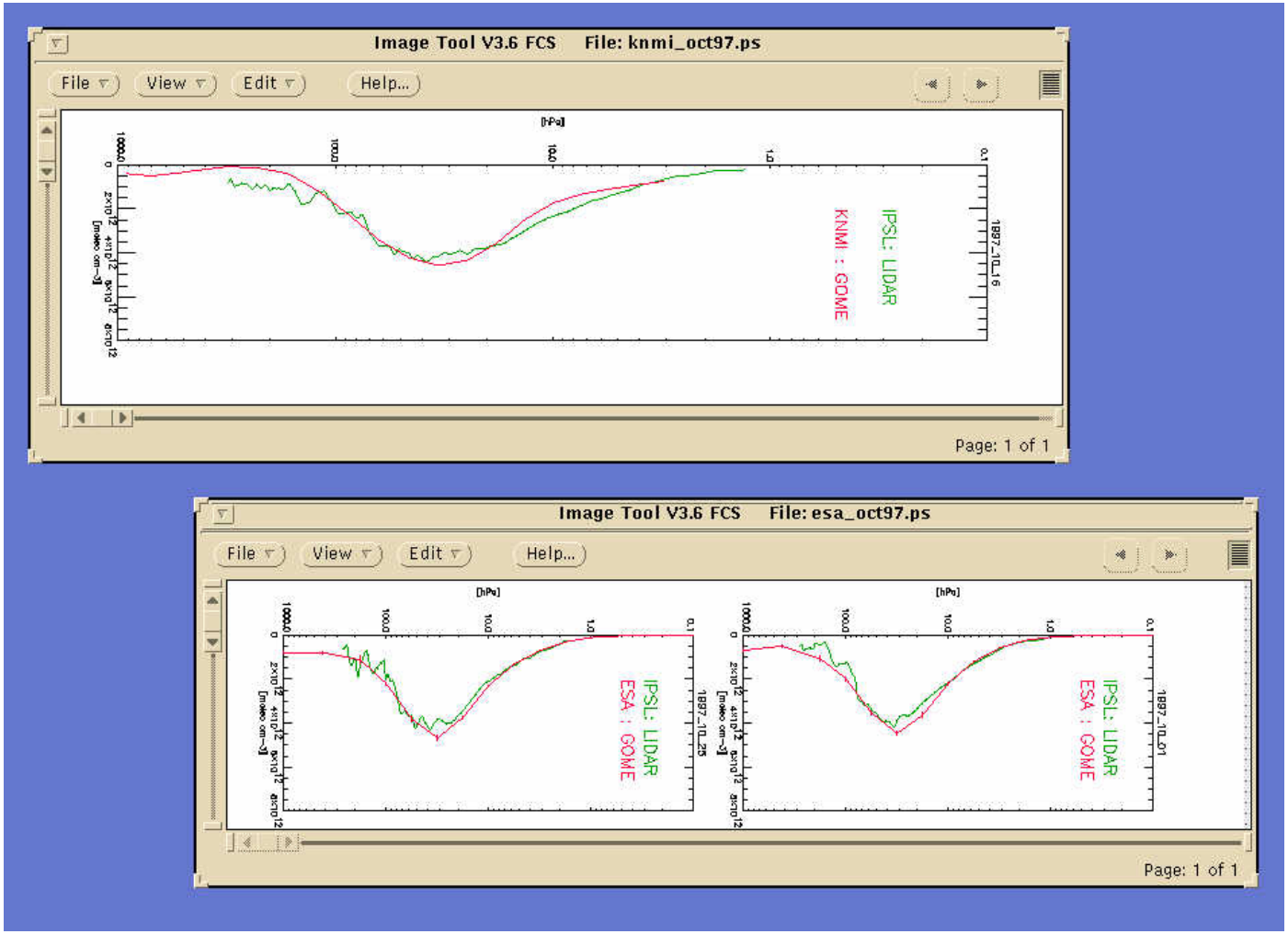
Step 3: Submit jobs to process Level1 data, to produce Level2 data

