



The Abdus Salam
International Centre for Theoretical Physics



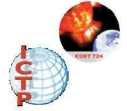
310/1749-26

ICTP-COST-USNSWP-CAWSES-INAF-INFN
International Advanced School
on
Space Weather
2-19 May 2006

EGSO
A Maturing VO for Solar Physics

Robert BENTLEY
UCL Department of Space and Climate Physics
Mullard Space Science Laboratory
Hombury St. Mary
Dorking
Surrey RH5 6NT
U.K.

These lecture notes are intended only for distribution to participants



EGSO

A maturing VO for Solar Physics

R.D. Bentley, UCL-MSSL
(and the EGSO Team)

3 May 2006
ICTP, Trieste



Overview

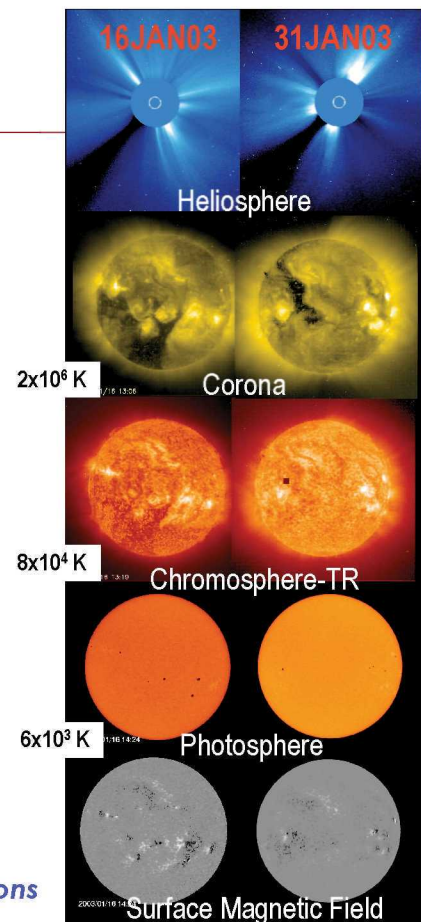


- Review of objectives of EGSO
- EGSO system
- Special Providers
- Interfaces into EGSO: GUI & API
- Example of application software
- Lessons learnt

- **EGSO is a Grid test-bed related to a particular application**
 - Aimed to improve access to solar data for the solar physics and other communities
 - Addresses generic problem of a distributed and 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)**
- **Involves 12+ groups in Europe and the US, led by UCL-MSSL**
 - 4 in UK, 3 in France, 2+ in Italy, 1 in Switzerland, 2 in US
 - **Several associate partners, mainly in the US**
- **Objectives include:**
 - Building enhanced search capability for solar data
 - Support of user community scattered around the world
 - Provide access to data sources around the world
 - Where possible, provide ability to process data at source

Generic Problem

- **Identify suitable observations**
 - Need to identify when and where events and features of interest occurred
 - Many observations serendipitous
 - Want to access as many different types of data as are available
 - Both space and ground based data important
- **Locate the data**
 - User only wants to know if data addressing a scientific problem exists, not where or how they are stored
- **Process the data**
 - Extraction and calibration of a subset of the *raw data*
 - Growing problem with complexity of environment user needs to do this
- **Return results to the User**
- **Compare data from different instruments**
 - Assume scientist responsibility using existing analysis tools



Problem similar for other communities using observations



Need to build on existing situation



- **Data sources located around the world**
 - Large and small providers; varying resources and means of access
 - Need to make it as easy as possible to add new data sets
 - Planned data volumes **much larger** than for current instruments
 - Virtual observatory approach will be key in handling them
- **User community scattered around the world**
 - Capabilities of users & their computing facilities vary greatly
- **Well established analysis environment**
 - SolarSoft (IDL) is a standard platform for analysis of solar data
 - Use what exists and require minimum change by the users
 - EGSO needed to tie back into these capabilities
- **Other virtual observatories in solar and related disciplines are under development - beneficial to collaborate**
 - VSO, VSPO, VHO, VSTO and CoSEC
 - Working together provides a more capable system



