WP9 Earth Observation applications

Final Project evaluation of EDG middleware, and summary of workpackage achievements

Luigi Fusco
luigi.fusco@esa.int

DataGrid is a project funded by the European Commission
under contract IST-2000-25182

Outline

- Objectives and Achievements
- Lessons learned
- Future & Exploitation
- Questions

WP9 Objectives

- Demonstrate how Grid infrastructure can respond to complexity / constraints imposed by EO applications
  - → Development of EO Grid application interfaces and tools
  - → Deployment of EO applications on the EDG framework
- Dissemination & promotion of this new technology
  - → To EO/space scientific teams, services providers and operational people
  - → Participation to outreach activities
  - → Publication of results obtained on Grid in scientific papers

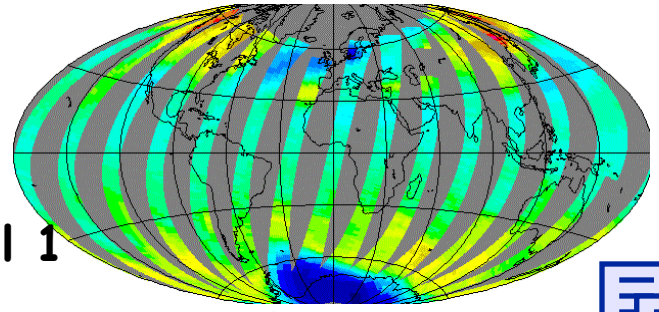
Main EO requirements for EDG

- Handling of large number and large volume of files from different satellite instruments and ground based measurements
  - → Processing with complex algorithms (e.g. Neural Network approach, Inversion approach)
  - → Creation and secure access to metadata catalogues and data
  - → Interface grid MW with the already developed operational satellite tools and infrastructure
- Tested using Ozone data processing and validation use-cases

The EDG Ozone processing and validation test-case

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

Level 1

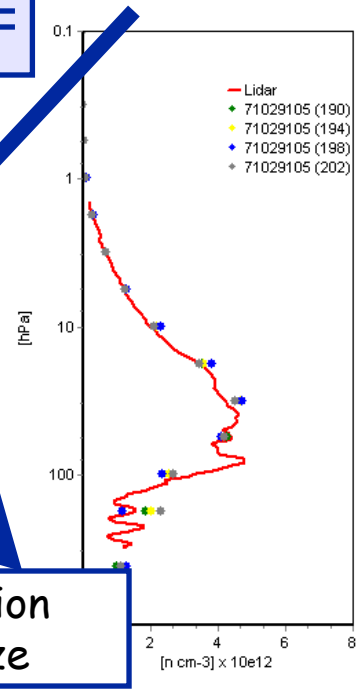


(example of 1 day total O_3)

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

Level 2

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

Achievements: Evaluation of EDG Testbeds-1

- Job management
  - → 1000's of job submissions - single jobs and simultaneous 'job storm' tests
  - → different job types - short/long duration, small/large datasets, with/out replica optimisation, MPI, etc.
  - → middleware has been tested to the limits of its capacity
- Replica & Data management
  - → ~10,000 entries registered in 1.4 RC, ~25,000 entries in 2.1 LRC
  - → over 35,000 data files transferred to EDG 'classic' and 'SRM' SEs
  - → EO products Metadata inserted in Spitfire & RMC catalogues
  - → Spitfire installed at IPSL and KNMI, RMC interface created at IPSL
  - → data migration testing for EDG upgrade from v1.4 to v2.0
- Information System
  - → Both MDS/LDAP and R-GMA used for application grid interfacing

Achievements: Evaluation of EDG Testbeds-2

- Fabric management
  - → LGFG, LCFG-NG fabric management tools used for EO site installations
  - → CE and SE at IPSL and at ESA-ESRIN
- Network management
  - → EDG Network monitoring tools installed & used at EO sites
- Security
  - → VOMS tested for AWG EO security use case
  - → VOMS tested in combination with Spitfire
- Tests were designed for maximum exploitation
  - → of EDG middleware features
  - → of available CE/SE resources

