European Grid of Solar Observations

EGSO

Infrastructure for the Space Weather Community

R.D. Bentley and the EGSO Team

Presentation to Space Weather Workshop
16-18 December 2002, ESTEC

European Grid of Solar Observations

Outline

- Overview of EGSO
- The problems EGSO will solve
- Details of how EGSO will function
- Enhancements for Space Weather

EGSO – European Grid of Solar Observations

European Grid of Solar Observations

- EGSO is a Grid test-bed related to a particular application
  - Designed to improve access to solar data for the solar physics and other communities
  - Addresses the problem of a distributed heterogeneous data set and a scattered user community
- Funded under the Information Society Technologies (IST) thematic priority of the EC's Fifth Framework Program (FP5)
  - Started March 2002; duration of 36 months
- Eleven groups in Europe and the US, led by UCL-MSSL
  - 4 in UK, 2 in France, 2 in Italy, 1 in Switzerland, 2 in US
  - Several associate partners, mainly in the US
  - EGSO, the US VSO & CoSEC working closely together
    - Successful joint meeting in October 2002 at MSSL
  - EGSO also collaborating with ESA's study project SpaceGRID
- *Currently working on details of the architecture and developing demonstration testbeds*

The EGSO Team

European Grid of Solar Observations

- UK
  - *UCL-MSSL*, UCL-CS, RAL, Univ. Bradford, Astrium
- France
  - IAS (Orsay), Observatoire de Paris-Meudon
- Italy
  - Istituto Nazionale di Astrofisico, Politecnico di Torino
    - INAf includes observatories of Turin, Florence, Naples and Trieste
- Switzerland
  - Univ. Applied Sciences (Windisch)
- US
  - SDAC (NASA-GSFC), National Solar Observatory

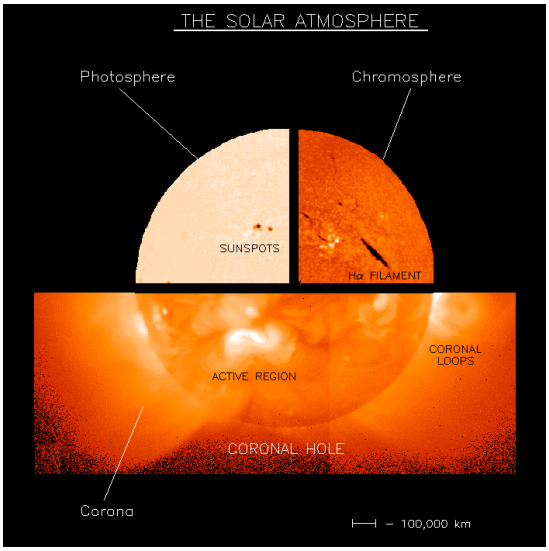
European Grid of Solar Observations

The Extended Team

- **UK**
  - *UCL-MSSL*, UCL-CS, RAL, Univ. Bradford, Astrium
- **France**
  - IAS (Orsay), Observatoire de Paris-Meudon
- **Italy**
  - Istituto Nazionale di Astrofisico, Politecnico di Torino
    - INAf includes observatories of Turin, Florence, Naples and Trieste
- **Switzerland**
  - Univ. Applied Sciences (Windisch)
- **Netherlands**
  - ESA – Solar Group
- **US**
  - SDAC (NASA-GSFC), National Solar Observatory (VSO)
  - Stanford University, Montana State University (VSO)
  - Lockheed-Martin (CoSEC)

European Grid of Solar Observations

Nature of solar observations



- The appearance of the Sun changes dramatically with wavelength
- For a complete picture, need to use as wide a range of observations as possible
- Identifying what observations were made and then retrieving them is a major obstacle

Overview of the generic problem

European Grid of Solar Observations

- **Observations used to build up a picture of the plasma in multi-dimensional parameter space (incl. *x, y, z, t, T* & ρ)**
  - Users need access to as many wavelengths as possible
  - For technical and practical reasons:
    - UV, EUV, X-rays and γ-rays observed from space
    - Radio and optical wavelengths observed from the ground (coverage)
- **Data centres and observatories located around the world**
  - Increasing data volumes, etc. require new methodology
  - Large and small facilities (with varying resources)
  - Aim is to make it easy to add new data sets
- **Users scattered around the world**
  - Do not need to know where the data is located
  - Capabilities of users computing vary greatly
  - Authentication issue needs serious consideration
    - Want to minimize how this affects the user