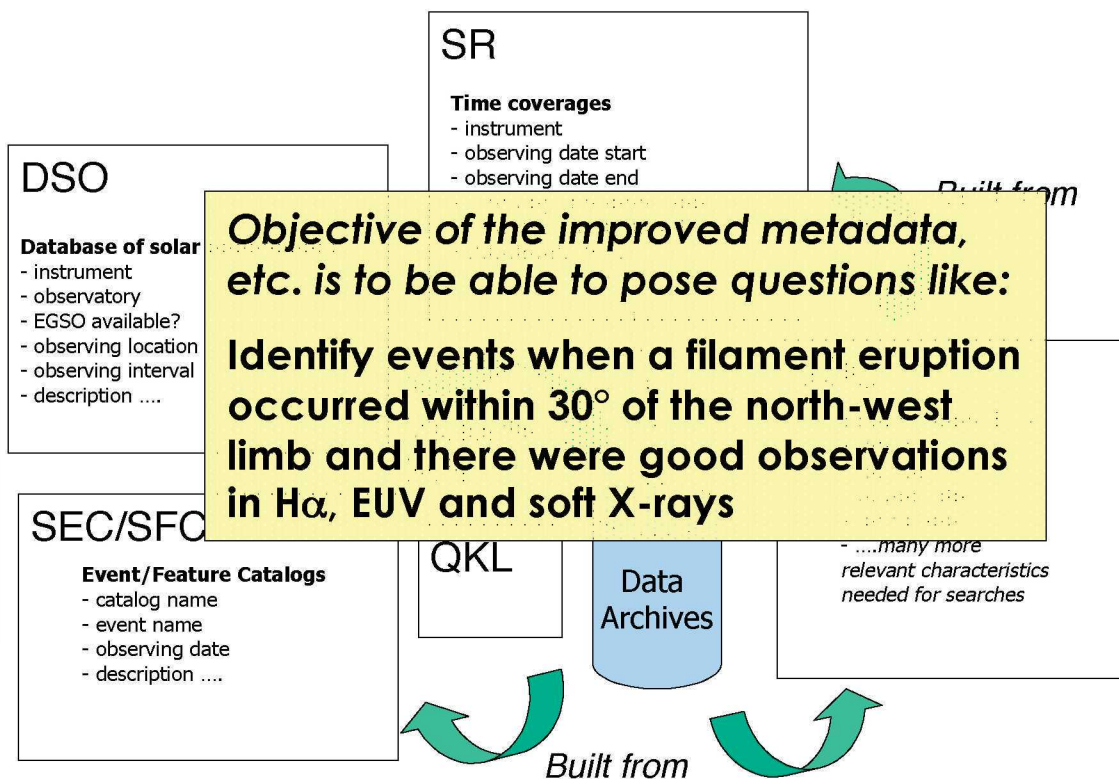
Collaborations between Projects

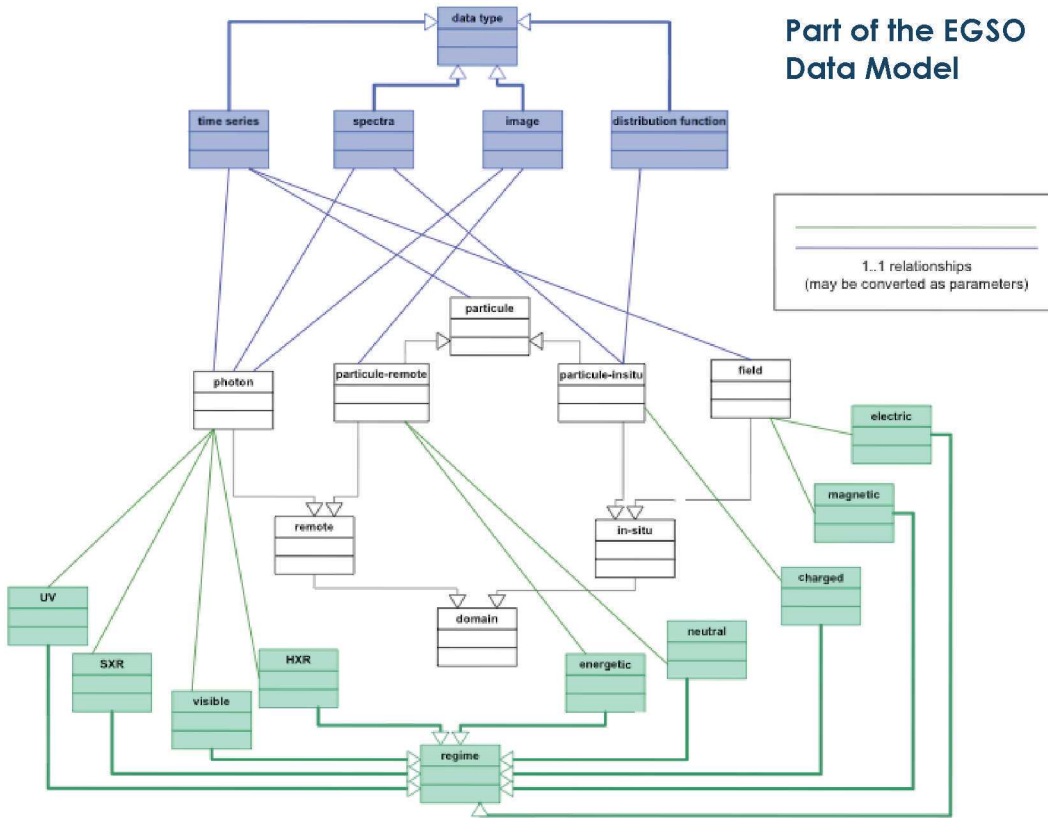


- **Virtual Solar Observatory (US-VSO)**
 - Funded by NASA as a replacement for SDAC
 - Differences wrt EGSO in scale of objectives – big/small box...
 - SDAC (NASA-GSFC) and NSO are partners in EGSO
 - EGSO Coordinator in Chair of VSO Steering Committee
- **Sun-Earth Connector (CoSEC)**
 - Funded by NASA under the LWS programme
 - Significant synergies between EGSO and CoSEC
 - EGSO using CoSEC services to provide additional capabilities
 - EGSO Coordinator is a Col in CoSEC continuation grant
- **Virtual Space Physics Observatory (VSPO)**
 - Funded by NASA and recently established
 - Discussions helping ensure EGSO provides full access to STP data
 - The SPASE Data model particularly helpful
- **See URL:** <http://www.mssl.ucl.ac.uk/grid/iau>

- **New types of catalogue allow searches that identify events, phenomena and features of interest**
 - Solar Event Catalogue (**SEC**) - derived from published lists
 - Solar Feature Catalogue (**SFC**) - generated by feature recognition
- **Enhanced cataloguing & Registries describe the data**
 - **Static Search Registry (sSR)** describes the type of observations an instrument makes
 - **Dynamic Search Registry (dSR)** identifies which instruments were observing during intervals of interest - allows narrowing of the search in initial stages
 - Standardized versions of observing catalogues (**UOC**) tie together the heterogeneous data sets
 - dSR can be an abstraction of the UOC, or derived from Provider
 - **Ancillary data** used to provide additional search criteria
 - DSO Server gives Yellow Page information on instruments, etc.
 - QLK Server provides Phone book access to images, time-series, derived products, etc.; can also do limited processing

