esa

SpaceGRID

EAE03-A-10724; GI4-1TH3P-1542

C.H. Perry[1], A.P. McDermott[1], P.J. Richards[1], R. Stamper[1],
C. Morris[2], E. Purvis[2] and S. Beco[3]

[1)] Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX, UK, spacegrid@rl.ac.uk
[2)] SciSys, Chippenham, Wiltshire, SN14 0GB, UK
[3)] DATAMAT S.p.A., Via Laurentina, 760 - I-00143 Rome, Italy

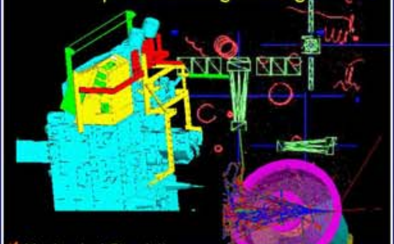
ABSTRACT

The European Space Agency's SpaceGRID study has been aimed at identifying the potential benefits of "Grids" to the Solar System Research Community and defining a road map for the possible use of this technology within ESA's future programme.

Grid technology is an emerging computing infrastructure that is intended to provide uniform access to a set of distributed resources such as computational systems, data-archives or shared facilities constituting a collaborative working environment.

In this presentation we review the outcome of the study and in particular consider the prototyping activities that have been taking place over the last six months. These have been based around the test federation of several heterogeneous Soar and STP data sets using service based interfaces. A simple XML schema has been developed for metadata handling and the VOTable format (http://cdsweb.u-strasbg.fr/doc/VOTable/) used for standardised exchange of tabular data between the participating systems. We look at the lessons learnt from this study and consider how a large scale deployment of this technology might be applied.

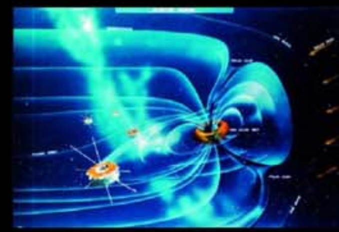
Spacecraft Engineering



- Mechanical Simulation Thermal control, structural and aero-thermodynamic engineering analyses will benefit from multi-disciplinary and collaborative engineering.
- Radiation Transport Monte-Carlo simulation of high energy particle interactions inside a spacecraft system, components or detector with large numbers ($>10^7$) of source particles are candidates for adaptation to a parallel processing system.
- Spacecraft Plasma Interactions Monte-Carlo 3-d temporal electrostatic kinetic Particle-in-Cell method with large number of particles ($>10^9$). Strong coupling between cells. Challenge the development of independent subtasks for parallel processing.

Contact: Hans-Peter.de.Koning@esa.int

Space Weather



- Wealth of science data from different instruments is/will become available.
- Using these distributed data sets in scientific analysis is not trivial.
- Simulation of Sun-Earth connection is computationally intensive
- Collaborative research involving international research teams could be facilitated with an advanced GRID infrastructure.

Contact: Eamonn.Daly@esa.int

esa

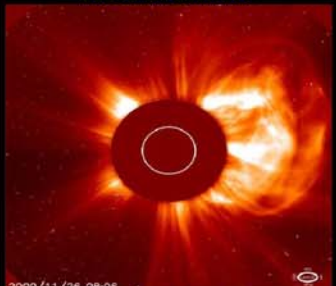
SpaceGRID

A study funded by the General Studies Programme, © ESA, 2001
Contact-point: Pier.Giorgio.Marchetti@esa.int

The goals of SpaceGRID are:

- *To assess how GRID technology can serve requirements across a large variety of space disciplines (spacecraft mechanical engineering, space weather, space science, earth observation).*
- *To foster collaboration and enable shared efforts across space applications.*
- *To sketch the design of an ESA-wide (and common) GRID infrastructure.*
- *To demonstrate proof of concept through prototyping.*
- *To involve both industry and research centres.*
- *Keep Europe up with GRID efforts !!*

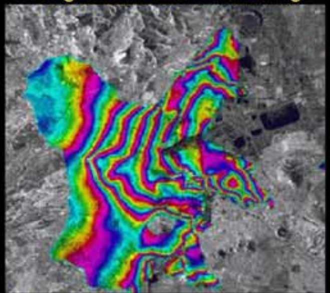
Space science and solar system research activities



- Access to multiple distributed data archives.
- Online data manipulation, analysis and visualisation of complex data sets.
- Interoperability between legacy systems and heterogeneous databases.
- Collaborative environment (e.g. sharing of algorithms for analysis and visualisation).