Step 4: Transfer Level2 data products to the Storage Element

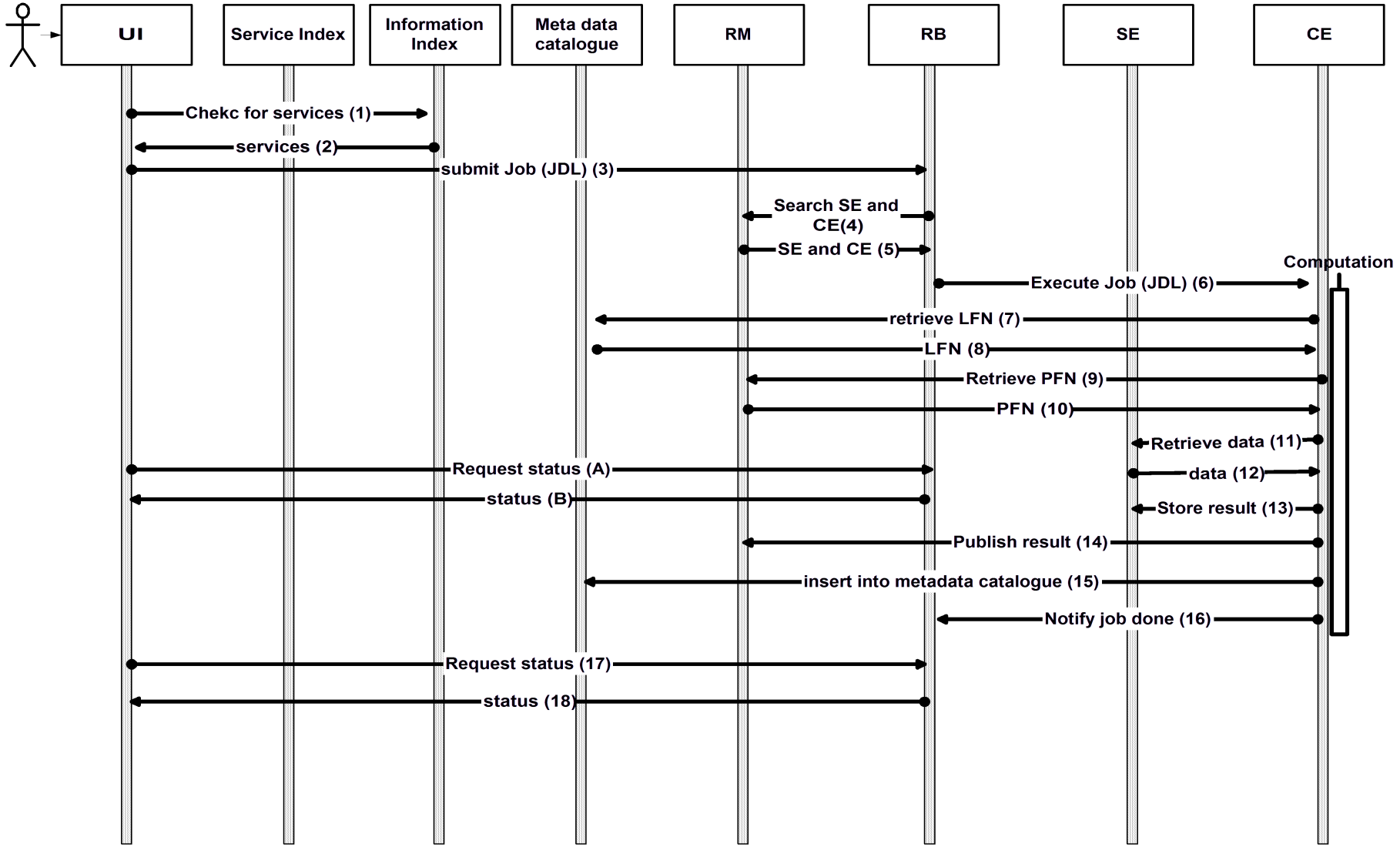


Validation of two different Ozone processing algorithms vs Lidar profiles using the testbed