Similar hierarchical cataloguing required in other data Grid projects

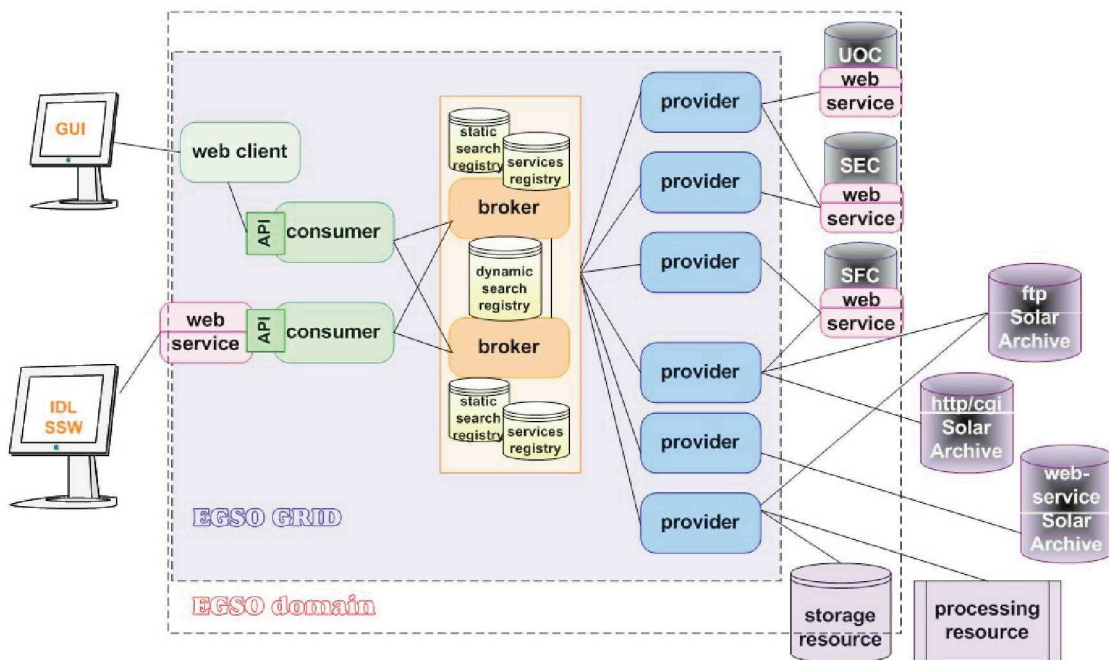




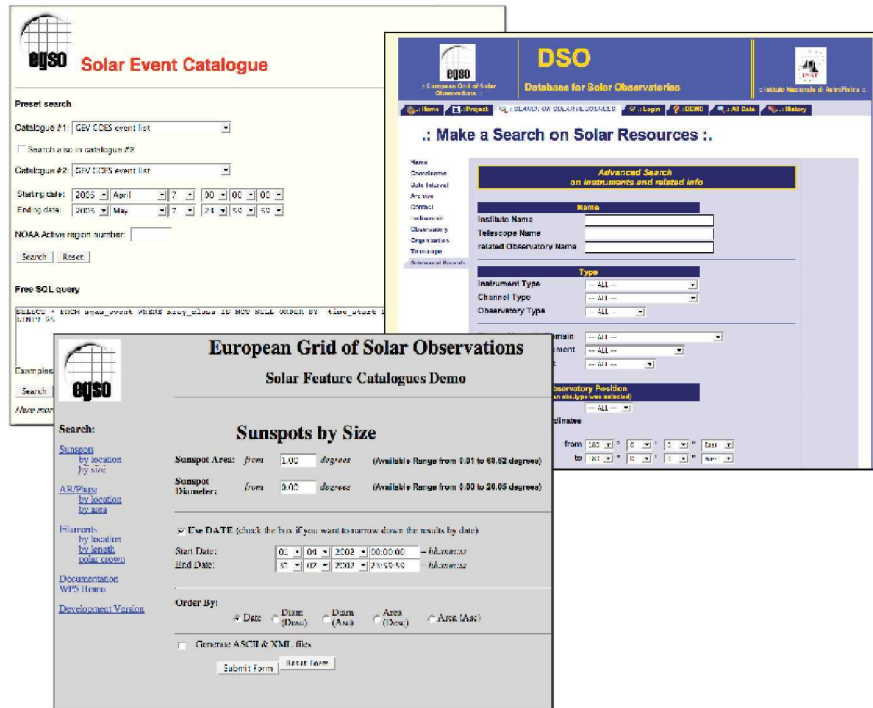
Architecture - simplified



European Grid of Solar Observations



Implemented as 3 roles - Consumer, Broker & Provider - with a communications layer



- **Solar Event Catalogues (SEC)**
 - Lists of the occurrence of energetic events and phenomena
 - Important criteria for selecting intervals of interest
 - Built from information contained in published lists
 - Includes 17 lists from late 1970s onwards
 - Flare lists, CME lists, lists in Solar Geophysical Data (SGD), etc.
 - Specialized lists prepared by individuals, e.g. EIT Waves

- **Solar Feature Catalogues (SFC)**
 - Lists of the occurrence of features and phenomena
 - Contains sunspot, filament and AR plage data for 1996-2004
 - Created using image recognition software developed in WP5
 - Provides an alternate means of selecting data

- **Database of Solar Observatories (DSO)**
 - Provides details of all instruments and observatories whether they are integrated in EGSO or not
 - Removes need to carry this information within other metadata
 - sSR is a very flat abstraction of the type of information in the DSO
 - Allows identification of instruments that could have made observations depending on date, location, etc.



EGSO Special Providers



- **Stand-alone servers providing special services for main system**
 - Relational database with Web Service interface
 - Supports SQL queries; results returned in VOTable format
 - Contents depends on server, but standardized where possible
 - Complexity of creating/updating database hidden from system
 - Servers can also be accessed by other organizations
 - Web interface also provided for the user
- **Systems developed in parallel to the main system**
 - Reduced complexity of main system
 - Allowed better deployment of effort within the project
- **Resilience of system increased by deploying multiple instances**
 - Widely separated locations to reduce network problems
 - SEC (3; Ts, UCL, Na), DSO (2; Na, To), UOC (2; To, Ts)
 - SFC (installed at BU, 2nd instance to be at Obs. Paris-Meudon)
 - Automatic fail-over by system if a Special Provider has problems
- **UOC & QLK are similar, but intended for internal use**



Solar Event Catalogue

Results as VOTable File

The VOTable (XML) result file is [sec1_20050405_123810.xml](#)
The TXT result file is [sec1_20050405_123810.txt](#)

SQL query: `SELECT * FROM goes_xray_flare WHERE time_start>='2002-07-15 00:00:00' AND time_start<='2002-07-15 23:59:59' ORDER BY time_start;`

Query using presets or free SQL

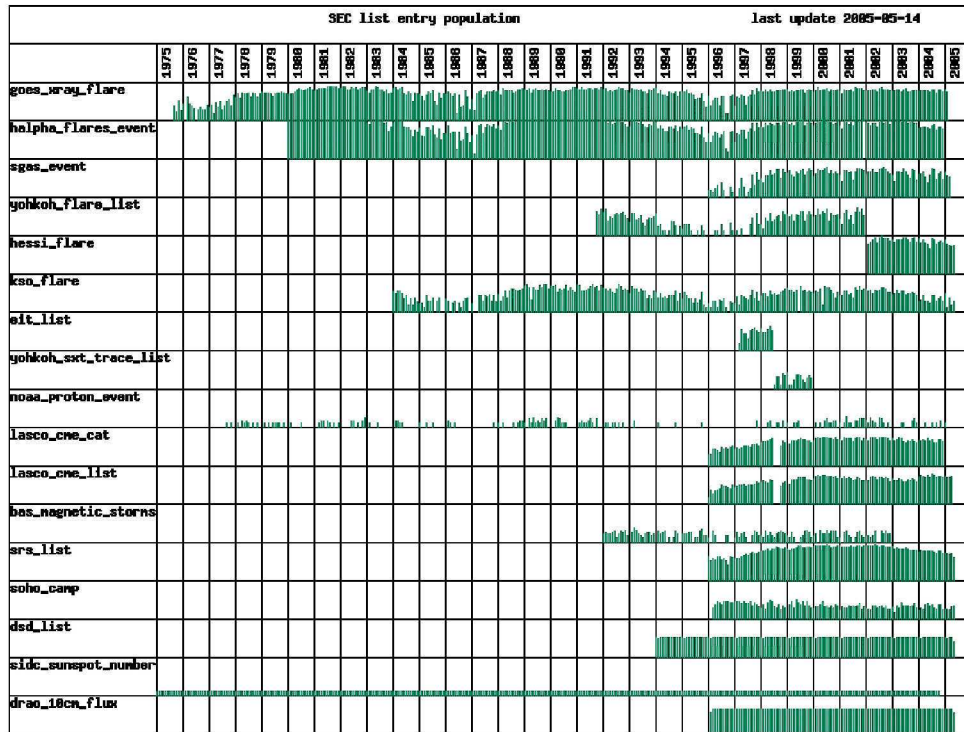
goes_id	ntime_start	time_start	time_peak	ngitude	long_carr	xray_class	optical_class
53524	2002-07-15 00:42:00	2002-07-15 00:46:00	2002-07-15 00:50:00			C1.8	
53525	2002-07-15 01:58:00	2002-07-15 02:00:00	2002-07-15 02:02:00	10030 24	-8	16.6	C1.6 sf
53526	2002-07-15 04:05:00	2002-07-15 04:14:00	2002-07-15 04:23:00	10030 19	-8	15.37	C1.6 sf
53527	2002-07-15 04:54:00	2002-07-15 05:03:00	2002-07-15 05:12:00	10030 21	-9	13.92	C3.8 sf
53528	2002-07-15 09:55:00	2002-07-15 10:12:00	2002-07-15 10:29:00	10030 19	-4	16.08	C1.6 sf
53529	2002-07-15 11:23:00	2002-07-15 11:39:00	2002-07-15 11:55:00	10030 20	-4	15.28	C9.1 sf
	2002-07-15 14:52:00			10030 18	-1	16.99	C3.6 sf
	2002-07-15 19:52:00			10030 17	1	16.44	C1.9 sf
	2002-07-15 20:56:00			10030 19	1	15.69	X3.0 3b
	2002-07-15 23:40:00			10030			M1.8

```
<?xml version="1.0" ?>
<!DOCTYPE VOTABLE (View Source for full doctype,...)>
<VOTABLE version="1.0">
  <DEFINITIONS />
  <RESOURCE type="results">
    <DESCRIPTION>EGSO SEC Server</DESCRIPTION>
  <TABLE>
    <FIELD name="goes_id" datatype="int" />
    <FIELD name="ntime_start" datatype="char" arraysize="3400" />
    <FIELD name="time_start" datatype="char" arraysize="3400" />
    <FIELD name="time_peak" datatype="char" arraysize="3400" />
    <FIELD name="time_end" datatype="char" arraysize="3400" />
    <FIELD name="ntime_end" datatype="char" arraysize="3400" />
    <FIELD name="nar" datatype="int" />
    <FIELD name="latitude" datatype="float" />
    <FIELD name="longitude" datatype="float" />
    <FIELD name="long_carr" datatype="float" />
    <FIELD name="xray_class" datatype="char" />
    <FIELD name="optical_class" datatype="char" />
  </TABLE>
</VOTABLE>
```

