Grid technology and EO: experience at ESA

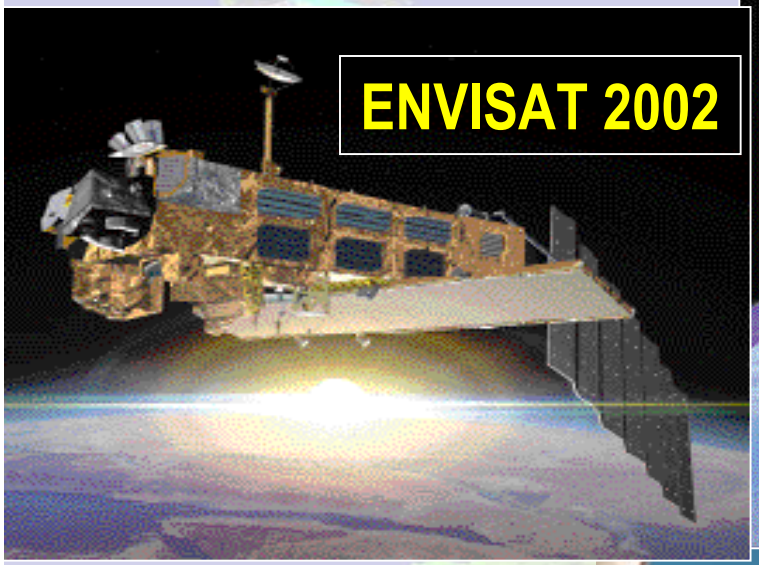
luigi.fusco@esa.int

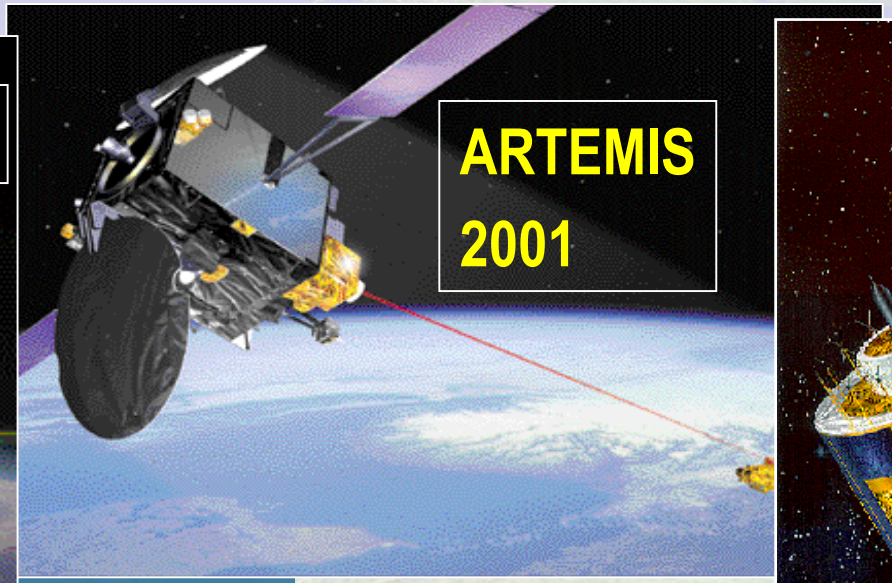
MAMA meeting - Roma, 3 June 2003

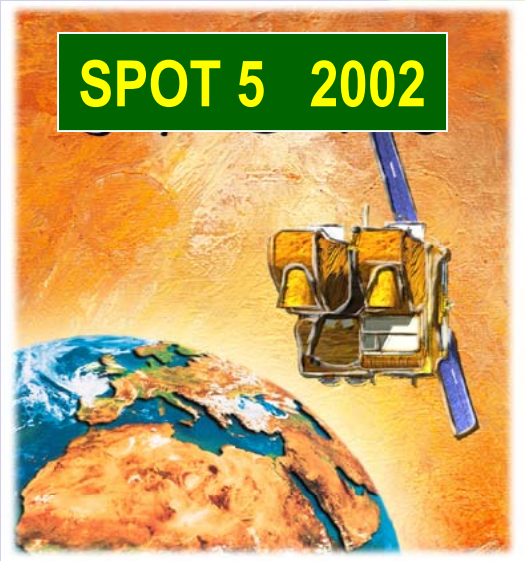
Summary

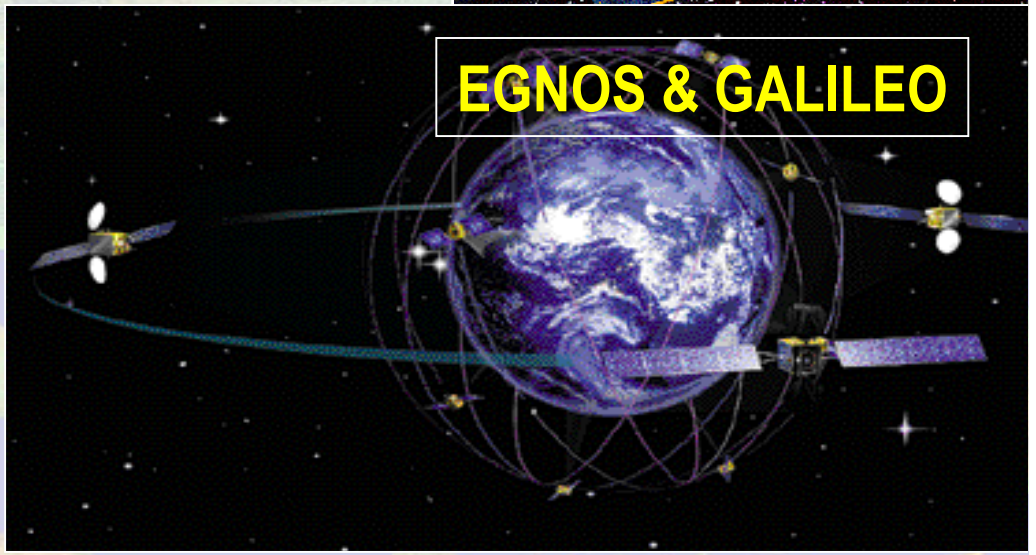
- EO recent facts, applications, technologies
- EDG Overview, Middleware and Status
- EDG services hosted at ESRIN, EO Applications services
- Demo