Step 5: Retrieve and visualise results



Detailed Use Case for L1-L2 processing



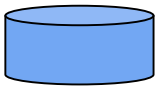
EO Web Portal Prototype

Querying metadata

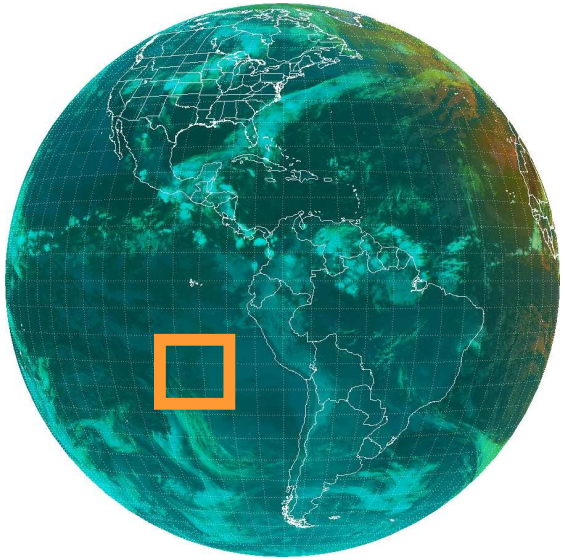
User

Using GRID Portal user selects world region and time frame to be processed

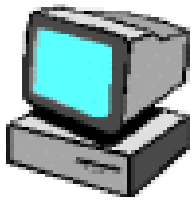
Queries L1 Catalogue and retrieves orbit number and logical name file



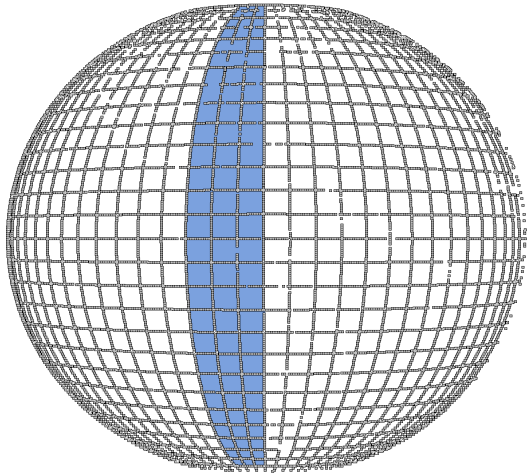
Level 1 Catalogue



Level2 Pixel Data

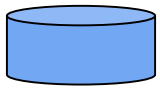


GRID Portal



Level1 Orbit Data

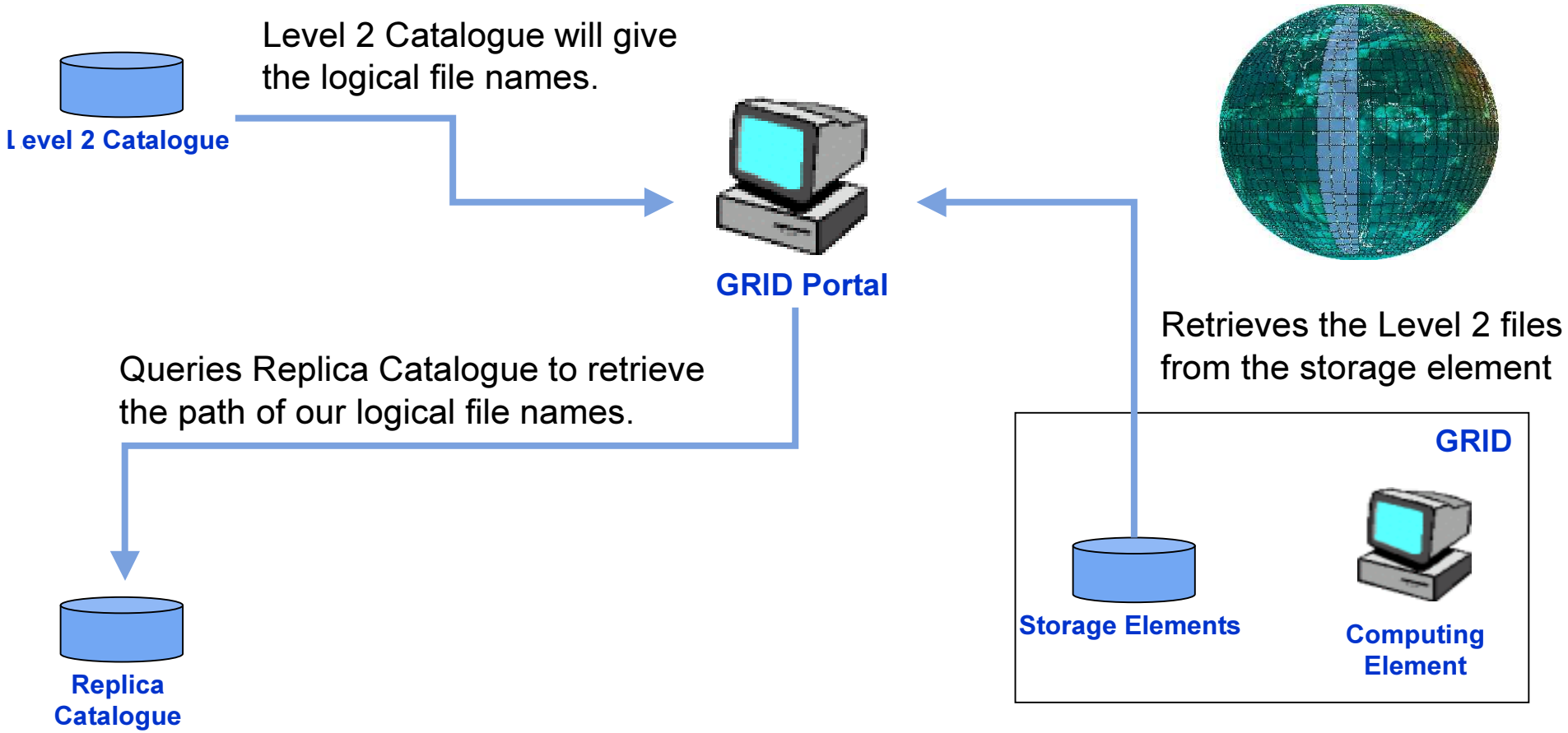
Queries L2 Catalogue to check if the data was already processed