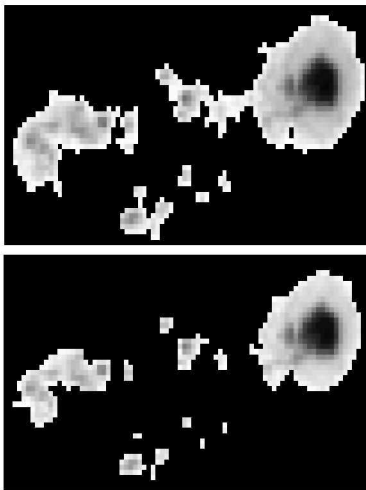
VOTable standard developed within IVOA

Designed and maintained by INAF - Trieste Astronomical Observatory

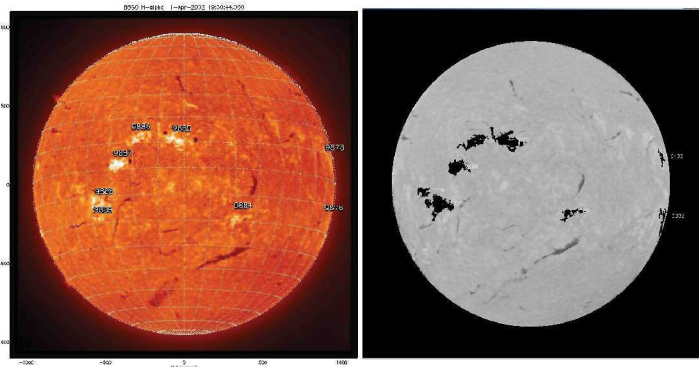




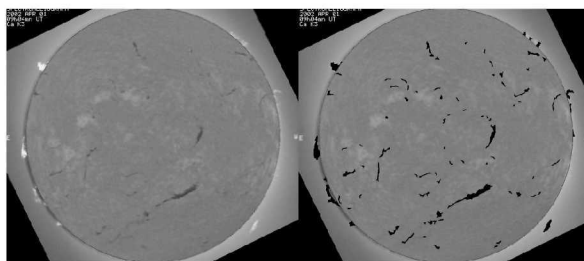
- Sunspot detection – edge detection



- Active region detection -RG



- Filament/prominences detection -RG



SFC includes data from 1996-2004
Results as VOTable File

- **Data could be located anywhere in the world**
 - User only needs to know observations exist, not where located
 - System isolates user from the intricacies of access
 - Burden on providers minimized to encourage participation
- **Provider role uses adaptor modules to handle the different access protocols of the data providers**
 - Supports HTTP, FTP, Web Services, SQL...
 - Sources used include other VO's - **VSO**, **VSPO** - and **CoSEC**
 - Data source interface appears in standardized way to the system
 - Simplifies addition of new sources
- **If possible, desirable to process the data "at source"**
 - Involves extraction and calibration of a subset of the *raw data*
 - Software for processing defined by instrument team (IDL, C...)
 - Processing reduces volumes of data moved around
 - Simplifies requirements on user's own system
 - **Resource problem** - some providers unable to support this
- **Standard (pipe-line) processing adequate in most cases**
 - Use **CoSEC** SolarSoft Servers for TRACE data

Observatory/ Network	Instrument(s)	Archive (1)	Status
SOHO	CDS, EIT, LASCO, MDI, SUMER, SWAN, UVCS CELIAS, COSTEP, ERNE, GOLF and VIRGO (2)	GSFC-SOHO (V)	available
RHESSI	RHESSI <i>(returns data plus observation summary files)</i>	GSFC (D+F)	available
Yohkoh	BCS, HXT, SXT and WBS	GSFC-SDAC (F)	available
TRACE	TRACE <i>(EGSO UOC used to refine access) (data extraction and calibration using CoSEC under study)</i>	GSFC (D+H)	in prep.
Coriolis	SMEI	NSO (H)	available
GOES-12	SXI <i>(EGSO UOC used to refine access)</i>	NGDC (D+H)	available
Coronas-F	SRT (SPIRIT), DIFOS <i>(DIOGENESS, RESIK under study; data scattered)</i>	IZMIRAN (H)	available
ACE	EPAM, SWEPAM, MAG	GSFC (H)	available
Ulysses	EPAC, VHM/FGM	GSFC (H)	available
GHAN	H-alpha from BBSO, KANZ, OACT, YNAO & HSOS (Global H-alpha Network)	BBSO (F)	available
GONG	Summary data from BBSO (Global Oscillations Network Group)	BBSO (F)	available
Meudon	Spectroheliograph and White Light Telescope	BASS2000 (D+H)	available
Pic-du-Midi	Coronagraph	BASS2000 (D+H)	available
Cesco Stn.	H-Alpha Solar Telescope for Argentina (HASTA)	OAFA (H)	available
Kitt Peak Sac Peak	KPVT and SOLIS (VSM, ??) Spectroheliograph	NSO (V)	available
San Fernando	Cartesian Full Disk Telescopes	KSFO (H)	in prep.
MLSO	Coronagraph (Mk IV), CHIP and PICS	HAO (V)	available
Nobeyama	Radioheliograph	NOAJ (H)	available
Nancay	Radioheliograph, Decametric Array, TRFA	BASS2000 (D+H)	available
OVRO	Owens Valley Solar Array (radio)	OVRO (V)	available



