egee

ESR Database Access

K. Ronneberger,DKRZ, Germany
H. Schwichtenberg, SCAI, Germany
S. Kindermann, DKRZ, Germany
J. Kraus, SCAI, Germany
J. Biercamp, DKRZ, Germany

www.eu-egee.org

Information Society
and Media

- Data Requirements of ESR

- Example Climate workflow:
  - Access via Webservice-interface/Amga
  - Missing pieces
  - Future challenge

- Example Satellite Data:
  - Access via OGSA-DAI
  - Implementation
  - Evaluation

Enabling Grids for E-sciencE

- Metadata and data bases are commonly large data sets, handled by different teams. The RDBMS generally used are MySQL, PostgreSQL or Oracle

- Many databases already exist the aim is the implementation of an interface with EGEE or at least to access a copy of them.

- If new bases are created on EGEE they need to be accessible outside Grid.

- Some metadata and data are only accessible to authorized persons. Others available on web site have rules for publications (acknowledgement, co-author).

- Many queries concern matching in time and/or space, expressed in geographical coordinates.

Enabling Grids for E-sciencE
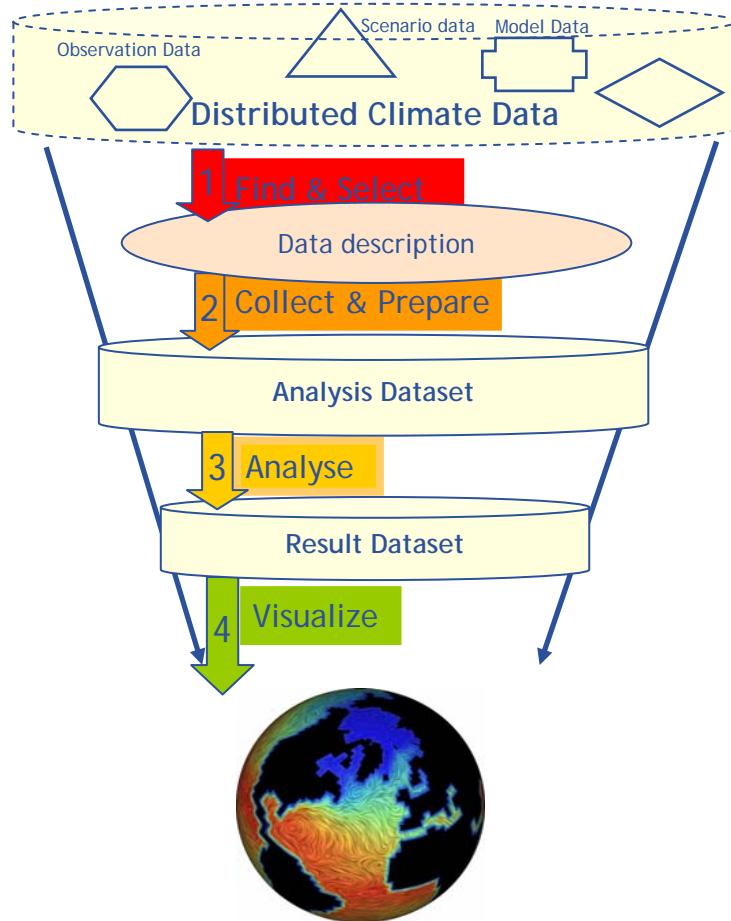
What is needed