Contact: Gerhard.Kreiner@esa.int

Earth Observation, e.g., interferometry for ground-motion monitoring



- Analysis of GRID-aware earth and environmental monitoring application infrastructure.
- Extend GRID access to European environmental and earth science application to large science communities - to value adding and commercial communities.
- Demonstrate collaborative environment for EO.

Contact: Pier.Giorgio.Marchetti@esa.int

Solar System Research Domain
Why GRIDs?

- Solar System Research is a multi-disciplinary science
  - Solar Physics, Solar Terrestrial Physics, Planetary
- Complex 3-D environment
  - Phenomena occur over a range of temporal and spatial scales
- Complex set of instrumentation, data and formats
  - Particles, fields, waves and imagers
  - Scalers, vectors, tensors, images, multi-dimensional arrays
- Data processing is often responsibility of PI
  - Heterogeneous data handling systems
- Researchers need to combine and manipulate multiple data sets
  - This is where a Data Grid and collaborative working environment can help

Solar System Research Key User Requirements

- Facilities for locating online sources of data based on a general query
- Standardisation in the delivery format of data/metadata from different sources
- The ability to query the catalogues of multiple distributed data archives
- The ability to query the data across multiple distributed data archives
- The ability to manipulate and process data remotely prior to download
- A web portal to access distributed resources from a single web site
- A Grid server allowing users to link their own data into SpaceGRID
- A software library allowing programs to access to SpaceGRID facilities
- An online collaborative working environment

Solar System Research Technology and Standards

- The "Middleware" is the underlying software providing the building blocks for the distributed system

- Open Grid Service Architecture (OGSA) is an extension to web service technology (XML, SOAP, WSDL) and defines standard semantics and mechanisms for locating and accessing systems in a heterogeneous environment.

- OGSA is standards based, building on work from the Global Grid Forum (GGF) and World Wide Web Consortium (W3C).

- For the Solar System Research prototyping activities the Globus Toolkit 3 (alpha 2, core) implementation of OGSI has been used.

http://www.globus.org/ogsa/releases/alpha/
http://www.w3.org/
http://www.ggf.org/

Solar System Research Technology and Standards

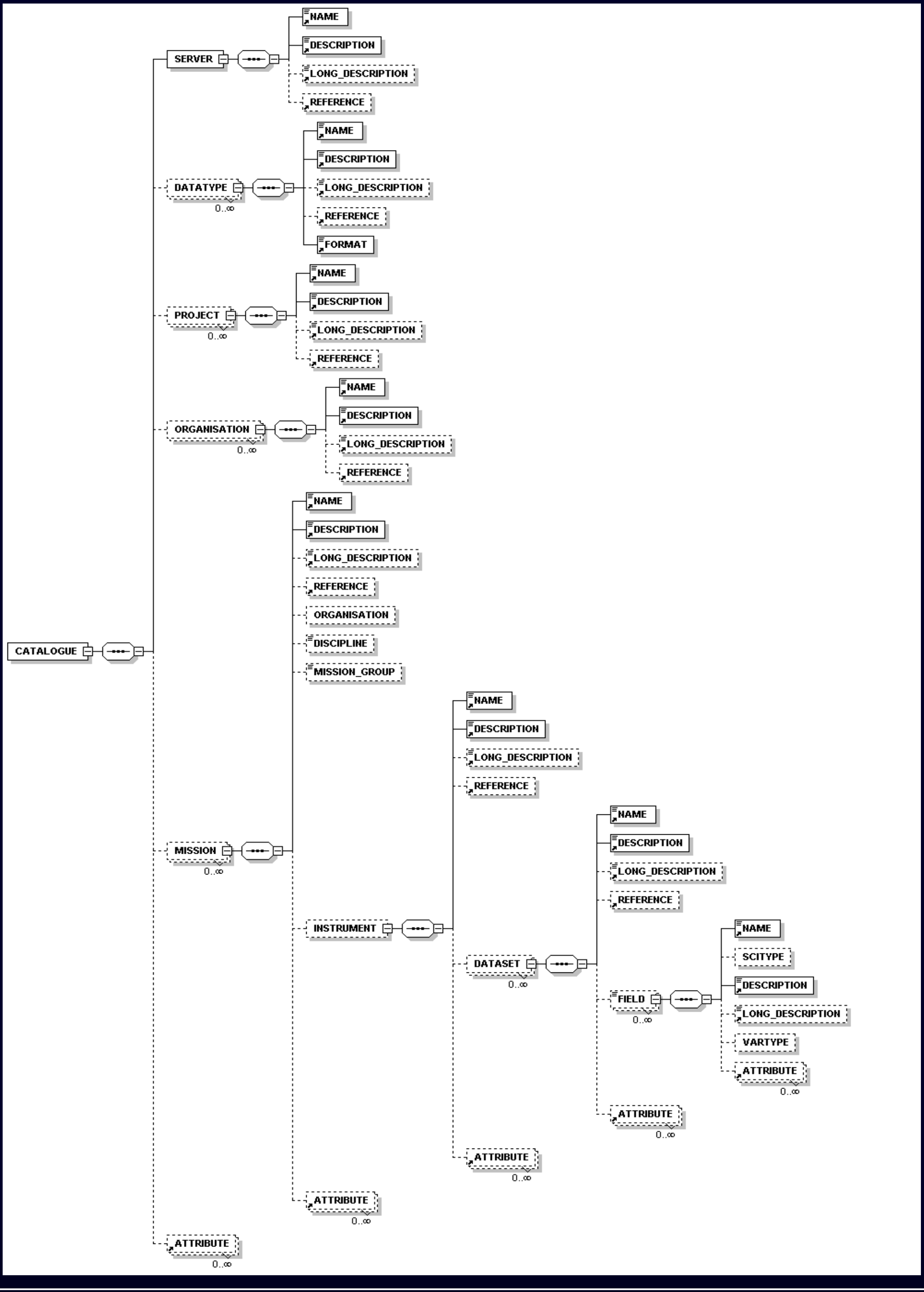
• Standardisation of the scientific Metadata is vital when developing interoperable systems.