European Grid of Solar Observations

Introduction | Help | Tools

EGSO GUI - BUILD QUERY

New query | **Build query** | List of events | List of observations | List of files

QUERY TYPE

Standard Query

Registry Based Query

SEARCH PARAMETERS

Observable Entity

RESULT PARAMETERS

Select Plots

Select Fields

DATE/TIME

Start Date: 2002-07-15 00:00:00

Format: 2003-12-31 23:59:59

End Date: 2002-07-16 00:00:00

Format: 2003-12-31 23:59:59

EVENT

None

GEV GOES Event List

NOAA Proton Events

LASCO CME Catalogue

LASCO Preliminary CME List

BAS Magnetic Storms

NOAA SRS Active Regions

X-Ray Class

from: [-]

to: [-]

OBSERVABLE ENTITY

Gamma Rays

XRays: SXR HXR

Ultra Violet: EUV UV

Visible (inc. H alpha and 10830A)

Infrared

Microwaves

Radio Waves

SELECT PLOTS

none

GOES X-Ray Plot

European Grid of Solar Observations



EVENT LIST

Download list as VOTable

10 events listed

<input type="checkbox"/>	Date	Time Peak	Nar	XRayClass	Long. /Lat.	Detail
<input type="checkbox"/>	2002-07-15 00:46:00 - 2002-07-15 00:55:00	2002-07-15 00:50:00		C1.8		
<input type="checkbox"/>	2002-07-15 02:00:00 - 2002-07-15 02:10:00	2002-07-15 02:02:00	10030	C1.6	-8.0 / 24.0	
<input type="checkbox"/>	2002-07-15 04:14:00 - 2002-07-15 04:34:00	2002-07-15 04:23:00	10030	C1.6	-8.0 / 19.0	
<input type="checkbox"/>	2002-07-15 05:03:00 - 2002-07-15 05:22:00	2002-07-15 05:12:00	10030	C3.8	-9.0 / 21.0	
<input type="checkbox"/>	2002-07-15 10:12:00 - 2002-07-15 10:39:00	2002-07-15 10:29:00	10030	C1.6	-4.0 / 19.0	
<input checked="" type="checkbox"/>	2002-07-15 11:39:00 - 2002-07-15 12:17:00	2002-07-15 11:55:00	10030	C9.1	-4.0 / 20.0	
<input type="checkbox"/>	2002-07-15 14:00:00 - 2002-07-15 14:10:00	2002-07-15 14:04:00	10030	C3.6	-1.0 / 18.0	
<input type="checkbox"/>	2002-07-15 18:37:00 - 2002-07-15 18:49:00	2002-07-15 18:40:00	10030	C1.9	1.0 / 17.0	
<input type="checkbox"/>	2002-07-15 19:59:00 - 2002-07-15 20:14:00	2002-07-15 20:08:00	10030	X3.0	1.0 / 19.0	
<input type="checkbox"/>	2002-07-15 21:03:00 - 2002-07-15 21:48:00	2002-07-15 21:32:00	10030	M1.6		

VIEW QUERY

VIEW CONTEXT

New query | **Build query** | **List of events** | List of observations | List of files

European Grid of Solar Observations



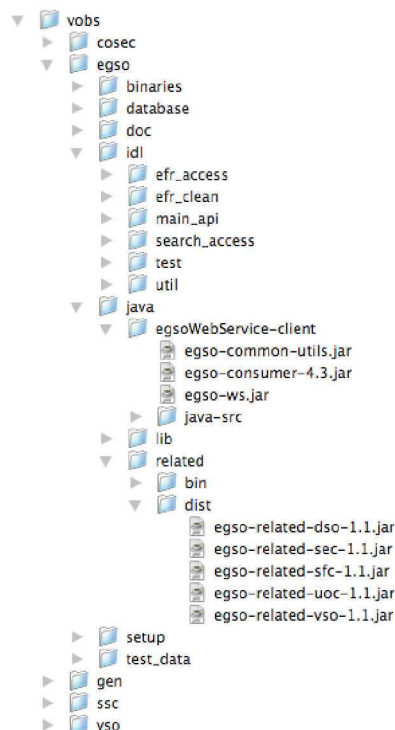
GUI Capabilities



- **Supports time intervals defined by:**
 - Simply entering a date/time range
 - An event selected from the Solar Event Catalogues
 - An entry in an ingested VOTable
- **Supports refinement of instrument selection based on:**
 - Domain: photon (HXR, SXR, EUV, UV...), particle (enegetic...)
 - Type of observation: spectrometer, coronagraph...
 - Object being observed: disk, corona, heliosphere
- **Allows user to make selection at each stage of the search**
- **Provides assistance to the user:**
 - Ability to examine details related to a flare
 - Ability to view image files - includes a Web-based FITS Viewer

