Earth Observation inputs to ATF

Annalisa Terracina

EU-DataGrid Project
Work Package 9 – EO Applications
15-16 April 2003 CERN

esa

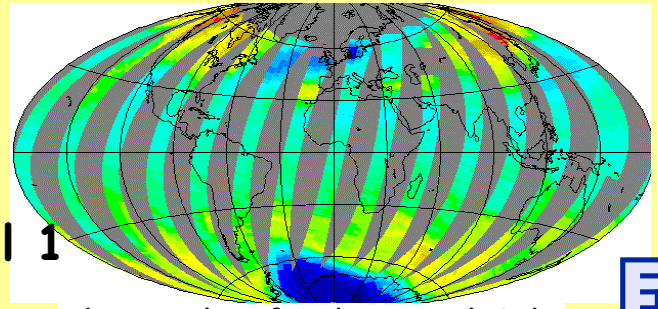
Summary

- ✓ WP9 use cases
- ✓ What is satisfied until now
- ✓ What we expect in release 2.0
- ✓ What can be done in release 2.1

High level use case 1/2

Raw satellite data
from the GOME instrument
(~75 Gb - ~5000 orbits/y)

Level 1

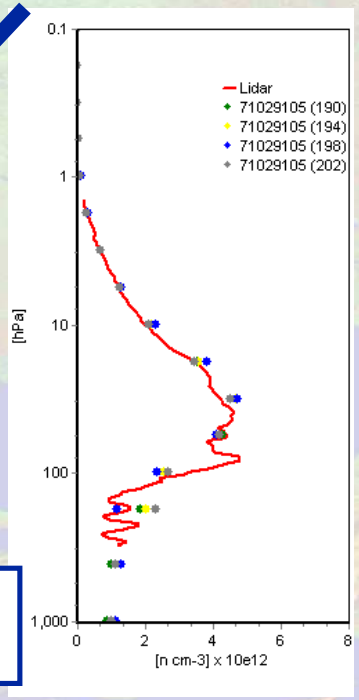


(example of 1 day total O_3)

ESA(IT) KNMI(NL)
Processing of raw GOME
data to ozone profiles.
2 alternative algorithms
~28000 profiles/day

Level 2

LIDAR
data (7 stations,
2.5Mb per month)



IPSL(FR)
Validate some of the
GOME ozone profiles (~10^6/y)
Coincident in space and time
with
Ground-Based measurements

DataGrid environment

Visualization
& Analyze

High level use case 2/2

File numbers for one year of GOME data :

	Number of files (to be stored and replicated):	File Size
Level 1:	4,724	15 Mb
NNO:	4,724	29 Mb
Opera :	9,448,000	12 kb
Lidar	12	2.5 Mb
Total:	9,457,460 files	267 Gbyte

New applications

- GOMOS "Operational" Validation in EDG
  - Interest in Ozone profile science community to extend GOME validation experience to GOMOS (and possibly to other ENVISAT instruments)

  Partners : ESA, ESRIN/DataGrid, IPSL/DataGrid, IPSL/SA, ACRI(company)

- Internal geophysical studies (IPGP/Paris)
  - Validation or treatment with data distributed in several seismic databases located in France, England, Germany and United States… (update : around 100 Mo/day)
  - Detection of the seismic deformations using SAR data
  - Earthquakes simulation (MPI needed)

Very sensitive data : very restricted access → very high priority

Lower level use cases 1/2

1. Use case 1:
   - Loading the GOME level-1 data on a GRID storage element
2. Use case 2:
   - Running the OPERA algorithm on the Grid
     - Grid job execution
     - Data handling
3. Use case 3:
   - Storing the OPERA output on a GRID storage element
4. Use case 4:
   - Storing the OPERA metadata in a database using spitfire
5. Use case 5:
   - Running the NNO algorithm on the Grid

Lower level use cases 2/2

6. Use case 6:
   - Storing the NNO output on a GRID storage element
7. Use case 7:
   - Lidar data and metadata catalogue on the GRID
     - Lidar data
     - Metadata catalogue in a database using spitfire
8. Use case 8:
   - Running the validation algorithm
     - Data handling: Ozone profile file packaging
     - Data handling: running validation algorithm
     - Grid Job execution

What is satisfied until now

- The running of EO application on the EDG testbed 1.4 was almost successful
- All applications from the PM24 usecase could be executed but due to well-known limitations of the testbed the usecase could not be finished entirely

The RC in 1.4 could handle up to 10,000 entries per logical collection: PM24 Use Case needs 9 million entries

What we expect in release 2.0

- Logical File collection problem solved, register millions of files
- WP5 storage space reservation interface for Storage Element, Worker Node and Archive System
- Extension of WP2 Replica Catalog service to provide application metadata services
- Interface between ESA AMS archive system and the WP5 Storage Element
- Programmable API (C++, Java) to interface programs with middleware services
- Procedure for creating new VOs and deleting obsolete ones, procedure for creating and deleting groups within Vos
- Submit thousands of jobs with 99 % of job submitted ending in successful way

List of priorities for release 2.1 1/2

The requirements list that follows is ranked by WP9 needs.The AWG is making an effort to provide a unique list of common requirements and prioritize them.

Very high priority

- Security inside spitfire by certificate (Vo, Group, user)
- Need to control file access rights at the user name level (ACLs)
- Need a comprehensive grid-security implementation (e.g. control access to VO data/resources by multiple user groups with different levels of privilege within the VO)

List of priorities for release 2.1 2/2

High priority

- Implementation of fast turnaround queues (i.e.for small jobs which require instant execution and near-real-time response)
- Improve fault detection and fault tolerance
- Recent system release support (redhat 7.3, redhat 8.0)
- Easy grid file access from running application (gridopen /gridclose / gridread/ gridwrite)
- Better support for directory structure handling inside Replica Manager

Intermediate priority

- MPI support