Level 2 Catalogue

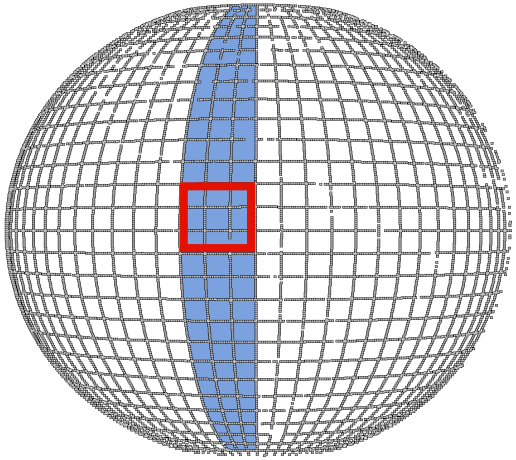
EO Web Portal Prototype

If the data already exists …



EO Web Portal Prototype

If data was not yet processed…



Level 1 Catalogue will give the orbit number*

Level 1 Catalogue

Places Level 1 Data into the storage element

GRID Portal

Launches the Job in the Computing Element

Retrieve Level 1 Data orbit file, extract the requested geographic area

Level 1 Storage (AMS)

GRID

Storage Elements

Computing Element

When completed announces a new level 2 data into the catalogue

Level 2 Catalogue

* Not confirmed

WP9 Plans for using TB1.2

- New dedicated EO Graphical User Interface will investigate use of multi processing and SE distributed datasets
  - multi-job submission capability
  - data transfer & replication interface
  - job / data partitioning capability
- ESRIN CE installation will interface EDG to ENEA Grid (AFS/LSF)
  - first case of single CE to manage two different batch queuing systems
- ESRIN SE will Interface to ESA archive system
  - investigate use of GDMP_STAGE_TO/FROM_MSS
- IPSL CE, SE and graphical user interface installation
- EO Applications will interface to the Grid using EO Web-services
  - integration of Grid services and GIS web-mapping services
- Several demonstrations planned in 4th quarter of 2002
  - Re-run of previous demonstrations using 1.2
  - WP9.4 Use Case - processing and validation chain for 1 year of GOME data
  - EO Application User Interface will be used to submit & monitor execution of hundreds of jobs simultaneously
  - Web-based GOME processing on-demand via access to dedicated EO SOAP services