- A central metadata catalog based on common and standardized metadata schema

- Uniform data access interfaces with transparent AA policies

eGee

Enabling Grids for E-sciencE
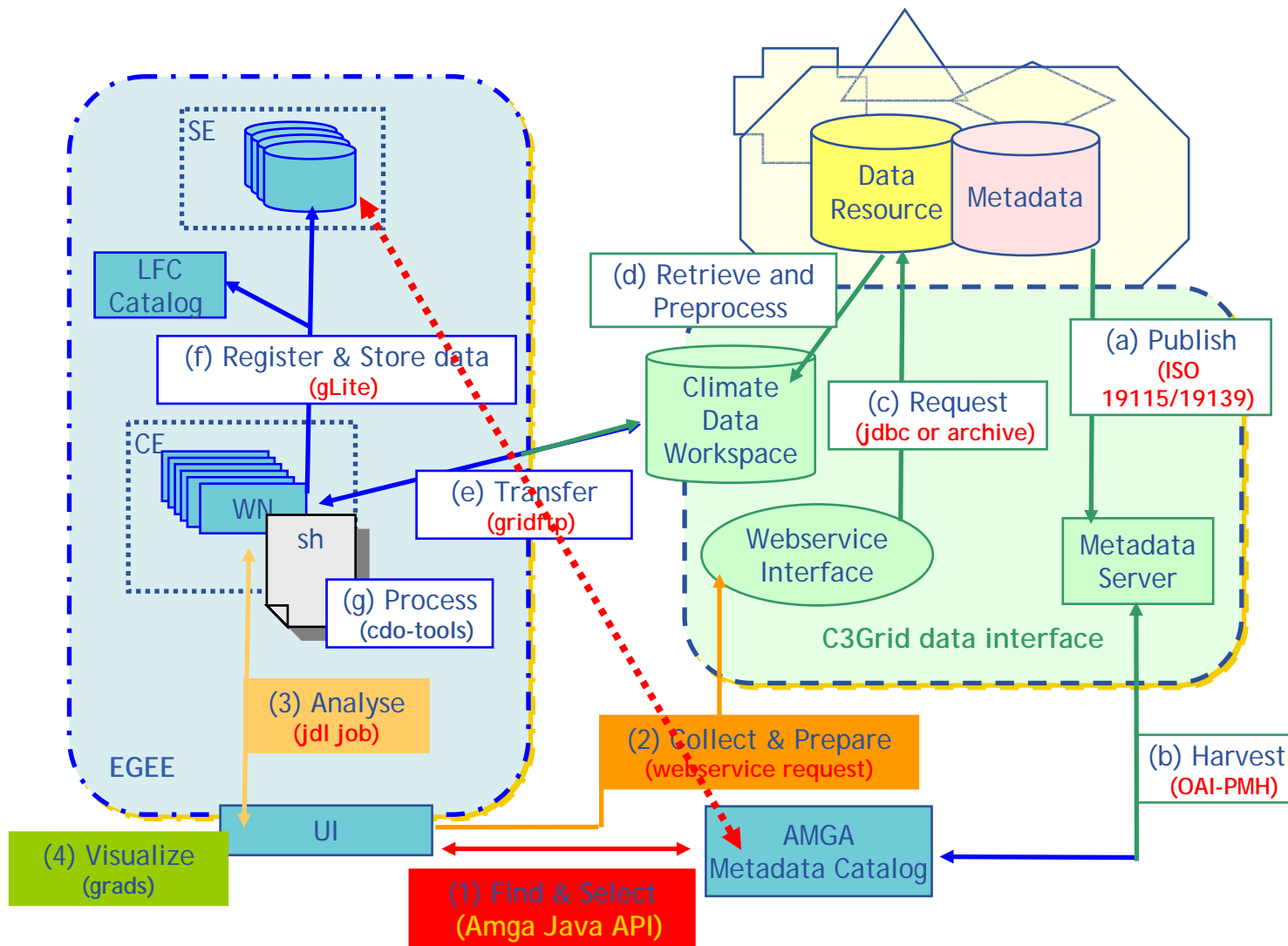
- Offering an alternative to current solutions for the daily workflows

- Additionally a common platform is provided to share data, tools and resources, supporting collaboration

