



Enabling Grids for E-science

Cooperation between GENESI-DR and the EGEE Earth Science Cluster

André Gemünd
Fraunhofer SCAI

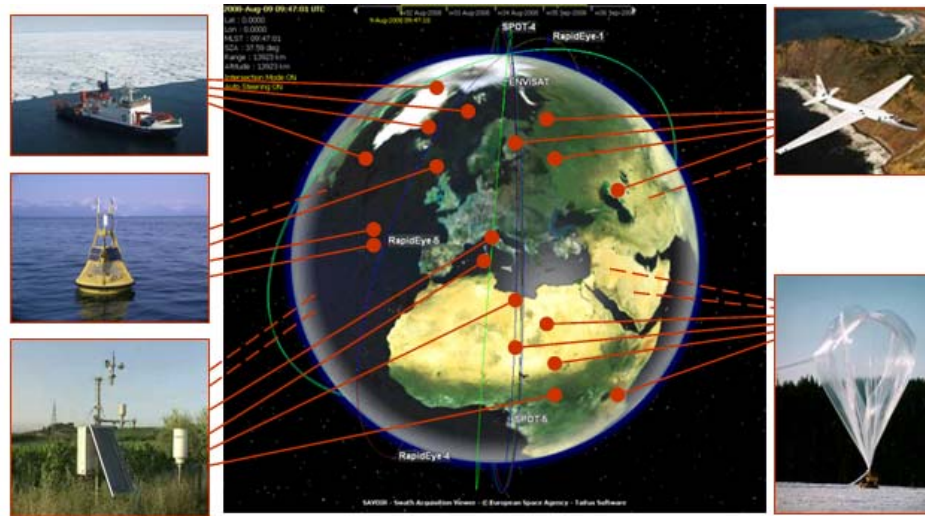
www.eu-egee.org



- **One of the 7 strategic discipline clusters of EGEE III**
 - Improving the situation for earth scientists using the Grid
 - Consulting, Porting applications, deploying services
 - Supporting users
 - Coordinating collaborations with other earth science projects
 - Dissemination in Earth Science community



- Heavily data dependant workflows
 - Diverse set of instruments / sensors
 - Satellites, seismographs, radar stations, weather stations etc.
 - Produced, stored and managed distributedly around the world



- **Usual, manual way:**
 - „discover“ the data
 - Register with the DR and download locally
 - Local preprocessing of data sets
 - Manual upload to Grid storage element of VO with personal certificate
 - Register in LFC and save identifier
 - Send LFN/guid with jobs or in job description



- **Problems thereof**

- User needs to download (from DR) and upload once again (to Grid)
- No automatic discovery
- no easy way to reuse („forgotten“ data)
- Redundant storage / cluttering of storage elements
- Meta data only stored in custom meta data bases (solutions that access a classical data base from the Grid)
 - Not really standardized / centralized
 - Custom schemas
 - Often need additional user registration
 - Support spatial queries



- **Solution to discovery problem**
- **Besides the GENESI-DR Web Portal there are Web Services offered**
 - OpenSearch based interface
 - XML describing the Search engine
 - Results presented in RSS/ATOM
 - Easy aggregation / integration
 - Adapted to scientific domain
 - Search for spatial or temporal extent, name of sensor, etc.



- **Cooperation**

- A.1 : Enable EGEE ES community access to GENESI-DR data.

- A.2 : Enable the EGEE ES community and GENESI-DR users to share applications across the two infrastructures



- **Discovery & Access with GENESI-DR**
 - Discover the data series through the central site
 - Extract the information service URL of the data repository storing the data series
 - Query the information service of the DR more detailed
 - Retrieve information about available data sets, their URL and additional meta data.
 - Invoke a client that is compatible with the possible data sources in GENESI

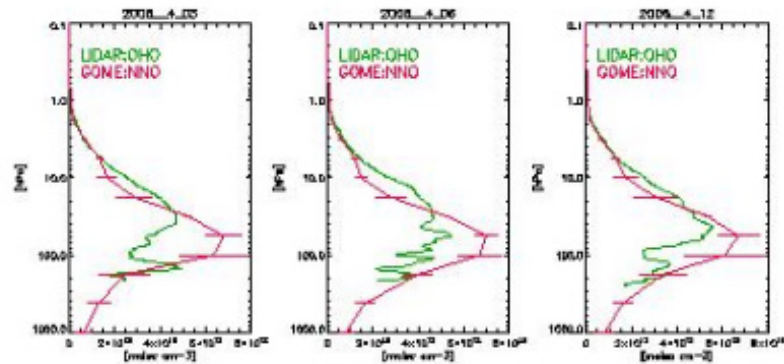
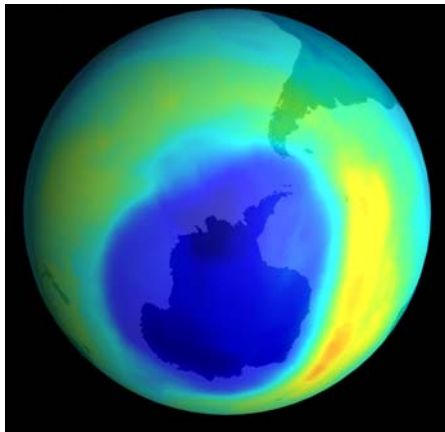


- GOME (Global Ozone Monitoring Experiment)
 - **Sensor carried by MetOp satellite**
 - **Measures (a.o.) the total column of atmospheric Ozone**
 - **Algorithms allow the scientists to estimate the vertical profiles of Ozone starting from the total column**

- LIDAR (Light detection and Ranging)
 - **Ground station using Laser beam**
 - **measure the vertical profile of Ozone**



- While satellites provide wide coverage, LIDAR stations can provide measures only in a narrow beam around them
 - spatial query in meta data
- Validate the Ozone profiles obtained processing GOME acquisitions with LIDAR measures



- **Using GENESI-DR it's straight forward**
- **Our example implementation**
 - Submit GOME processing job to Grid
 - Extract temporal and spatial extent from the input file
 - Query the GENESI-DR to get a corresponding LIDAR file
 - Access LIDAR file and compare profiles
- **Simple CLI was enough**
 - Implemented in Java using standard XML/RSS libraries and JavaGAT



- **Advantages**

- Simple Implementation
- Use existing XML & RSS components or even OpenSearch Implementations (see opensearch.org)
 - e.g. Perl/Python libraries, plugin for java ROME, Apache Abdera
 - Module for Drupal CMS to present online
- Backend independent
- Discovery and access can (theoretically) be done directly from WN
 - e.g. for spatially decomposing domain and assign to WN



- **Open questions**

- Not specified how to register data with catalogue ourselves (yet)
 - Meta data had to prepared
- Meta data might get cluttered
 - E.g. xml ns: rdf, opensearch, eop, dc, dct, dclite4g, foaf, ical
- Software not public (yet?) -> will it be maintained?
- Will more data repositories join?
- Too centralized? Replication mechanisms?
- A.2 (application sharing) w.i.p
- Complement it on Grid site with a cache in SEs / Replica handling?
- OWS Integration, e.g. WCS / CAT
- Integrate in Portals / Science Gateways for EGI



Questions?