Achievements: EDG EO applications

- GOME L2 NNO processing (TV-ENEA-ESA)
  - → 7 years GOME (30k orbits, 500GB) processed in EDG 1.4, EDG 2.0 and in local grid;
  - → Integration of IDL (cots) licenses.
- GOME processing OPERA (KNMI)
  - → 7 years of GOME data loaded to the Grid, Use of application metadata DB (Spitfire and RMC) under Grid for large number of files
- GOME validation (IPSL) and validation portal (ESA) – 2 versions
  - → 7 years from 7 sites LIDAR profiles (MDSC), use of application metadata (Spitfire and RMC). Validation of Opera and NNO
- GOMOS reprocessing and validation (ESA-IPSL)
  - → 5 months data in ESA local grid. 40k files, 120GB
- GREASE: OMI simulation (Dutch Space-ESA)
  - → 1 month OMI instrument simulation (including development of Workflow Management System in GRID environment)
- CEOS-GRID (ESA-NASA-DS) - ongoing

Achievements: EDG & WP9

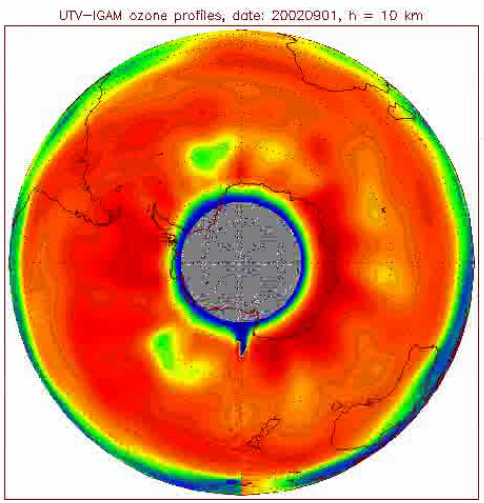
- Work closely with the MW developers and contribute to the working groups
  - → Contribution to Application Working Group (AWG): common usecase document (WP8, 9, 10), special usecases on security, metadata handling.
  - → Contribution to the Architectural Task Force (ATF)
  - → Contribution to the Quality Assurance Group (QAG)
- Full integration and participation of WP9 in EDG
  - → Interaction with WP2 on metadata handling and data replication
  - → Combined WP2/WP9 paper submitted for publication
  - → Test reports: D9.3 and D9.5, providing feedback to the MW developers
- Deployed vs Technical Annex planned effort
  - → ESA: 169mm (+80%), KNMI: 27(+13%), IPSL: 58(+61%), …

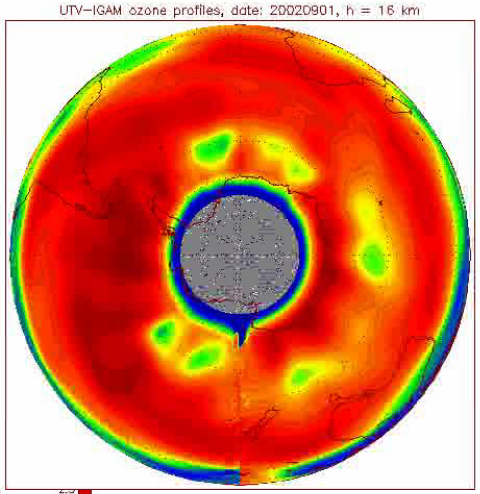
Achievements: EO VO and EO infrastructure