- The common metadata scheme, based on international standards can be adapted/extended
  - by other disciplines
  - by International partners (discussion with NDG (GB) and ESG (USA) are ongoing)

eGee

Enabling Grids for E-sciencE

- Registering of uploaded and processed files in Amga

- Grid-enabling the remaining data

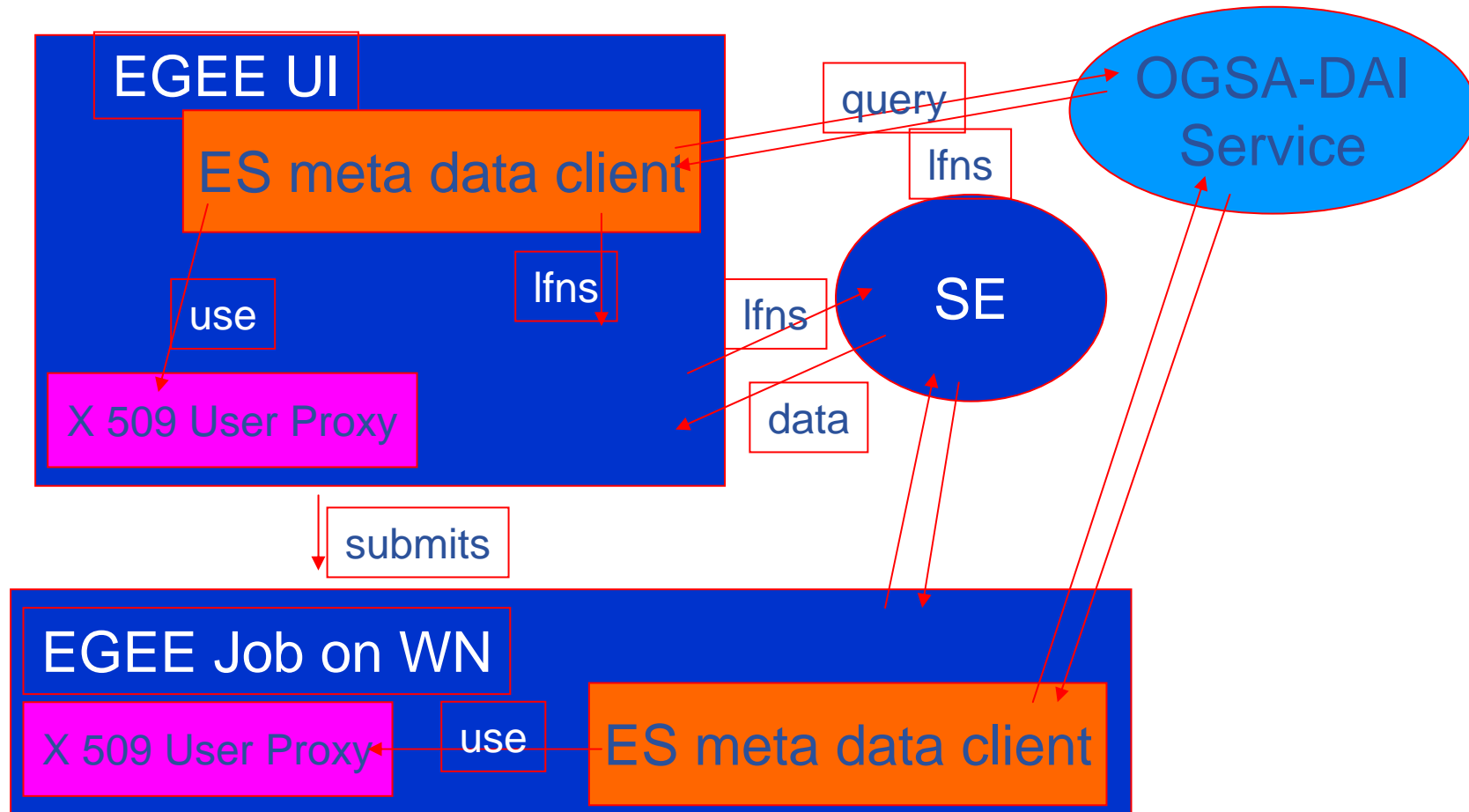| Data Centers	Current Volume	Grid enabled
DKRZ Archive	~4 PB	~3 TB
WDCs (Climate/Mare)	~200 TB	~5 TB
IFM Geomar	~1 TB	~500 GB
DWD	~200 GB	
FUB	~1 TB	The rest is coming soon...
PIK	~700 GB	
AWI	~300 GB	
DLR	~60 GB	