New European Space Applications Missions



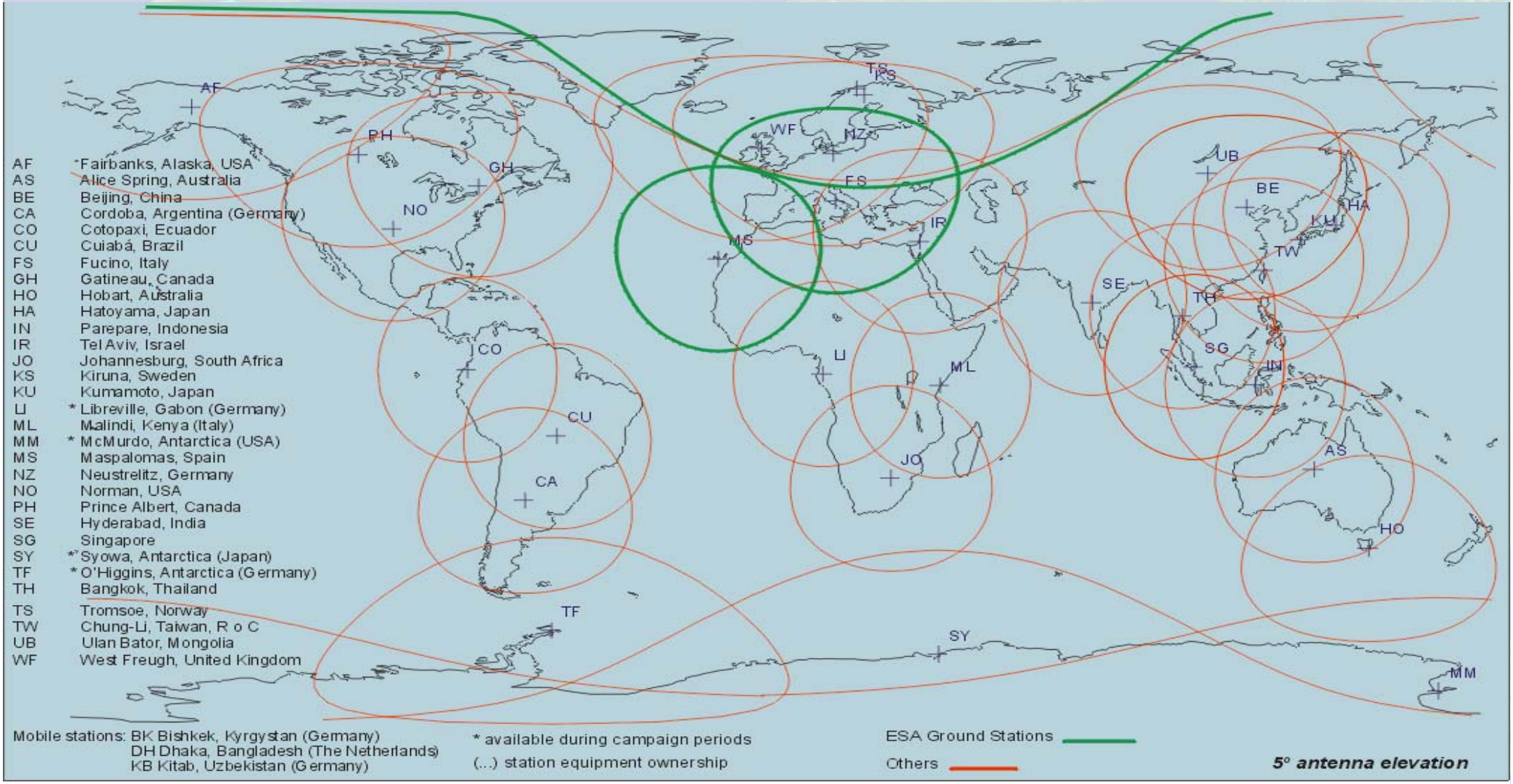




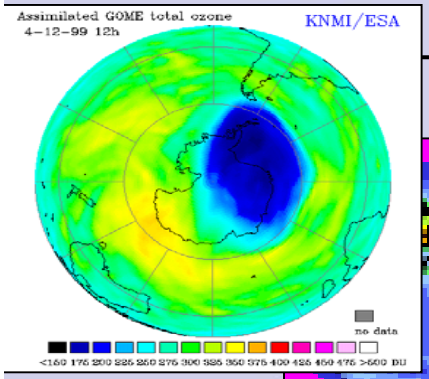


esa

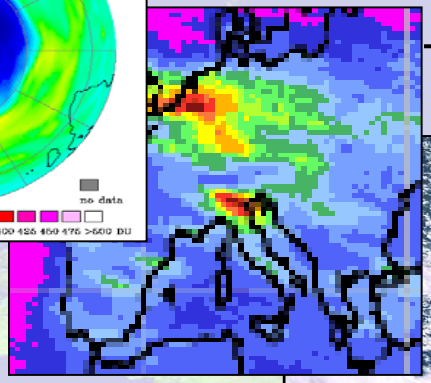
Major international EO Acquistion Facilities



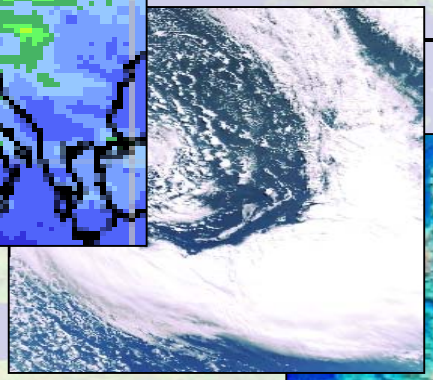
ENVISAT will provide very large data flows…



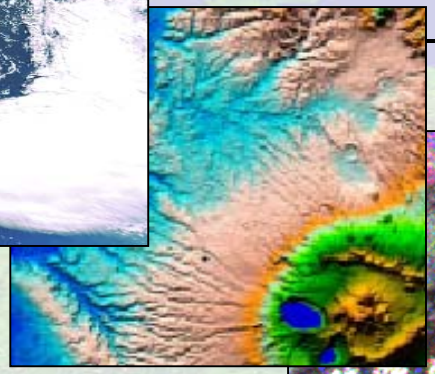
Altitude 0 to 100 km: GOMOS, MIPAS and SCIAMACHY are building a three-dimensional profile of ozone concentrations in the atmosphere.



Altitude 0 to 20 km: MIPAS and SCIAMACHY are detecting low levels of gases from industry, power generation and agriculture.

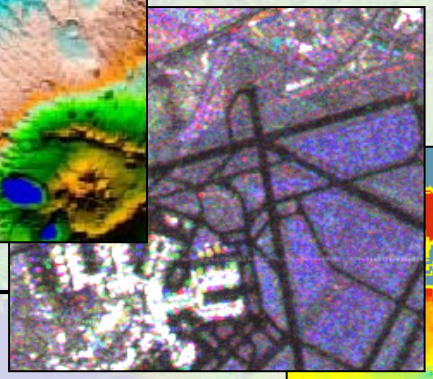


Altitude 0 to 10 km: MERIS obtains an image in which the clouds you see are but a part of a complex map of the concentration of water vapour.

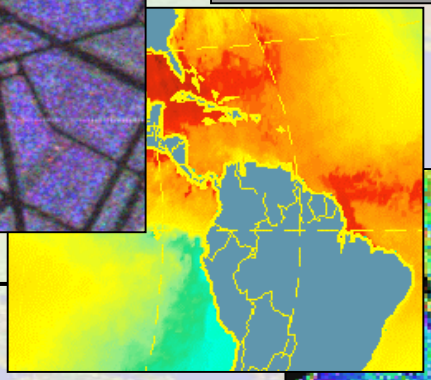


Altitude 0 to 4 km : ASAR and RA-2 create an accurate digital map of your surroundings, with height contours as accurate as 10 m.

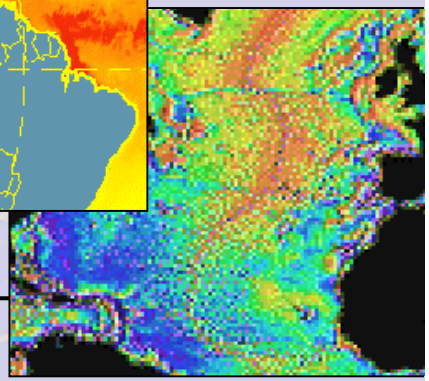
…400+ TB/year…



Ground level: ASAR, AATSR and MERIS map the vegetation and land use around you.



Sea level: AATSR measures sea surface temperature to 0.3 °C accuracy. MERIS precisely maps ocean colour, plankton and chlorophyll distributions. ASAR and RA-2 measure ocean currents, average wave-heights and wind velocities.



Underwater: RA-2 and DORIS combine to produce a detailed map of local gravitational strength, detecting the distribution of denser and less dense rock in the Earth crust beneath the oceans.

The Earth Observation Operational Community

- Largely distributed
- Each facility deals with multiple missions to serve distributed user science and commercial communities
- Strongly dependent on Information Technologies advancements (integration of what is "available")
- Often in place to meet institutional requirements (e.g. national, Civil Protections)

EO applications European plans

- Part of the ESA-European Commission Space Strategy
  - Integration of technologies, including Space Communication, Navigation and Earth Observation
  - Focus on GMES – Global Monitoring for Environment and Security
    - Serve institutional, operational and science communities
    - Dedicated funding by ESA and by EC
    - Ambition to develop European-wide "operational services"
  - Need to develop an European Earth Science Infrastructure

Identified priorities

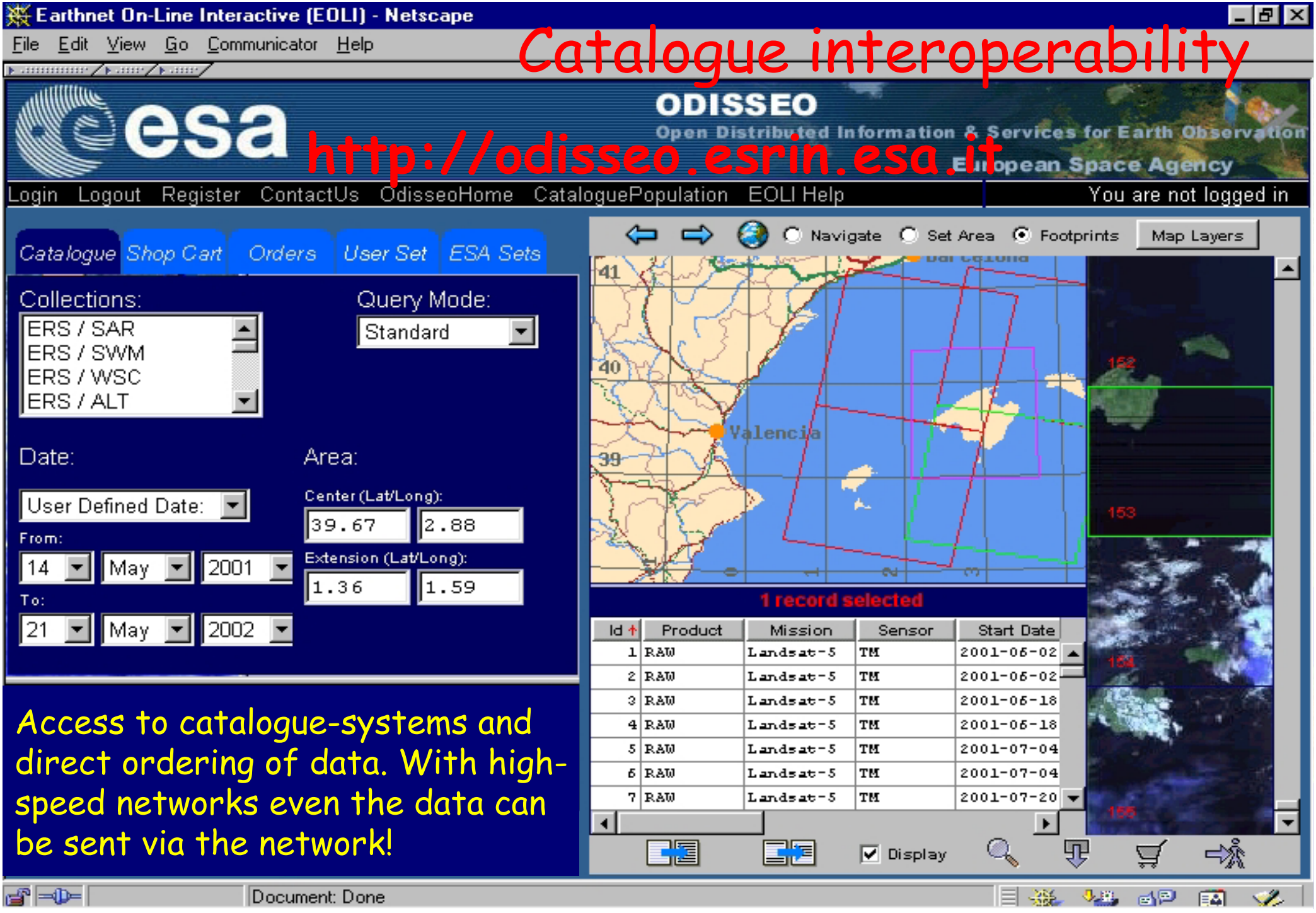
- Earth Science has requirements for emerging information technologies
- Beside the operational "Space" and "Ground" Segments the priority is in the development of a effective "User Segment", i.e.
  - Users needs to access distributed data, information and collaboration services
  - Dedicated Earth Science Research Infrastructure