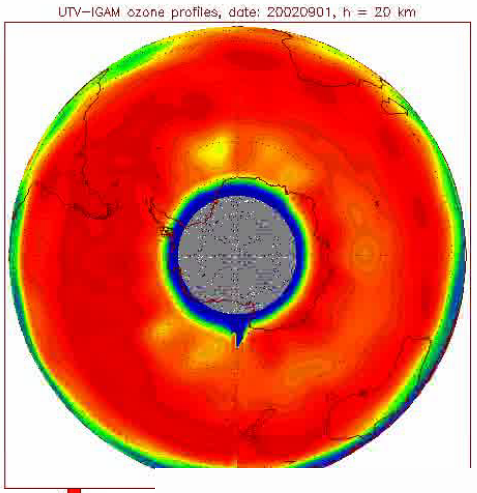
- Initial EO VO: ESA-ESRIN, KNMI, IPSL
- Present EO VO: extended to research, space related industry, International Space Agencies community
  - → At present some 25+ people, wish to extend it
- The present EO dedicated infrastructure (CE and SE): ESA-ESRIN, IPSL, connection to ENEA proprietary GRID infrastructure
- Upgrading infrastructure plans:
  - → extend it to ESA-ESTEC, Dutch Space, CNR, CEOS-NASA, new ESRIN CE+SE
  - → Mature elements considered for "operational" deployment

The latest GOME user results

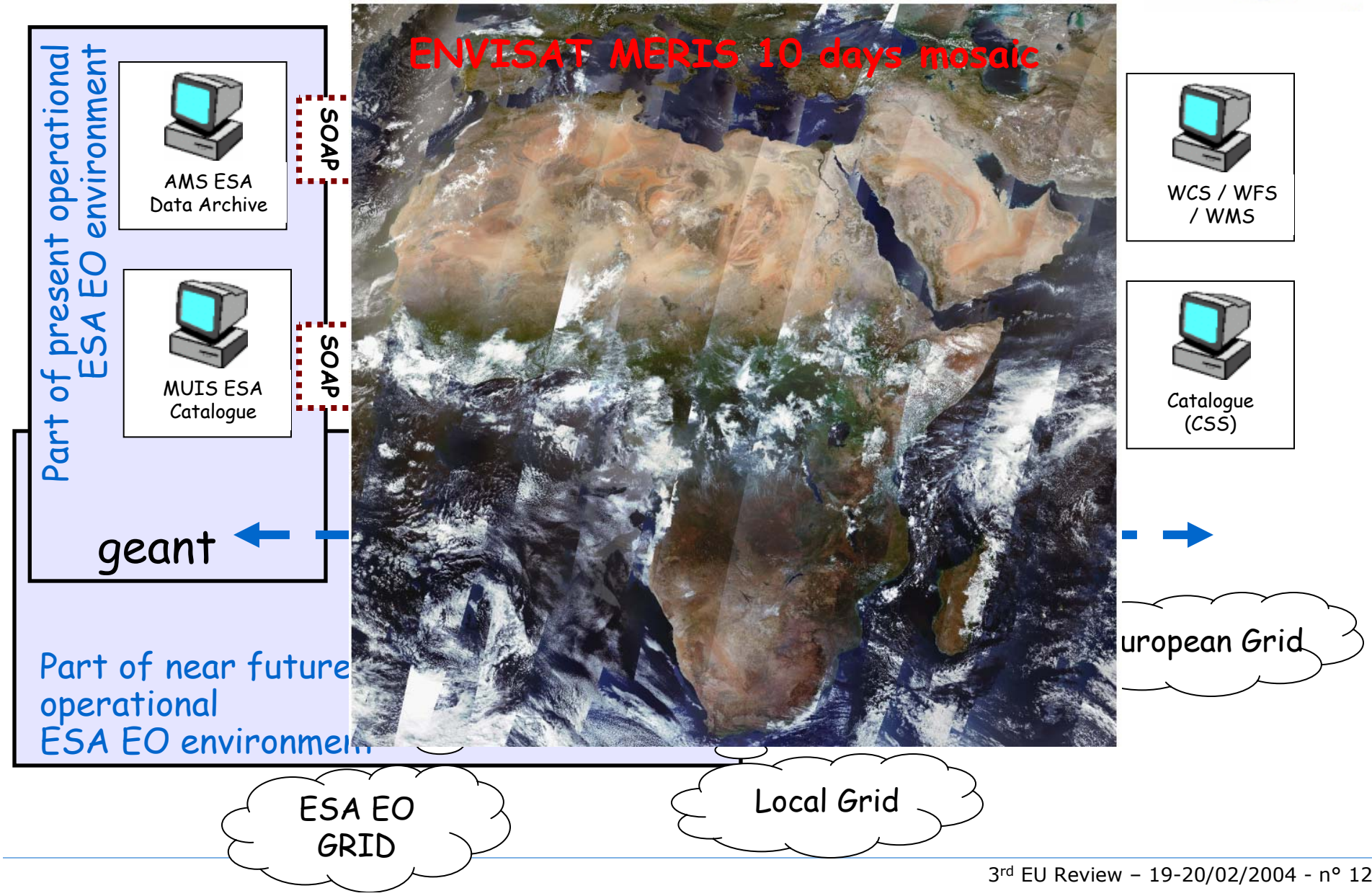
S. Casadio – ESA ESRIN
(GOME 3D Ozone volume over
Antarctica - Sept 02,
NNO Level 2 products generate in EDG)