Enabling Grids for E-sciencE

- Feedback from EGEE to C3 (publish updated metadata of AMGA for the C3 portal)

- Mapping and interoperability of the AA infrastructures of EGEE, C3 and DBs

- Direct and transparent transfer of external files to, and registration in EGEE

  - That is, automatic selection of a close and free SE for storage

eGee

Enabling Grids for E-sciencE
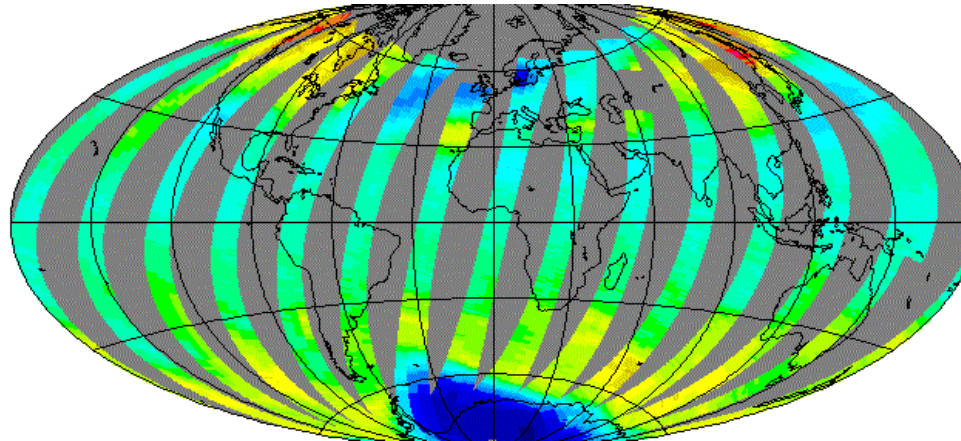
- The goal is to develop for a specific case a prototype that includes the needed tools:

- **Example**: Two different instruments : **Ground-based Lidar, spectrometer aboard the satellite, ERS**.

- The satellite data stored by orbit or pixel; different algorithms

- The Lidar data stored in monthly files with one profile/night

egee Enabling Grids for E-sciencE

- SL 4.1
- Web-Service Container: Tomcat 4.1.31
- OGSA-DAI OGSI 6.0 with GLOBUS 3.2.1 (TLS by Port 8443)
- Three different resources today

  - MySQL 4.1.10

    MySQL spatial extensions only support convex polygons

  - PostgreSQL 7.4.8 + PostGIS (production)

    PostGIS adds support for geographic objects to Postgres: http://postgis.refractions.net/

  - Oracle 10g   (also for Bio Applications)

eGee

Enabling Grids for E-sciencE

EGEE UI

ES meta data client

query

lfns

OGSA-DAI Service

use

lfns

lfns

SE

X 509 User Proxy

data

submits

EGEE Job on WN

X 509 User Proxy

use

ES meta data client

egee

Enabling Grids for E-sciencE

- straight forward installation by SCAI no integration

- fat client on nodes -- only for Authorisation (Globus)

- User Authentication
  - with grid proxy certificates
  - mapping to db roles for every user

```
<User dn="/O=GRID-FR/C=FR/O=CNRS/OU=IPSL/CN=David Weissenbach"
    userid="lidar_writer"password="***" />
```

eGee

Enabling Grids for E-sciencE

Advantage:
- access to existing databases - nothing to convert
- out-of-the box installation
- easy to extend by own classes
- "quasi industrial standard"
- multiple resources with multiple services

Disadvantage:
- not fast
- scalable over the resources ?
- not integrated in gLite