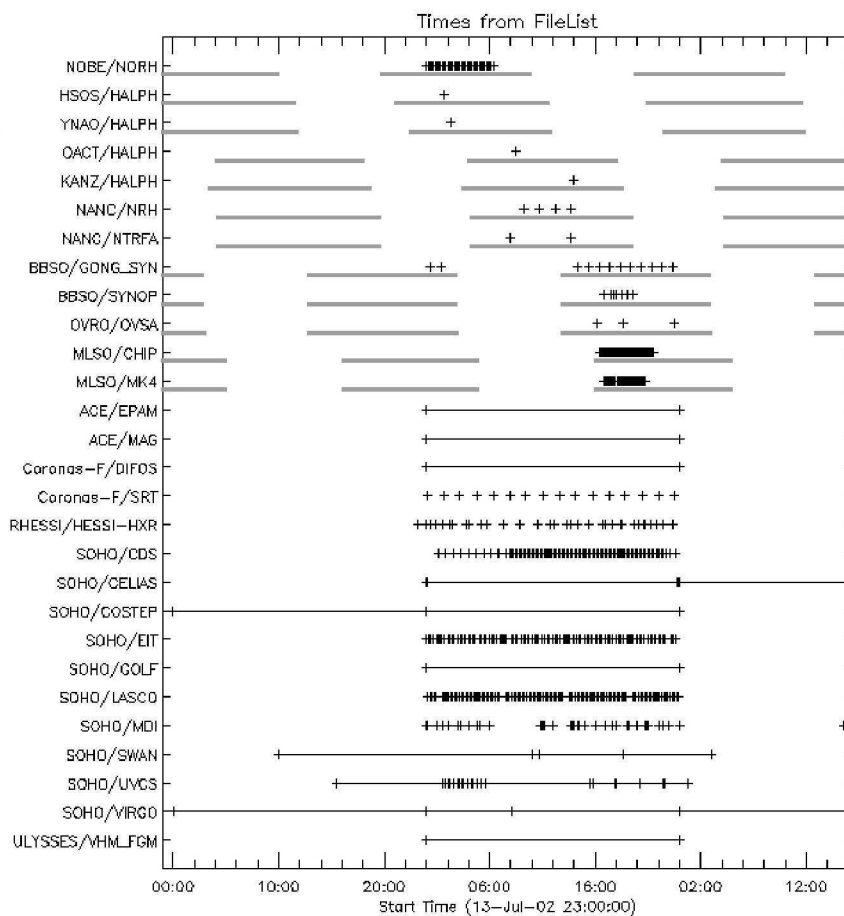


EGSO API and SolarSoft



- **Special Providers can be accessed directly from SolarSoft**
 - Uses same Java code as is used for their interface into main system
- **SolarSoft routines make requests to SEC and SFC; lists retrieved directly**
 - Applications software available to over-plot these data on FITS images
- **Access into main EGSO system from SolarSoft being finalized**
 - Supports same capability as the GUI using object oriented code
 - User will be able to request data directly from user SSW session
 - Interface uses same Java library as the Special Providers

- In general grids are “batch-oriented” systems
 - Difficult to display intermediate results, to go back and forth between data products, to perform own “client” analyses
- EGSO has developed access via an interactive environment.
 - In an interactive environments, users have better control
 - Tested within IDL/SolarSoft - widely used by the community
 - Can also be accessed by other system, VO, etc.
- Within EGSO API, users can:
 - Access the entire functionality offered on the GUI
 - Discover the parameter space (global schema) in real time
- API provides enormous potential for further development





EGSO Documentation



European Grid of Solar Observations

Help pages for the EGSO SolarSoft Code Available help: EGSO System API

How to work with the egso object

Last update: Help pages for the EGSO GUI/Consumer Available help: Standard Query: List-driven

This document describes how to use the egso object. Please check use the egso object you need to have solarsoft installed.

Note: the parameter values shown on the right side of the equals sign are the values by whatever your own values are.

- 1. Create an egso object:**
 - `o = egso()`
 - This is the access point (handle) which connects you to the system. It is used to get query results.
- 2. Get to know your environment**
 - `print, o->which(/feature)`
Tells you what feature catalogues you can search on (coming soon)
 - `print, o->which(/event)`
Tells you what event catalogues you can search on.
- 3. Set general parameters**
 - Now you can set the parameters that will narrow down your search.
 - `o->set, time_range = '2002/02/20 ' + ['10:00:00', '2002/02/20 ' + ['11:00:00']`
Sets the time range on which to search the data. All times are in UTC.
 - `o->set, event = ['goes', 'noaa']`
Sets the system to search in that specific event catalogue which command:
 - `o->set, event = [0,1]`
is equivalent to the former command.

Standard Query: List-driven

In a List-Driven Query, the EGSO Solar Event Catalogue (SEC) is queried for events, in the selected list, within the time range entered by the user. At any stage the results can be saved to a VOTable file.

- When you select an event list an extra bar will appear at the top called 'List of Events'.
- Select the time interval you want using the Date/Time dialogue box and then hit the 'List of Events' button.

New Query Build Query List of Events List of Observations List of Files

- After a short delay, the selected list of events will appear in the main window - this is generated by a query to the EGSO SEC Server. A plot of the GOES X-ray flux (or particle flux) covering the overall time interval defined in step 2, should appear above the table - this is generated by a call to CoSEC.
- Select one or more event(s) for examination by checking the boxes in the left-hand (select) column of the results table. You can find out more information about the event and active region it occurred in by clicking on the 'Detail' button in the right-hand column of the row of an event.
- Then click on the 'List of Observations' button at the top or bottom of the page to get the list of available data.

Help pages for the EGSO SolarSoft Code Available help: Using the SEC and SFC data

Using the EGSO Feature and Event Catalogues

Demonstration Routines

It is easier to explain how to use feature and event data from the EGSO Solar Feature Catalogue (SFC) and Solar Event Catalogue (SEC) if this capability is demonstrated. Data from these catalogues are ingested in VOTable format.

The routine `sefc_test3` does not assume an internet connection and requires that all data have been gathered ahead of time - this can be done by visiting the Web pages for the SEC and SFC Servers, etc. The routine also assumes that all the required data, including VOTable files (containing feature and event data) and selected FITS files, are stored in a single directory. The directory is addressed by the environment variable `TEST_SEFC_DIRS` - this can be a string defining multiple directories if desired. The VOTable files are read using the routine `get_votable`, the FITS file with `egso_get_fits`. Test data to demonstrate this routine, including images, are included in the EGSO distribution, so you should just be able to type:

```
sefc_test3
```

Keywords: `!select_image` allows the user to select a new reference image; `!set_zoom` zooms the field of view to 8x8 arcmins centred at the cursor selected position; `!reset_zoom` resets the zoom; `!new_dataset` allows the user to select a new dataset directory (if there are several defined).

The routine `qlk_sefc_test2` retrieves data on the fly and requires that you are connected to the internet. The routine allows the user to identify a FITS image of interest (currently limited to 2002) using the EGSO Quicklook Catalogue (QLK) and then submits requests to the SEC and SFC Servers for event and feature data matching the image. The only required parameter is the desired date - a default value is assumed if none is specified.

```
date = '14-jul-2002'
qlk_sefc_test2 [, date]
```

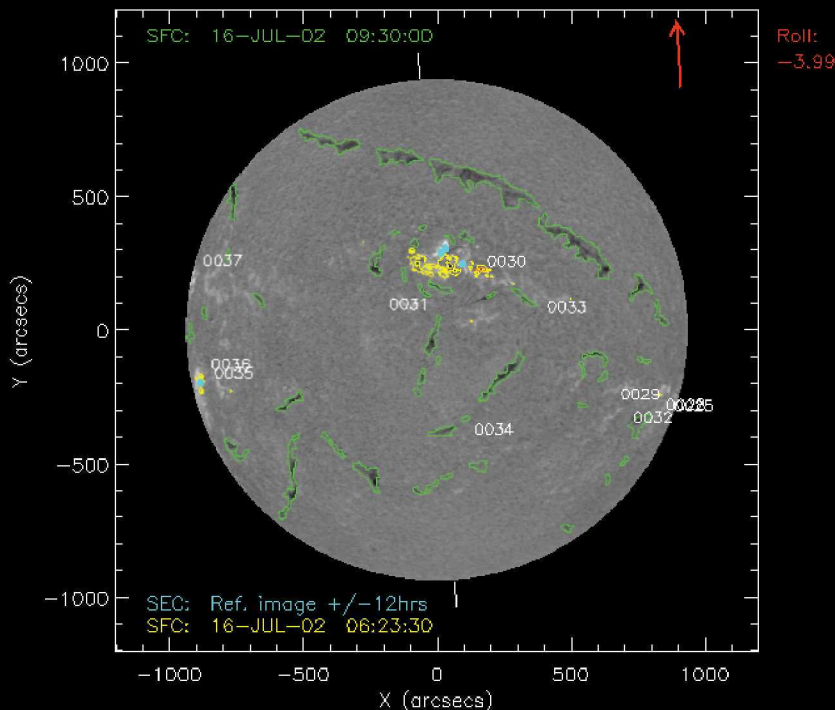
Note: Although the coverage of the SFC between 1996 and 2004 is very good, it is not 100% - it is possible that a selected date might not have any data in the Catalogue. If you find anything you do not understand with the data, contact: fr-egso@brad.ac.uk