Overview of generic application

European Grid of Solar Observations

- **Identify suitable observations** *(many serendipitous)*
  - Should be possible without accessing the data
  - Catalogues differ in quality, contents, and dependencies
- **Locate the data**
  - Data scattered, with differing means of access (some proprietary)
  - Often only need a subset of each data set
- **Process the data**
  - Involves extraction and calibration of a subset of data
  - Uses code defined by instrument teams (*SolarSoft*, C…)
- **Return results to the User**
- **Compare results from different instruments**
  - *SolarSoft* (IDL) provides a standard platform for analysis

Note the interchange in the order of bullets 3 and 4 in the Grid solution when compared to current practice

EGSO work packages

European Grid of Solar Observations

EGSO work package structure reflects the generic application

- **Catalogues (WP4 & WP5)**
  - Means of identifying what observations are available
- **Catalogue search and visualization tools (WP3)**
  - User interface to search catalogues for suitable observations
- **Data retrieval and processing mechanisms (WP2)**
  - Locate best source of requested data
  - Select subset, process and return them
- **System Definition and Integration (WP1)**
  - Define overall requirements and constraints (user input)
  - Architectural Design; define integration and testing
  - Integration of the components of other WPs
- **Project Control and Dissemination (WP0)**

EGSO – Search Engine

European Grid of Solar Observations

- **Enhanced "cataloguing" used to select observations**
  - Metadata versions of observing catalogues tie together heterogeneous data sets – data itself left untouched
  - New types of catalogues allow searches on events, features and phenomena rather than just date & time, pointing, etc…
- **Ancillary data (images, time series, etc.) provide additional search criteria**
  - Also enhance capabilities of the (Graphic) User Interface
- **Search Registry allows hierarchical optimization**
  - Entries describe metadata/data available for search
  - Registry replicated to provide resilience and load sharing
- **Alternate entry point (to User Interface) allows access by researchers from other communities or Grids**
  - Astrophysics, Climate Physics, Space Weather…

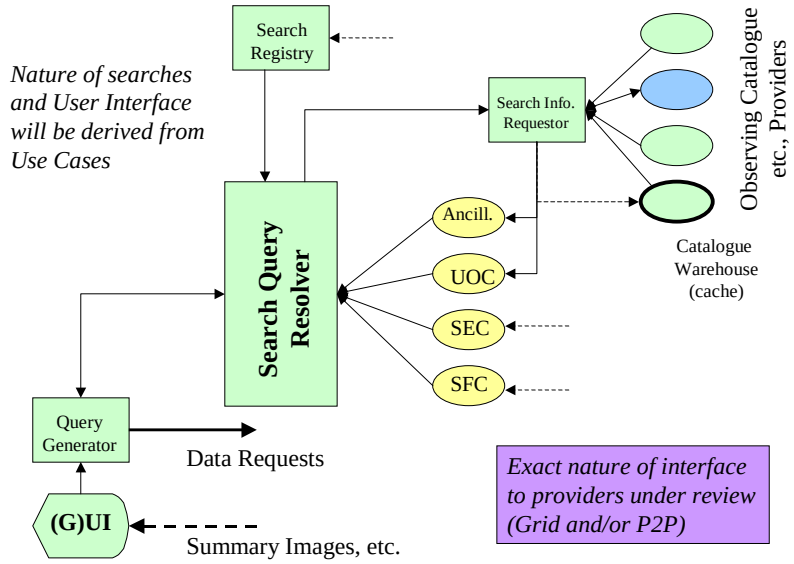
European Grid of Solar Observations

The enhanced solar catalogues

- **Unified Observing Catalogues (UOC)**
  - Metadata form of catalogues needed to tie together the heterogeneous data, leaving original data unchanged
  - Self describing (e.g. XML), quantised by time and instrument, with no dependencies on *ancillary data* or proprietary software and any errors corrected
  - *Standard* defined for future data sets (e.g. STEREO, iLWS, Solar-B)
- **Solar Event Catalogues (SEC)**
  - Built from information contained in published lists
  - Flare lists, CME lists, lists in SGD, etc.
- **Solar Feature Catalogue (SFC)**
  - Lists of the occurrence of events, phenomena and features provides an alternate means of selecting data
  - Derived using image recognition software

European Grid of Solar Observations

EGSO – Query Resolving



EGSO – Handling the data

European Grid of Solar Observations

- **An objective is to dramatically enhance access to the data**
  - User only needs to know observations exist, not where located
  - System able to optimize use of sources (closest, least used…)
  - Addition of new sources made as simple as possible
- **Process as much data at source as possible**
  - Extraction and calibration of subset of data
    - Solar data are usually stored in raw form
    - Software for processing defined by instrument team (IDL, C…)
  - Processing reduces volumes of data moved around
  - Simplifies requirements on user's own system
- **Standard (pipe-line) processing adequate for many users**
- **More complex problems require ability to upload code**
  - Used in analysis of extended data sets (helioseismology, etc)
  - System allocates resources; Security an issue
  - Models and simulations have similar requirements