• Much work on this is currently underway within the various Virtual Organisation (VO) projects. *The Space Physics Archive Search & Exchange (SPASE) group have a splinter meeting here at EGS (Thursday, 16:30, Gallieni 1) to present work on a data dictionary for space physics. We would encourage interested members of the community to come along and participate.*

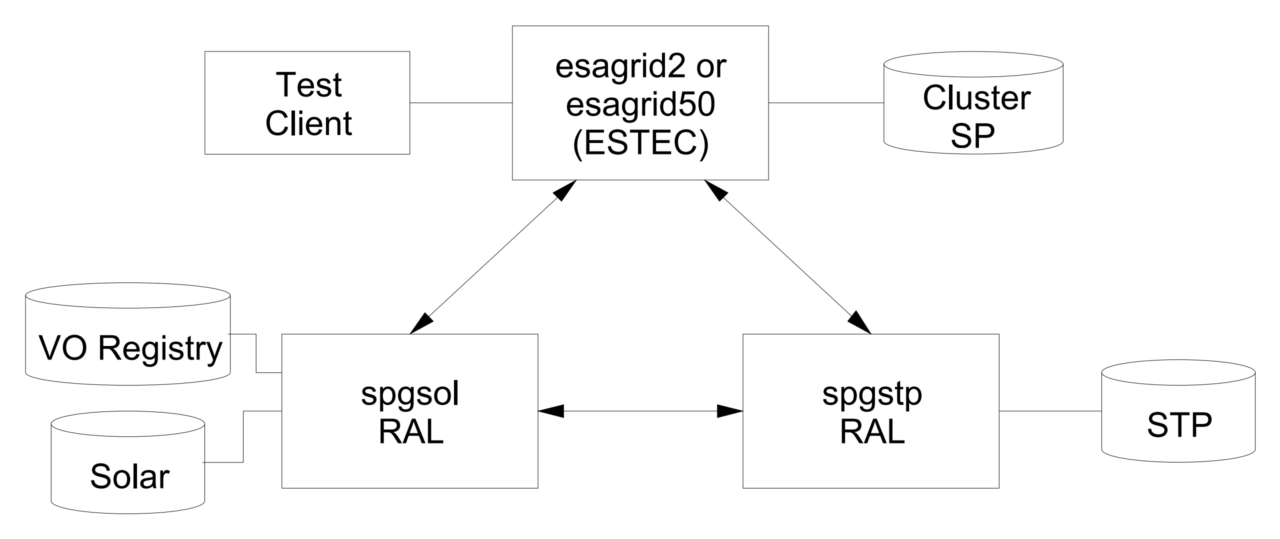
• For SpaceGRID a simple XML schema has been developed. Processing using XML Style sheet Translations (XSLT) has simplified the future migration as standards are agreed.