The EO technology view:

emerging technologies of CEOS interests

- Access to metadata and data, interoperability (syntax and semantics)
- Data Archive Models
- Internet services, including web services, web mapping
- High speed connectivity
- Distributed Processing Systems (GRID) for the EO community
- …

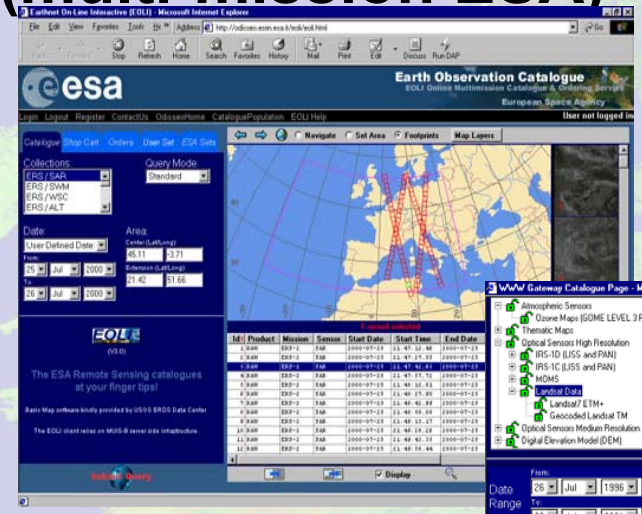
Catalogue interoperability

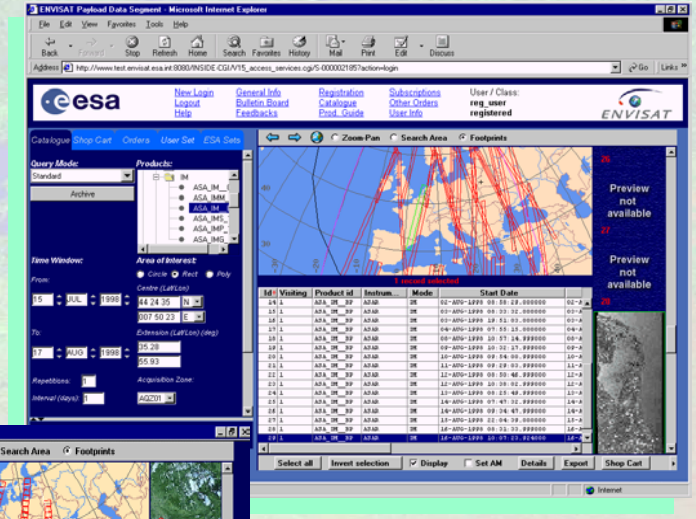


Access to catalogue-systems and direct ordering of data. With high-speed networks even the data can be sent via the network!

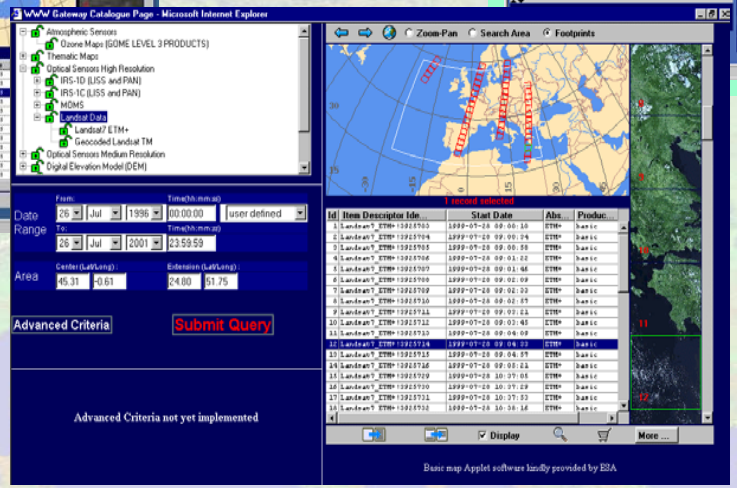
EO large databases and Interoperability across data servers

MUIS
(multi-mission ESA)





ENVISAT



INFEO
(international) – access to NASA, DLR, CCRS, SPOT, …

EO catalogue interoperability facts

- CEOS standards are in the ISO process
- Various services identified in underlined data model (directory, inventory, browse, …)
- Spatial Data Initiatives gets now political interest in Europe:
  - INSPIRE initiative supported by EC DG/ENV aims at defined European wide standards for environmental parameters

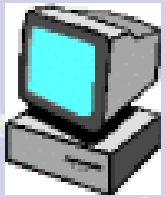
Web Mapping services

OpenGIS Consortium has developed WEB MAP SERVICE specifications

- International Consortium of 220 companies, governmental agencies and universities
- Consensus building for geographic information public access interfaces
- Interoperable solutions for geographic services in internet and mobile communications
- consider limitations on data volumes and available bandwidth

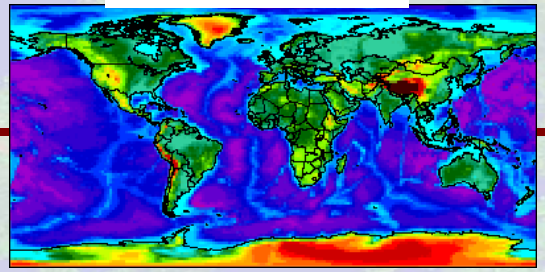
➔ EO data access / visualisation

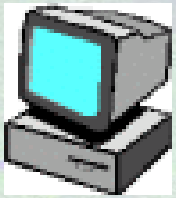
Web Mapping services



USER

URL





Webmap Service

Environmental Data

Location Based Services

Satellite Images

Catalogue Systems

Geographical Data

GAF - IONIC Geographical Application Framework - Netscape

File Edit View Go

Bookmarks

http://mapserv2.esrin.esa.it/map/wtf

esa Web map server

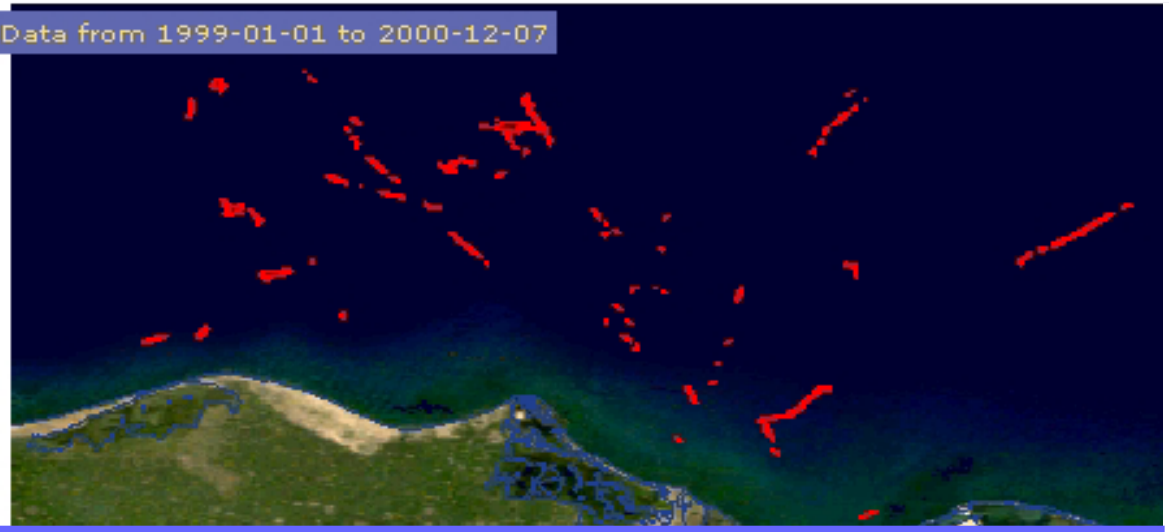
European Space Agency

ABOUT IONIC ESRIN

Search by

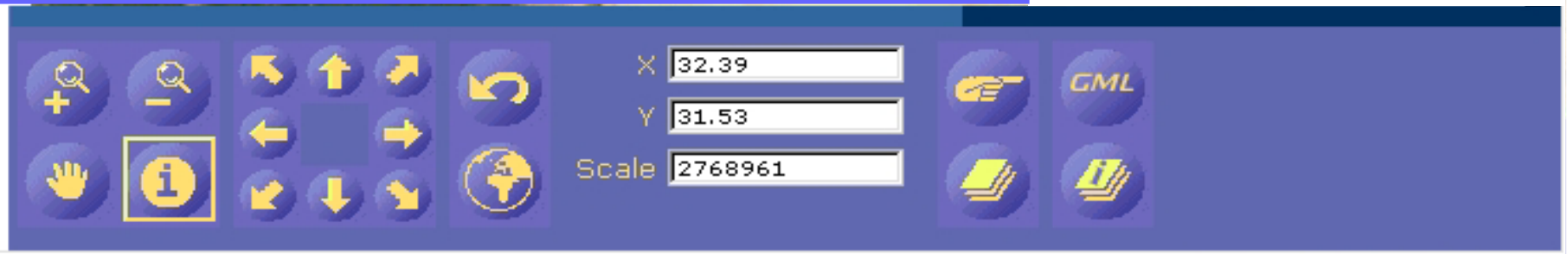
Country

Date



Oil spill monitoring near the Suez canal (ESA).
On the background 1x1 km² MODIS (NASA).