Achievement: "GRID on Demand"



Lessons Learned

- Substantial first-hand experience and understanding of Grid technology (as users)
  - → Grid middleware encompassing wide range of standards
  - → Technology now being used for real world applications
- Impact of Grid technology on EO community and applications
  - → unprecedented experience for many of the communities involved
  - → New approach to scientific collaboration, communication and sharing among participants with distinct backgrounds
- Testbed evaluation
  - → Still need time, experience and major effort to convince all potential operational users

Future & Exploitation: WP9 committments

- Continuation of EO Virtual Organization
  - → New EO Applications proposed in EGEE
  - → NIKHEF and KNMI to maintain the EO-VO
  - → ISPL/CNRS could lead participation to EGEE
  - → ESA-ESRIN willness to provide data and infrastructure
- Continue widespread dissemination and promotion of Grid solutions in the EO community
  - → Involve new users and applications in EO community
  - → Continue to establish pre-operational Grid services

Which operational services?

1. Support to science users
   - Support science communities for focused collaborations, e.g. cal/val, global products, new algorithms
2. Support to application projects
   - → Provide reference application processing environment for generation of products
   - → Generation of periodic global and regional products for immediate availability at ESRIN
3. Support to Specific Reprocessing
   - → Allow multiple re-processing of same dataset
   - → Consider long term evolution of EO ground segment

EO GRID plans @ ESA-ESRIN

1. Recent facts - references

- **Frame work for ESA–EC near future activities**
  - ✓ **EC-ESA framework agreement** …; EC COM(2003)673 White Paper on "Space: a new European Frontier for an expanding Union. An action plan for implementing the European Space Policy"
- EC- ESA **Global Monitoring for Environment and Security**, Final Report for the GMES Initial Period (2001-2003):

"A key feature of the GMES information architecture is the need **to support collaboration between geographically dispersed GMES users and service providers**. **Collaboration** has to be **supported by an electronic infrastructure** enabling GMES users not only to communicate but also to access resources such as very large data collections or archived information, scientific experiments and computing power. **For the data- and computationally intensive areas of GMES, such as real-time modelling based on Earth observation data or climate modelling, high-performance networks and GRID-based computing are essential for mining, sharing and analysing data and visualising results**." ….