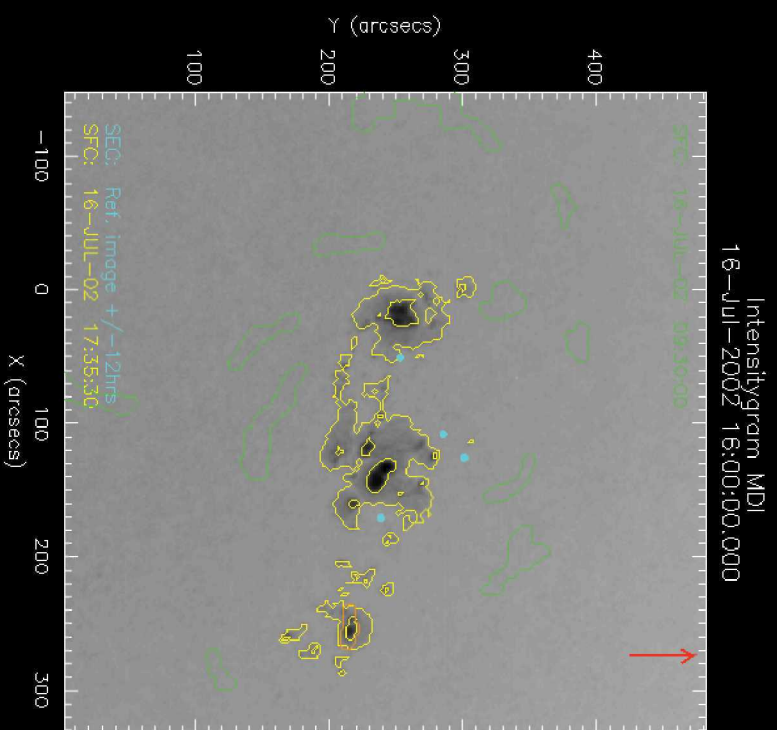
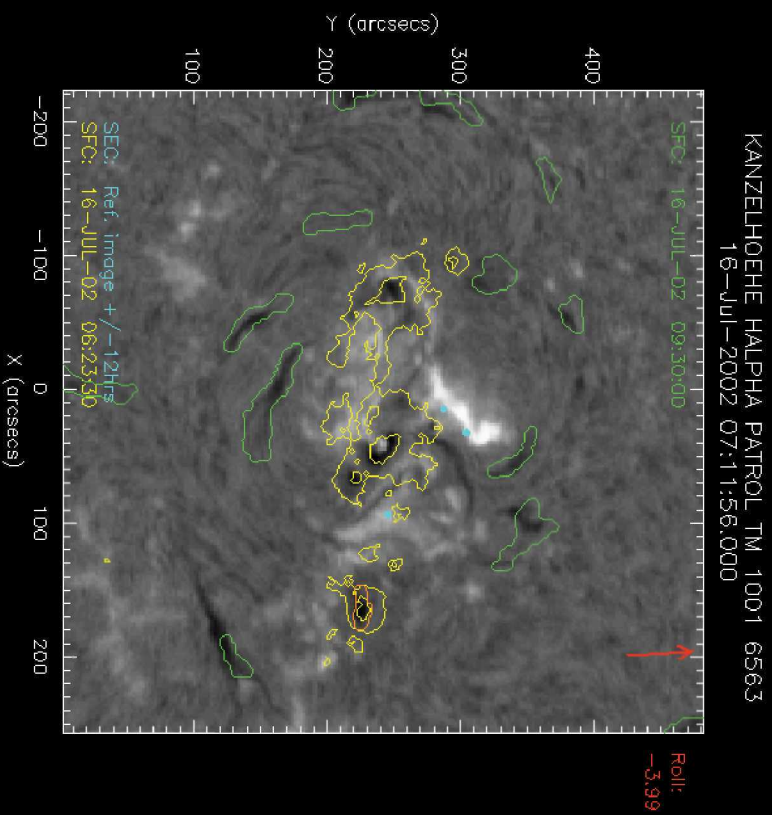