The ESA Web map server offers access to multiple globally distributed databases (e.g., NASA and ESA). The need for fast access to large volumes of data requires usage of high-speed networks..



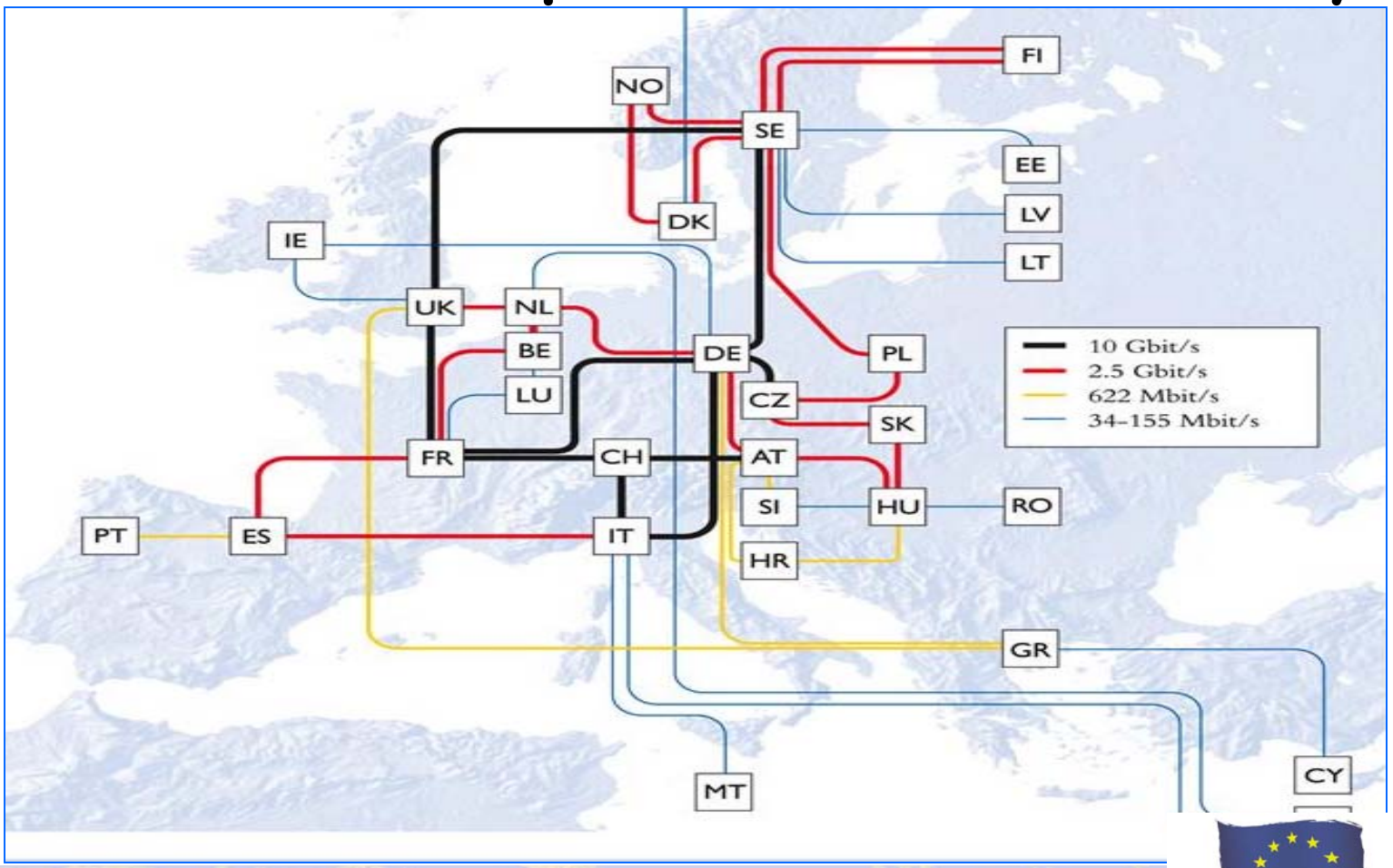
Document: Done

Summary

- EO recent facts, applications, technologies
- EDG Overview, Middleware and Status
- EDG services hosted at ESRIN, EO Applications services
- Demo

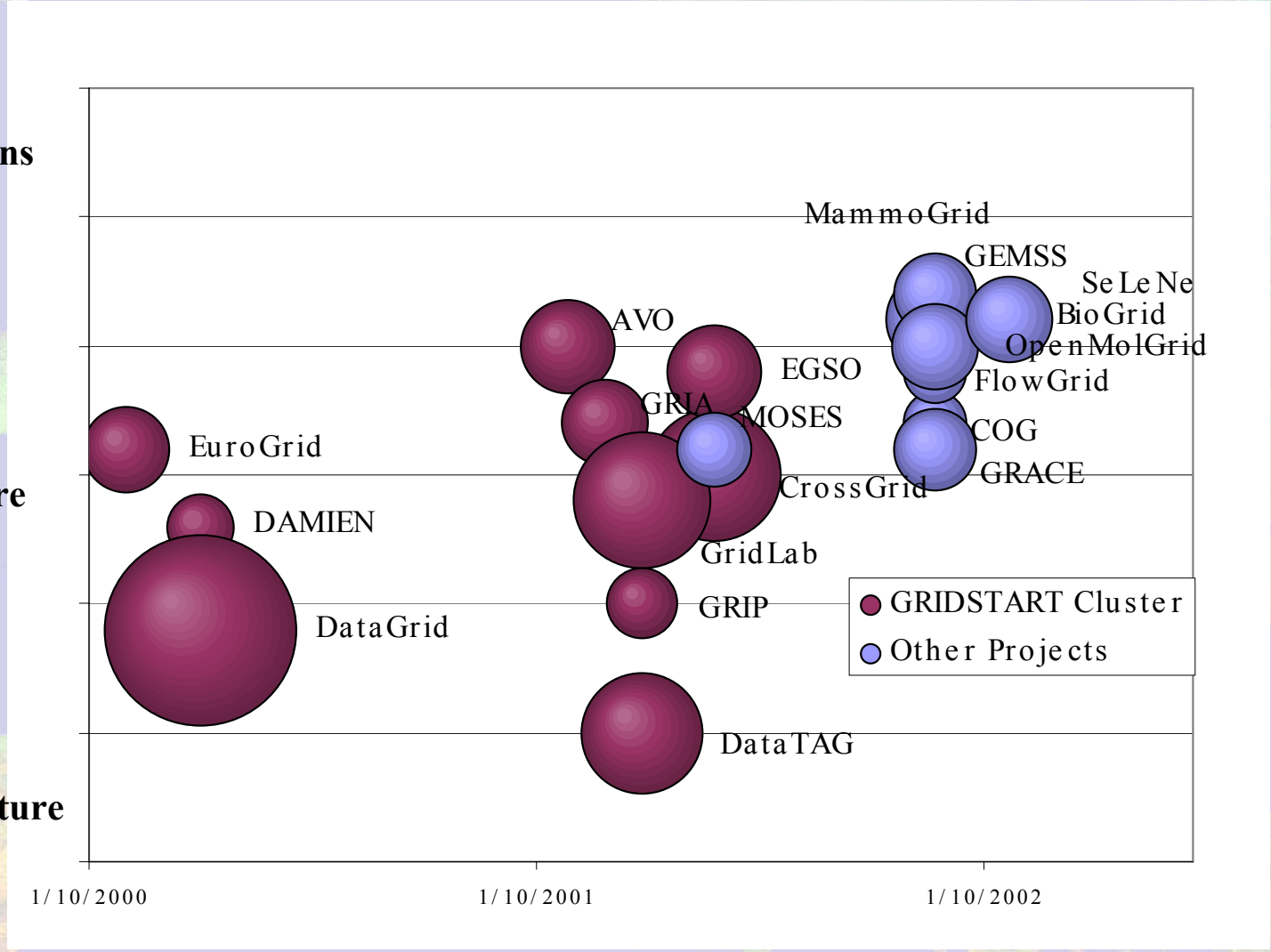
GEANT connectivity in Europe

- Not yet used by the EO community
- Across Atlantic testbed in place for HEP community