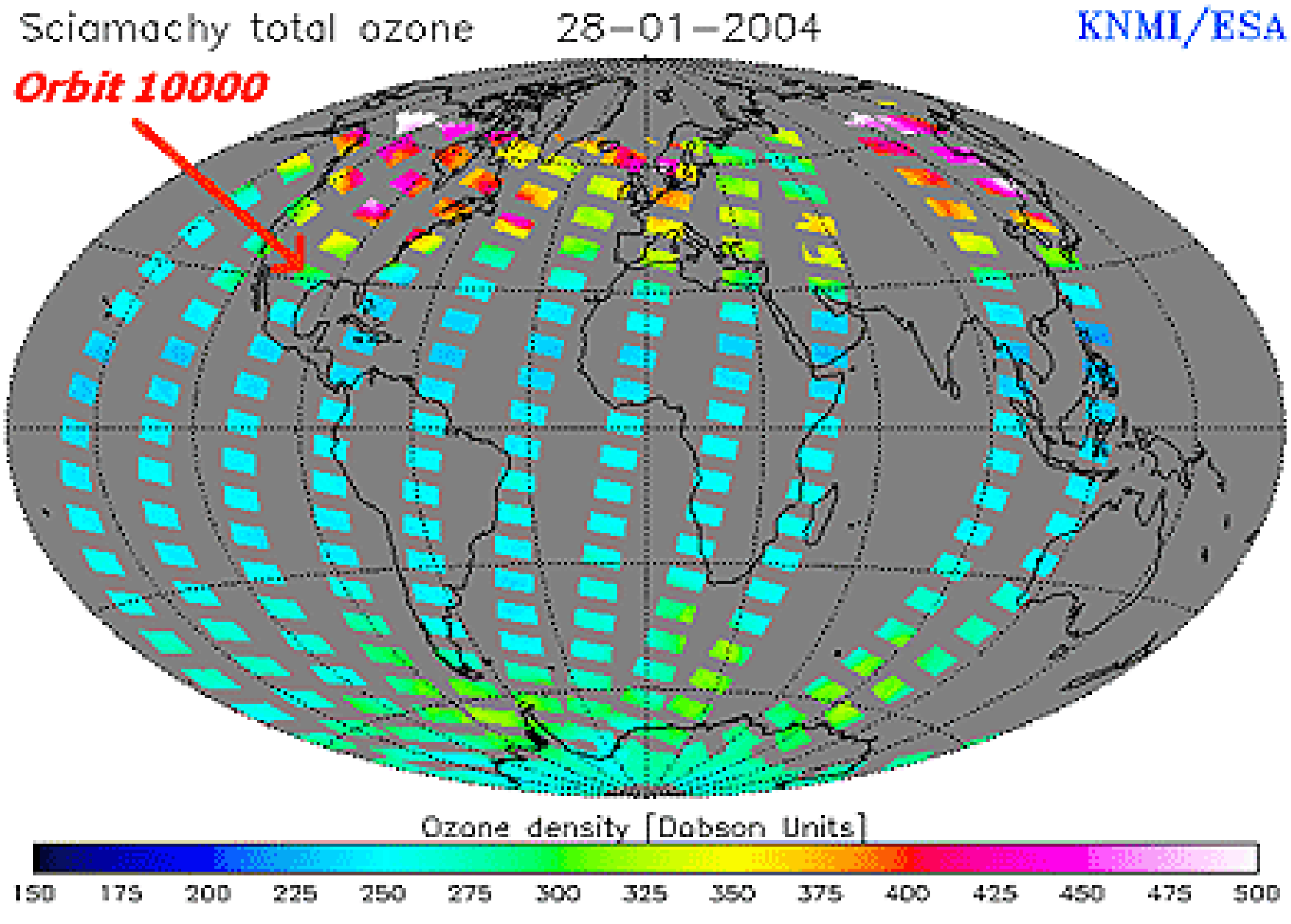
"The combination of an ESDI together with high-speed technology networks (**GRID&GEANT**), space and in-situ monitoring and data collection …"

- **ESA Agenda 2007** (ESA DG internal plan)
  - ✓ Technology innovation plans …; EO "Open & Operational" initiative

EO GRID plans @ ESA-ESRIN

2. New short term RTD projects
   - "The VOICE" – e-collaboration environment –funded by ESA (3+ new applications)
3. Preparation of "operational environment"
   - Integration of compatible and available ENVISAT software processing tools
   - Upgrading of Research Network Bandwidth, CE and SE
4. Preparation of ESA internal GRID technology plan
   - Consider internal and science operational requirements for coming 3-4 years

Important EO chang



Concluding comments

- Initial objectives are mainly reached.
  - → Although 10,000's of jobs, 35,000 files have been put on the grid ….. real mass production and validation of profiles not established due to instabilities
  - → Successfully deployed 7 applications on the testbeds
  - → Successful dissemination of Grid technology to other EO groups
- Full EDG collaborations established
  - → ATF, QAG, AWG, cooperation MW groups
  - → Interaction between different application domains (HEP, Bio, EO)

Questions?