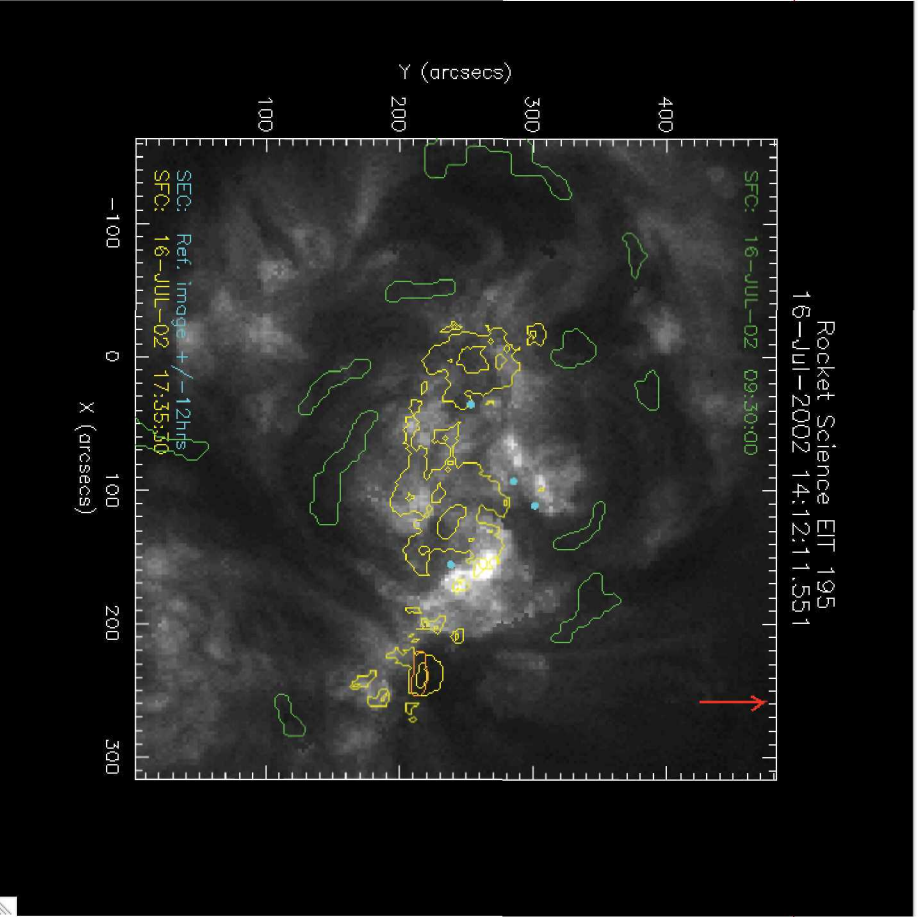
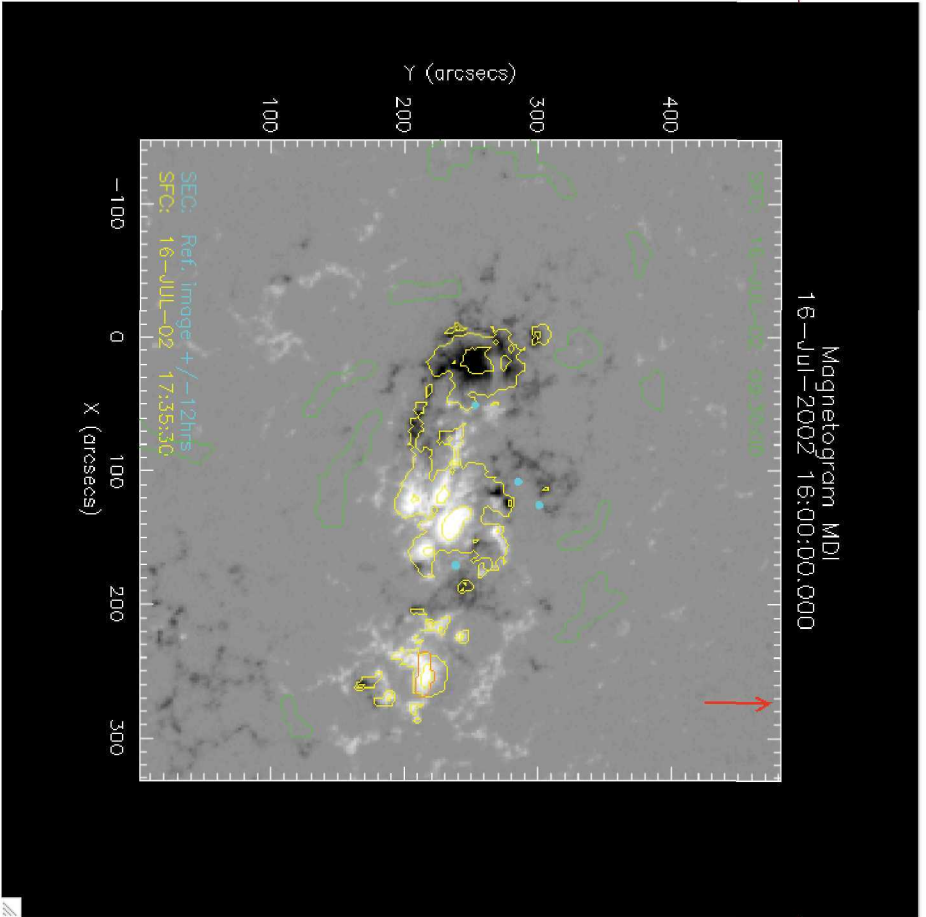
KANZELHOEHE HALPHA PATROL TM 1001 6563
16-JUL-2002 07:11:56.000



European Grid of Solar Observations









So, where are we?



- **Flexible GUI using selectable portlets has been deployed**
 - User able to conduct date driven and event driven searches
 - Event driven search accesses EGSO Solar Event Catalogue
 - System searches for datasets that match search criteria
 - User able to make selections at each stage of the search with aid of supporting data
- **Many popular data sources are being integrated**
 - More than 50 datasets from 20 observatories
 - Space-based: Yohkoh, SOHO, RHESSI, TRACE, SMEI, GOES/SXI...
 - Ground-based:
 - Optical: NSO, Meudon, Pic-du-Midi, Global H α Network, MLSO...
 - Radio: Nobeyama, Nancay
- **Search Registry (SR) is still being developed**
 - Simplified version currently installed in GUI
 - More complex version in preparation (Release 4b)
 - Fully functional Search Registry will allow comprehensive selection of the types of instrument, data, region observed, etc
 - Interoperable with the STP & heliospheric observation



So, where are we? (cont...)



- **Special Servers are becoming operational**
 - SEC Server is fully integrated into EGSO
 - Web Service accessible relational database - SQL in, VOTable out
 - Several lists are already included: Flares, particle events, CMEs...
 - DSO and SFC Servers functional, but integration not yet complete
 - Both available as Web and Web Service interfaces
 - Over 9 years of data processed for Solar Feature Catalogue
 - QLK Server is still under development
- **Processing using CoSEC partially integrated**
 - GOES X-ray & energetic particle lightcurves generated for GUI
 - In process of adding ability to:
 - generate quicklook images & movies
 - extract and process certain datasets (e.g. TRACE & RHESSI?)
 - create composite plots for use in searches and publications
- **EGSO API deployed for use through SolarSoft**
 - Seen as way to develop new capabilities
 - Java command-line capability also developed

- **Use cases for a virtual observatory are hard to define**
 - Users do not understanding the potential of the VO
 - In ESGO, we had to add cases to extended system requirements
- **The purpose of the virtual observatory is hard to explain**
 - Users remain unsure of the need for the VxOs
 - Same problem as before - understanding the potential
 - Currently, main purpose is to save users time
 - Not having to search for data leaves more for research
 - For future projects, management of data within the VO structure will be very important
 - Too much data to move around, easily assimilate
 - Need to provide ways into the data for users
- **Standards are a good idea, but you need to keep pushing**
 - Data models (FITS headers), Registries, databases...
 - Extremely important for interoperability between VOs
 - Will make things more compatible in the future
 - Funding agencies need to encourage this in projects

- **When building a VO, the combined need for expertise in the discipline and IT (information technologies) is a problem**
 - Communications can be difficult
- **Pushing the edges of technology can be frustrating**
 - Considerable evolution in technology over 3 years
 - In the next layers, semantics will be important
- **Be flexible, things change and can bear fruits.**
 - The concept of the Special Providers has proved useful
 - Hides complexity from the system; provides independent resource
 - Multiple instances of the Roles and Special Providers have produced a resilient system
 - Two layer Search Registry proving quite powerful
 - The Applications Programming Interface (API) gives enormous potential to use EGSO capabilities from other systems

- **EGSO can be accessed through:**
 - **Web Page Interface**
 - Main system and Special Providers have own interfaces
 - **Application Program Interface (API) using Web Services**
 - Includes access from SolarSoft/IDL
 - Access to Special providers functional
 - Main system interface under test
- **Useful URLs:**

The different parts of the EGSO system, including the main entry portal, can be accessed from URL:

<http://www.egso.org/software>

General information about EGSO can be found under URL:

<http://www.egso.org>