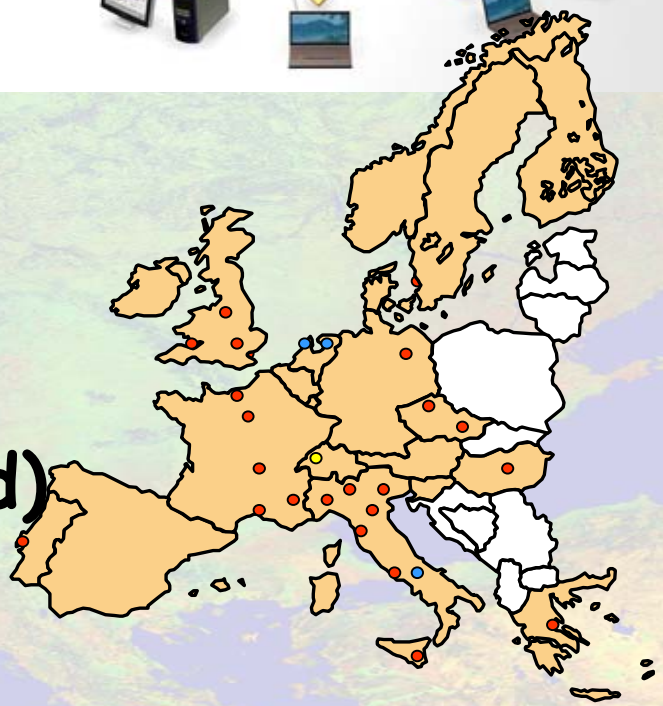
European Commission

IST FP5 Grid Projects



European DataGRID - EDG

- Project funded by the EU
- Enable the access to geographically distributed computing power and storage facilities belonging to different institutions
- Led by CERN together with 5 main partners (+15 associated)
- Active Industrial Forum



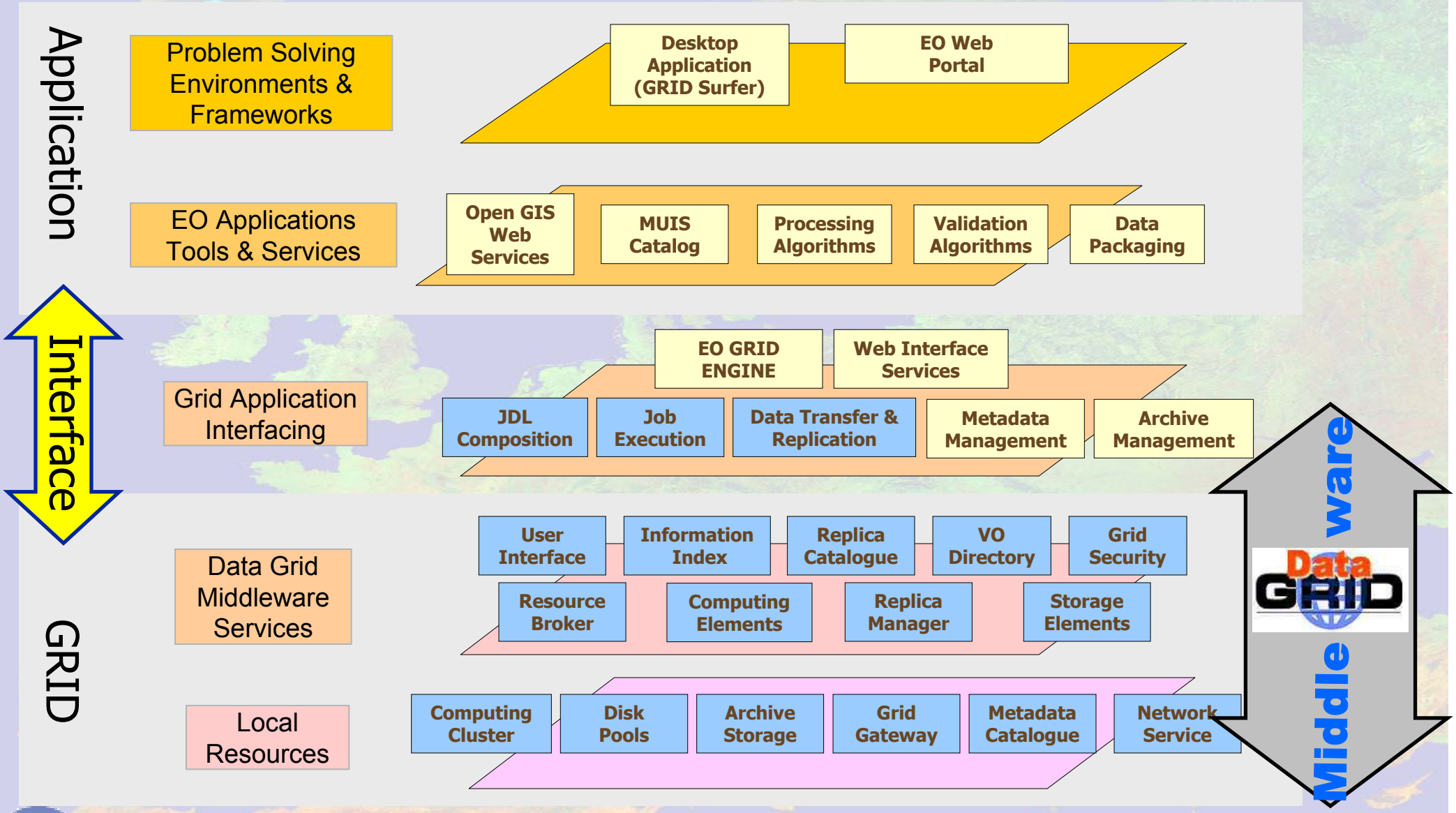
European DataGRID - EDG

- Provides production quality testbeds
- Demonstrates the possibility of building very large clusters of distributed resources out of low-cost computing commodities
- Three real data intensive computing applications areas are covered by the project
  - High Energy Physics (HEP), led by CERN (Switzerland),
  - Biology and Medical Image processing, led by CNRS (France)
  - Earth Observations (EO), led by the ESA/ESRIN (Italy)

EDG System Overview

- **Certificates, Users, VOs** : CAs recognised in 14 countries; ~300 users in 9 VOs
- **Middleware:** Workload Management, Data Management, Information & Monitoring, Fabric Management, Mass Storage Interfacing, Network Monitoring
- **Integration:** EDG central code repository; installation & testing on development testbed before release to Applications
- **Production Testbed (shared resources in 7 EU countries):** Resource Broker (CNRS-Lyon), Information System (RAL-UK), 47 CE, 17 SE, some 2000 worker nodes

EDG EO Architecture Layers



EDG Middleware -1

- **Information System:**
  - MDS widescale deployment: limited scalability of Globus solution;
  - EDG relational model (R-GMA) to replace;
  - Uniform GLUE schema to be adopted
- **Resource Broker / Job Submission System:**
  - builds on GT2 and Condor;
  - scalability/reliability issues (GASS-cache) ;
  - major re-design for release 2
- **Replica Catalogue:**
  - scalability, reliability issues in 1st Release;
  - new RLS architecture in release 2;
  - EDG/Globus joint development

EDG Middleware -2

- **Computing Element:**
  - interface to Globus Gatekeeper; staging of input/output sandbox;
  - logging & book-keeping : no major problems
- **Storage Element / Mass Storage:**
  - developed specific solution for CERN; needs to be analysed for compatibility / portability to other environments
- **Fabric Management:**
  - LCFG "New Generation" adopted from Edinburgh Univ. Configuration of EDG releases, automated installation & configuration independent of internal sites setup

EDG Status – EO use

- Release 1.1:
  - Delivered Oct 2001
  - EO application evaluation in D9.6 Grid Scaling Study (→ summary version passed to CEOS GRID TT)
- Release 1.4: - current version
  - Delivered Dec 02 -Jan 03
  - 1-year GOME dataset processed Feb 2003
  - EO Evaluation report D9.3 (Mar 2003)
- Release 2.0: Due July 2003
  - Final assessment & report due December 2003

Summary

- EO recent facts, applications, technologies
- EDG Overview, Middleware and Status
- EDG services hosted at ESRIN, EO Applications services
- Demo