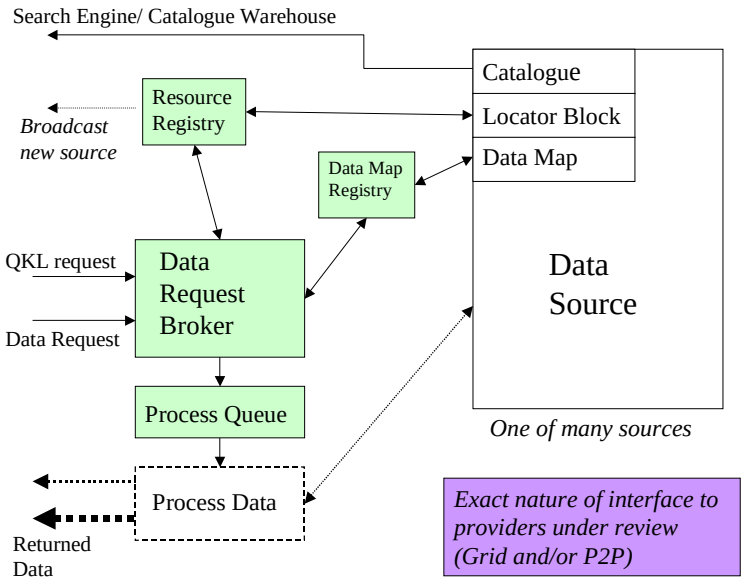
Resource Usage

European Grid of Solar Observations

- **Resources available to EGSO are described in the Resource Registry**
  - These include: data, processing, storage, service, …
  - Some providers can support multiple capabilities
  - Entry for a data resource just identifies what data stored where, how they are accessed, etc.
    - Allows handling of replicated data and aggregated sources
    - Data resources described in more detail in the **Data Registry**
- **Resource Broker allocates and monitors resources needed to satisfy user requests**
  - Controls processing of data & staging of results
    - Control how much being requested of a particular provider
    - Processing may be at different site to data provider
    - Plans to support use of multi-instance processing and HPC
  - Broker & Registries replicated to provide system resilience and permit load sharing

Locating & Processing the Data

European Grid of Solar Observations



Project Status

European Grid of Solar Observations

- **The EGSO project falls into four phases**
  - I. Project definition; consult with the community; explore and experiment with technologies
  - II. Architectural design; system integration and validation plan
  - III. Implementation of of design; development of middleware and catalogues
  - IV. Product commissioning and delivery
  - Note: There are no clean breaks between phases
- **Project currently in Phase II**
  - Detailed list of requirements gathered
    - Web survey of users and science cases from individuals
    - Used to define Technical Requirements
  - Working on 1st iteration of EGSO Architecture
  - Work packages drawing up detailed plans

Additional needs of Space Weather

European Grid of Solar Observations

- **More integrated access to STP and other data**
  - Types of data more important than specifics
  - Ensure that cataloging,etc. and registry compatible
- **Immediacy of access to data**
  - Most access to EGSO likely to be for older data
  - Near-casting needs data a few minutes old
    - e.g. CME passing L1 will hit magnetosphere 30-60 mins. later
    - Generally only need modest amounts of data
  - Forecasting needs data a few hours old
    - e.g. CME starting from Sun will arrive 36-72 hours later
    - More complex data requirements
- **Ability to run models using data or derived parameters as input**

Compatibility of current design

European Grid of Solar Observations

- **Aspects of design of EGSO already compatible**
  - Interface to data provider interface designed to be as simple as possible
    - Want to be able to add as many sources as possible
  - Unified Observing Catalogues formed in fragments
    - only update the latest one; provider defines update frequency
  - Ancillary data – e.g. time series – already used by search engine
  - Stripped down Search Registry could focus on space weather providers
    - STP data could be thought of as ancillary data or SEC data
  - Uploading and execution of code will support models and simulations
- **Need to ensure EGSO detailed design does not prohibit things required for Space Weather**

Conclusions

European Grid of Solar Observations

- EGSO is providing enhanced access to solar data for the solar and related communities
- Much of this infrastructure could be adapted as the basis of a European space weather system
- EGSO has already established close links with counterparts in the US, and relevant projects including iLWS and STEREO
- *Currently working on details of the architecture and developing demonstration testbeds*

- **For more information on EGSO see:**
  - http://www.egso.org
- **Or e-mail**
  - bentley@egso.org