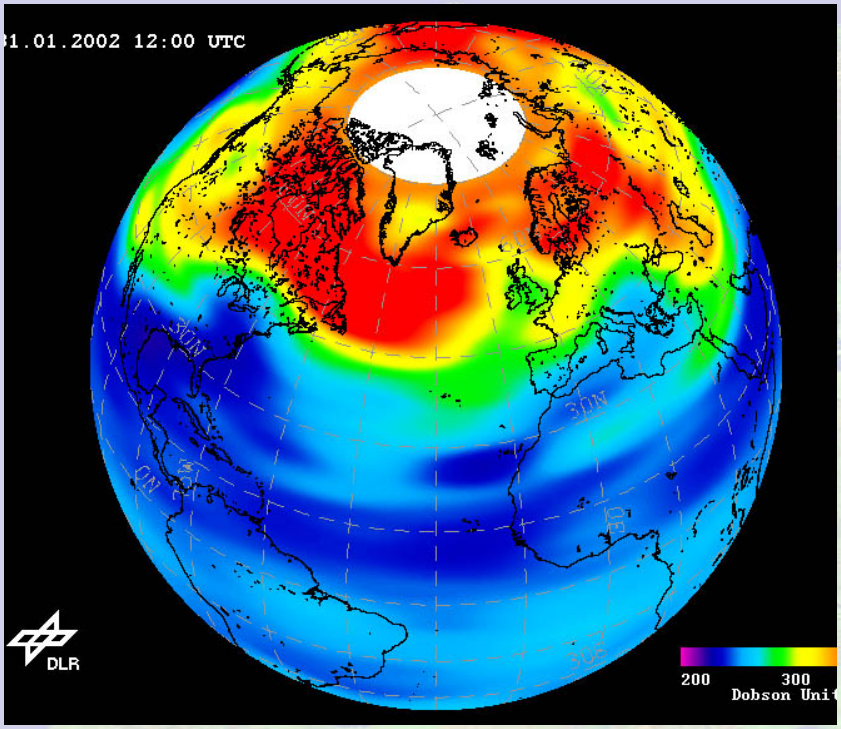
EDG services hosted at ESRIN

- User Interface to European Grid
  - logon to the grid via ssh to issue direct commands, or
  - interfaced to EO Grid Portal (GOME demonstrator)
- Computing Element (1)
  - grid0007.esrin.esa.int; 2 PBS batch queues; 30 CPUs
- Storage Element
  - grid0006.esrin.esa.int; 3.3 TB RAID
- Network Monitoring
  - 24h/7d performance monitoring
  - Present at 8Mbps, being upgraded to 34Mbps

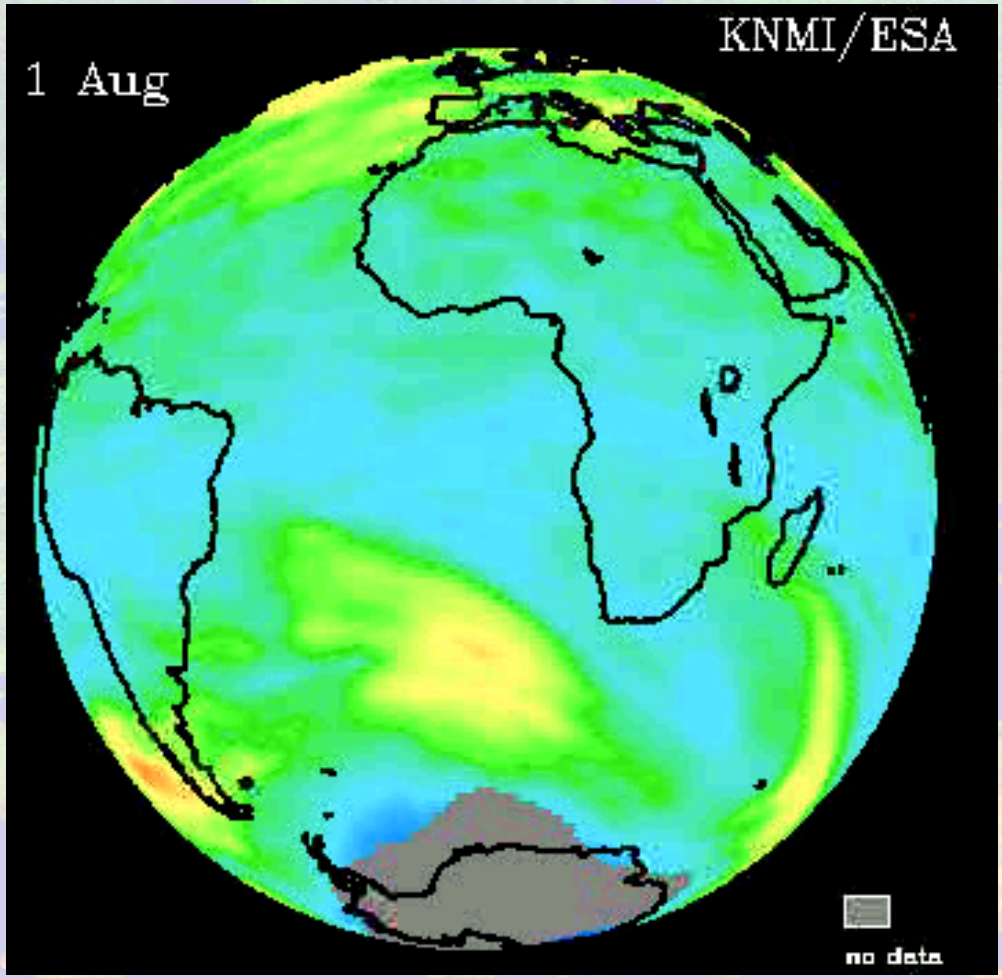
Extended services at ESRIN

- Local GRID Computing Element
  - Based on GT2 interfces
- Computing Element (2) – Campus Grid
  - gateway to ENEA Grid (Italian HPC network)
  - gigabit link operational
  - interfacing EDG with LSF/AFS (proprietary solution - work in progress)
  - Extension to other "Rome sites" – CNR, Univ Roma 2
- Integration with non-Grid systems
  - MUIS catalogue
  - AMS archive

OZONE: a case of Global Environmental Monitoring

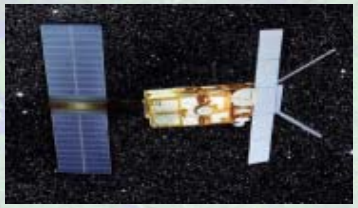


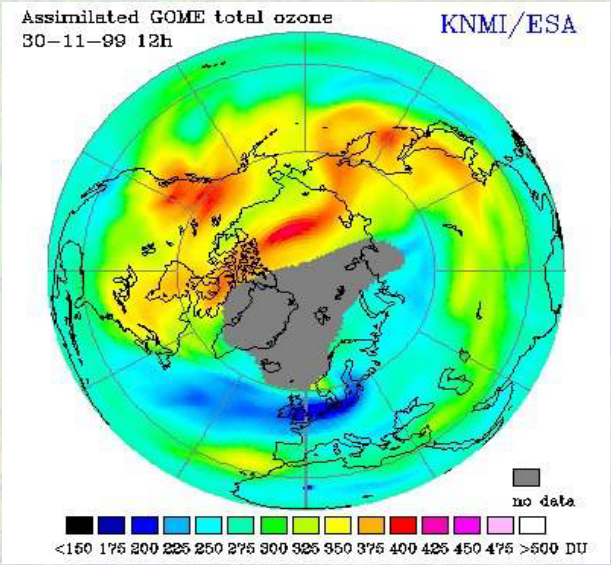
GOME analysis detected ozone thinning over Europe
31 Jan 2002



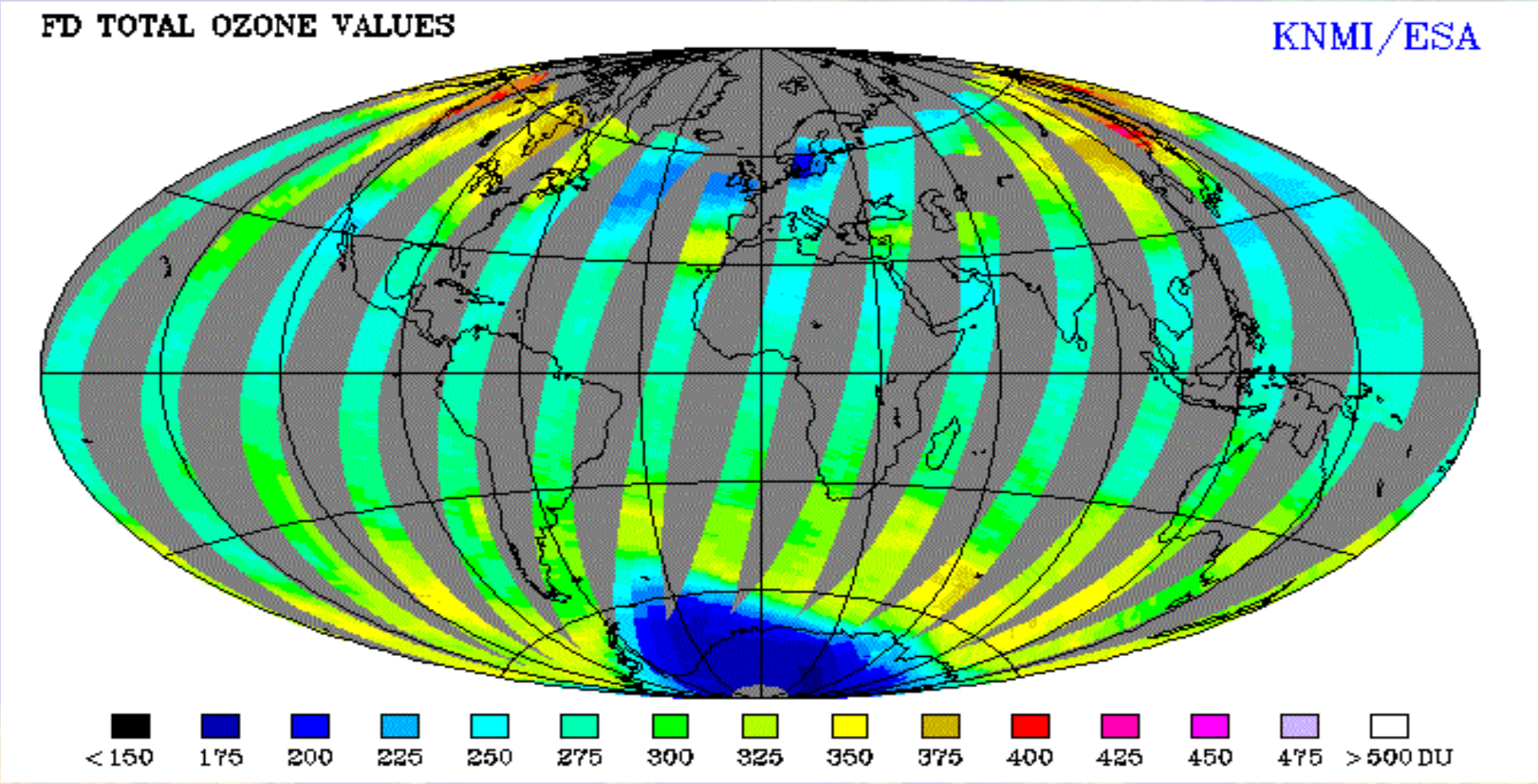
Ozone Application

- Wave spectra data measured by the GOME instrument on the ERS (level 1)
- Calculation of satellite ozone profiles (level 2 data)
- Two algorithms: OPERA (KNMI - modeling) and NOPREGO (Neural Networks)
- Data validation using ground based LIDAR measurements
- Collaboration among different institutes: France (IPSL), Italy (ESA, ENEA, UTV), Holland (KNMI)





GOME Instrument (1 day coverage)



Example of GOME level 2 output product

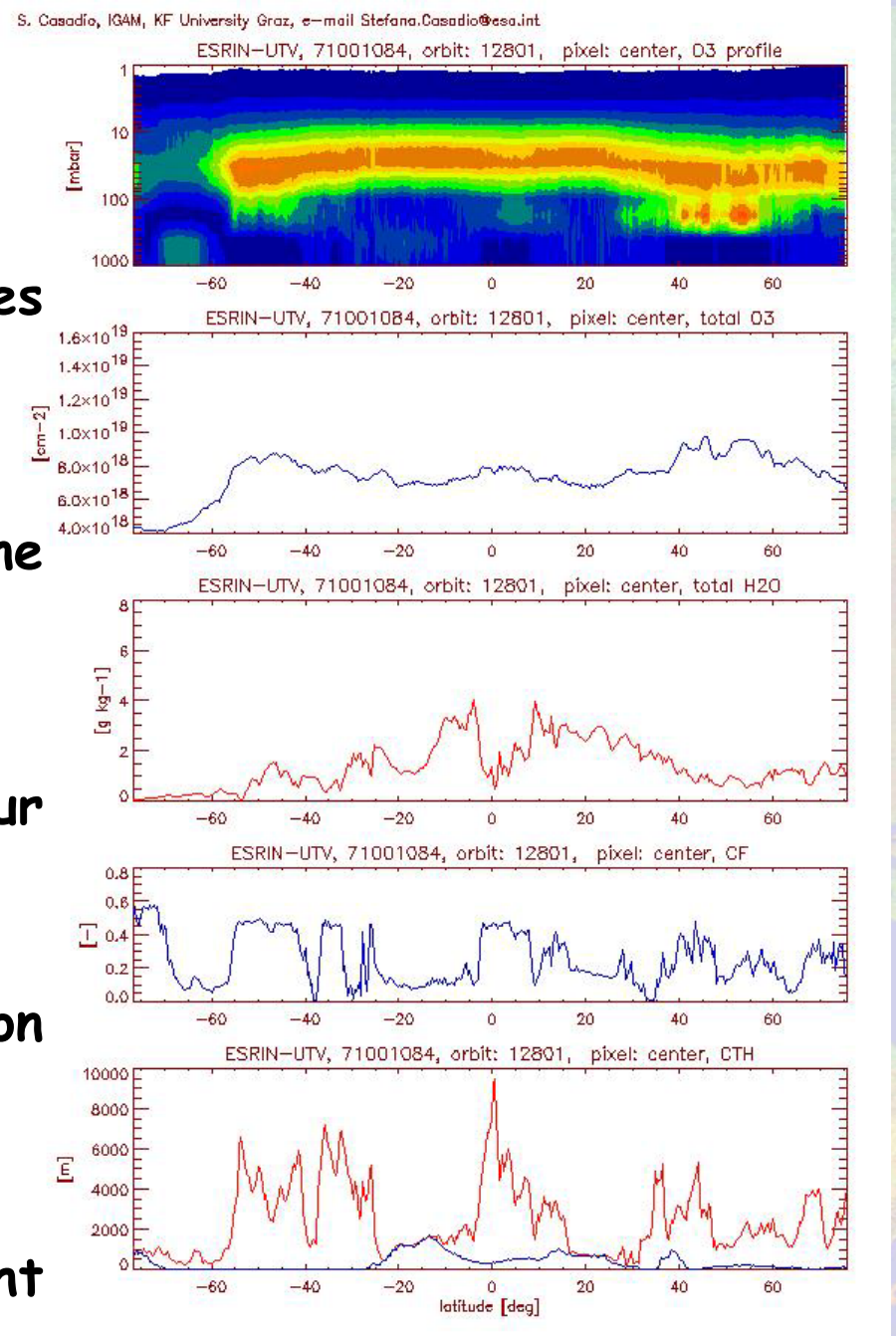
Ozone profiles

Total Ozone

Total Water Vapour

Cloud Fraction

Cloud Top Height



Validation application

- GOME Ozone profiles produced by OPERA and NNO are validated by IPSL using ground-based LIDAR measurements at Haute Provence observatory (OHP), France.
- The validation algorithm performs statistical analysis of the bias between GOME and LIDAR data at different altitudes
- The satellite observations which coincide geograpically and temporally with the ground-based LIDAR observations have to be extracted from the global GOME dataset
- A grid-enabled metadata catalogue (Spitfire) is used to locate the coincident data files

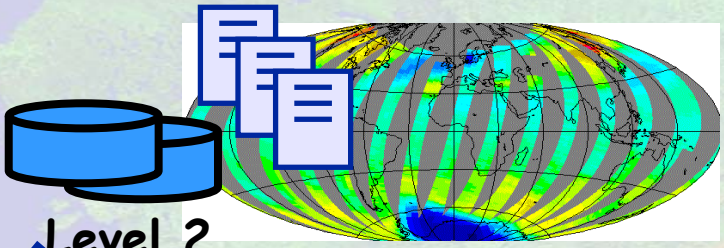
Computation of Ozone Profiles

- OPERA (KNMI - modeling) products takes 20x real time
- NOPREGO - Neural Network approach for level 2 products
  - GOME Input data (level 1)
  - Information used in neural network experiment (28 inputs)
    - Solar and Earth shine spectral radiances
    - Geometric info (solar zenith angle, line of sight)
    - No radiance calibration required
  - This is not an "approved" standard product!

Earth Observation Challenge

ESA(IT)
Raw satellite data - GOME
(~75 GB/y - ~5000 orbits/y)

Level 1



Level 2

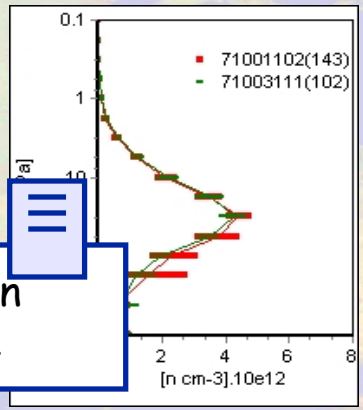
ESA(IT) – KNMI(NL)
Raw GOME data to ozone profiles
2 alternative algorithms
~28000 profiles/day

http://giserver.esrin.esa.int/grid-demo

IPSL (FR)
LIDAR data (7 stations,
2.5MB /month)

IPSL(FR)
Validate ozone profiles (~10^6/y)
coincident in space and time with
ground-based measurements

Visualization
& Analyze



Other EO Applications @ *ESRIN*

- EDG EO applications (ongoing and planned)
  - Re-processing 7 years of GOME data (1995-2002) on going
  - GREASE project (OMI simulation by Dutch Space)
  - GOMOS level 2 processing and validation
- Other actions under evaluation (not part of the EDG plans):
  - Integration in operational EO Ground Segment environment (SpaceGrid)
  - Access and use of high-res data for: Urban monitoring, large scale mapping, …
  - Integration of Meteo and ocean models,…

EDG Lesson learnt

- Globus GT2 been stress tested to reveal limitations
- Wide scale deployments never before attempted (at least in Europe)
- International cooperation demonstrated on middleware complex system developments (HEP community leads)

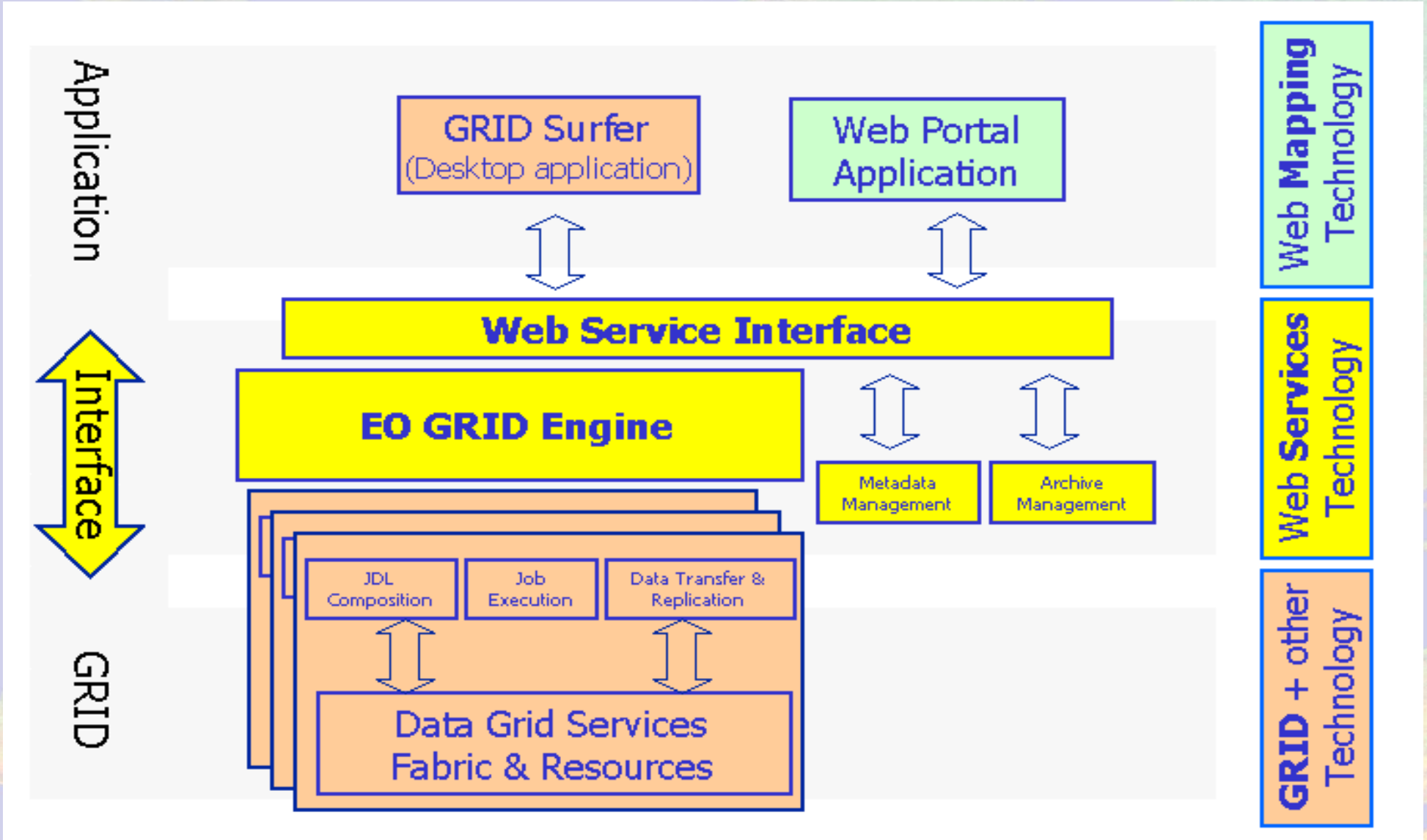
EDG Lesson learnt - 2

- Real application evaluation of all EDG features
  - Large number of files and distributed data volumes handling
  - Large number of simultaneous job execution
  - Multiple use of same data sets (simple and complex processing requirements)
- Integration of GRID with other technologies
  - Need to integrate "operational EO tools" with GRID environment
- Interaction of users from different locations / organizations
  - Access to distributed metadata and data sets (interoperability)
  - Use of Virtual Organization and security tools

Emerging requirements

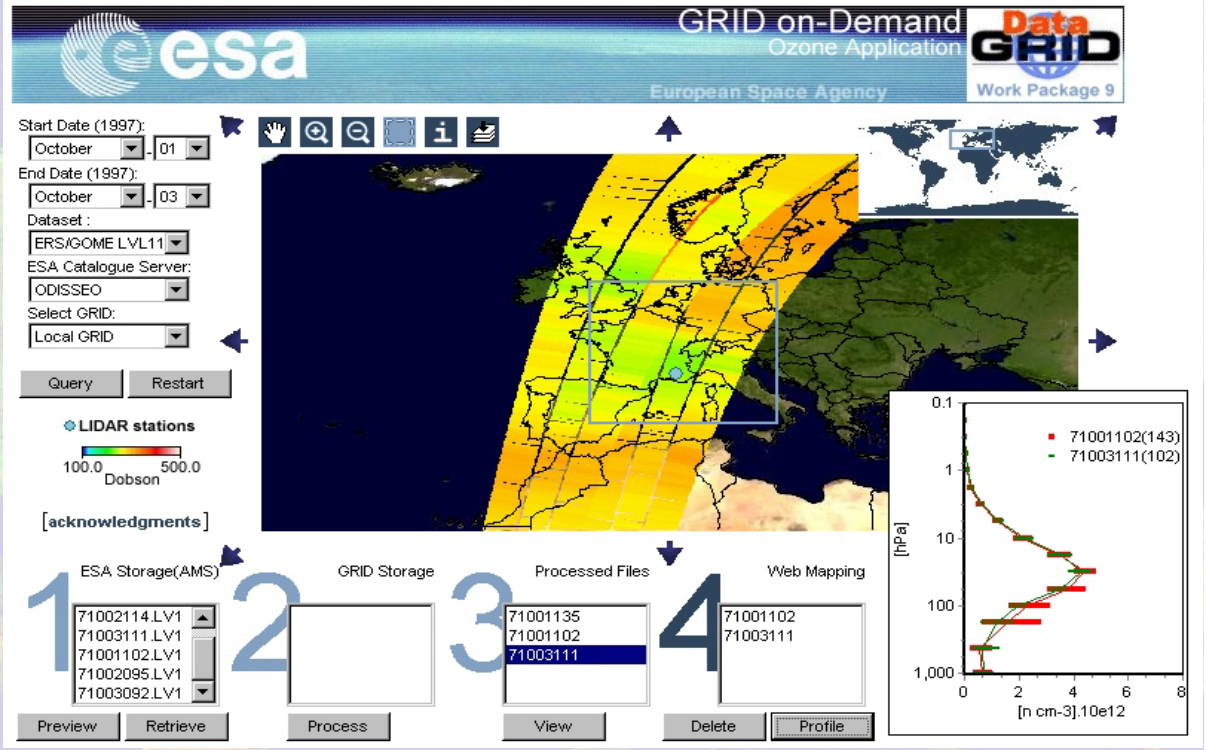
- Initial requirements detailed in EDG D9.1 and D9.6
- Key EDG implemented/demonstrated requirements
  - Store and retrieve files from different SE (outside or inside job scope)
  - Access distributed data via edg replica manager tools
  - Access metadata catalogue (spitfire)
  - Interface to web mapping via web services
- Key requirements not presently implemented in EDG
  - Workflow management
  - MPI (fast connections)
- Integration / interfaces with other GRID environments
  - as wished in CEOS demonstration (EDG and ESG-NASA interface)

Application and Grid Layers



GRID on Demand demo:
Ozone Application Portal

- Temporal and spatial selection of data
- Catalogue access and data transfer from ESA data warehouses to the GRID storage elements
- Job selection and status information
- Result retrieval and visualization in OWS
- Remote MySQL access (SOAP)
- Data validation w/ ground measurements



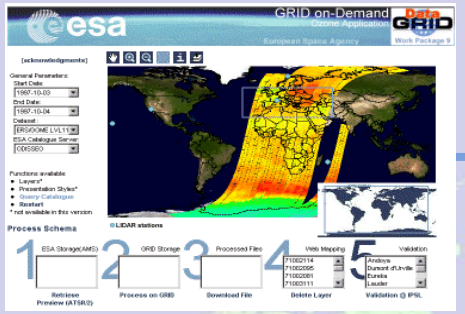
Conclusion

- The presented technologies have clear use for the ocean / marine applications

- Any interest for cooperation?

Thank you!

GRID on-Demand
Ozone profiles



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

GRID Portal

Queries L2 Catalogue to check if the data was already processed

If YES then queries Replica Catalogue to retrieve the path of logical file names.

Retrieves the Level 2 files from the storage element

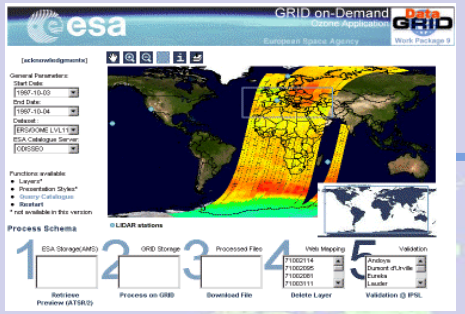
GRID

Replica Catalogue

Storage Elements

Computing Element

GRID on-Demand
Data not yet processed



Queries L1 Catalogue and retrieves orbit number and logical name file

Level 1 Catalogue

GRID Portal

Places Level 1 Data into the storage element

Launches the Job in the Computing Element

GRID

Storage Elements

Computing Element

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

Level 1 Storage (AMS)

When completed announces a new level 2 data into the catalogue

Level 2 Catalogue