• In SpaceGRID tabular results are encoded in the VOTable XML format. VOTable has been developed within the Astronomical VO projects and it is not yet ideally suited to handle some aspects of the Space Physics metadata. http://cdsweb.u-strasbg.fr/doc/VOTable/)

Solar System Research
Simple SSR XML Meta Data Schema



Solar System Research Prototype Configuration



SSR persistent factory services provided at each hub:-

- LocalCatalogueFactory –
  Access to local metadata information at one site
- CatalogueMiningFactory –
  Coordination of metadata collection from all sites
- LocalQueryFactory –
  Apply query on local dataset
- QueryMiningFactory –
  Coordinate distributed query
- LocalStorageFactory –
  Manage local storage element
- StorageFactory –
  Coordinate access to distributed storage elements

Solar System Research Test Data Federations

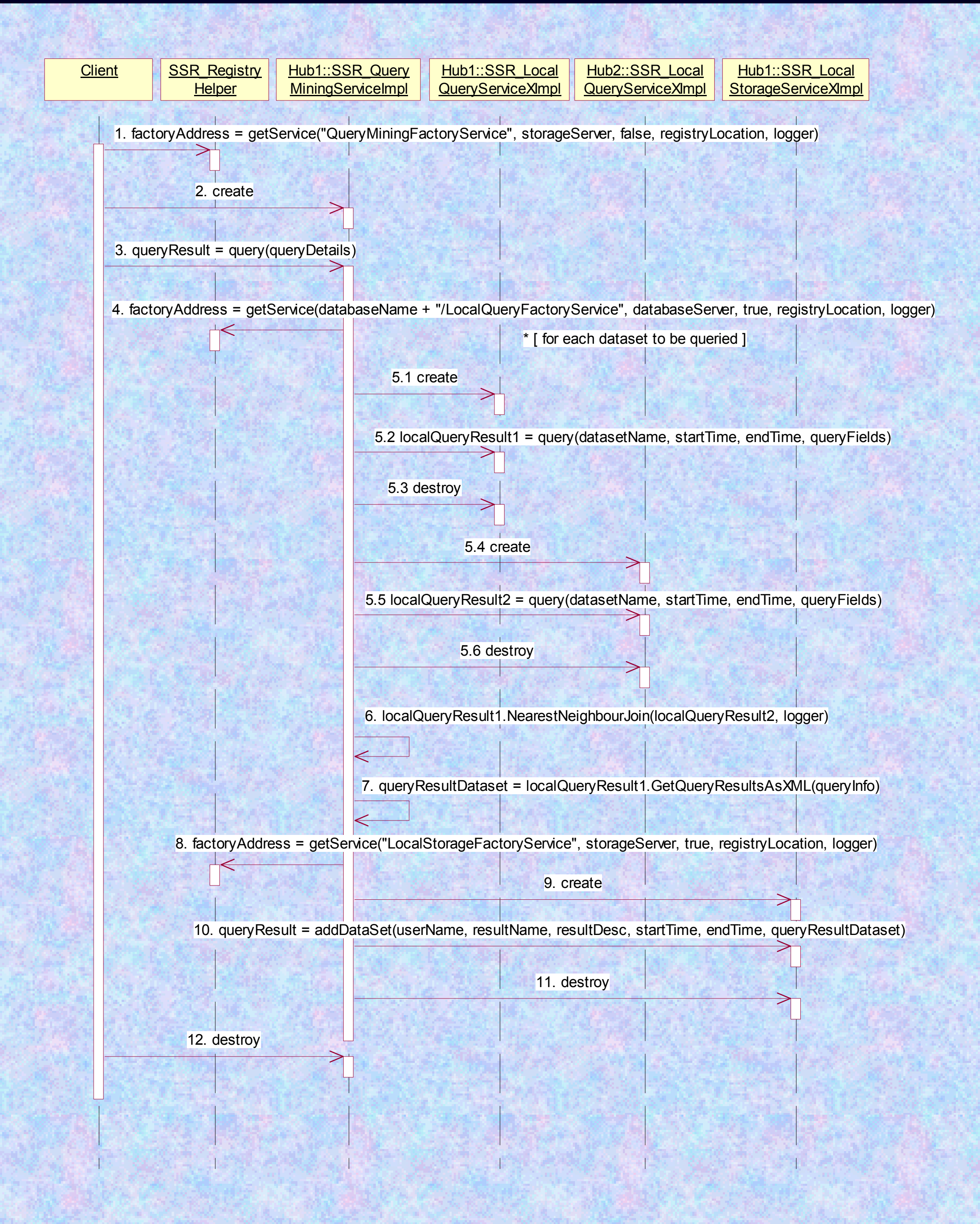
Data Federation	Fields (entries)	Server:Port
Cluster Summary Parameter Data	91	195.169.140.11:8080 (esagrid2)
Solar	52	130.246.35.161:8080 (spgsol)
World Data Centre	134	130.246.35.162:8080 (spgstp)

Cluster Summary Data – One minute resolution data from most instruments on one of the four Cluster spacecraft. Auxiliary data providing spacecraft locations and configuration are also included.

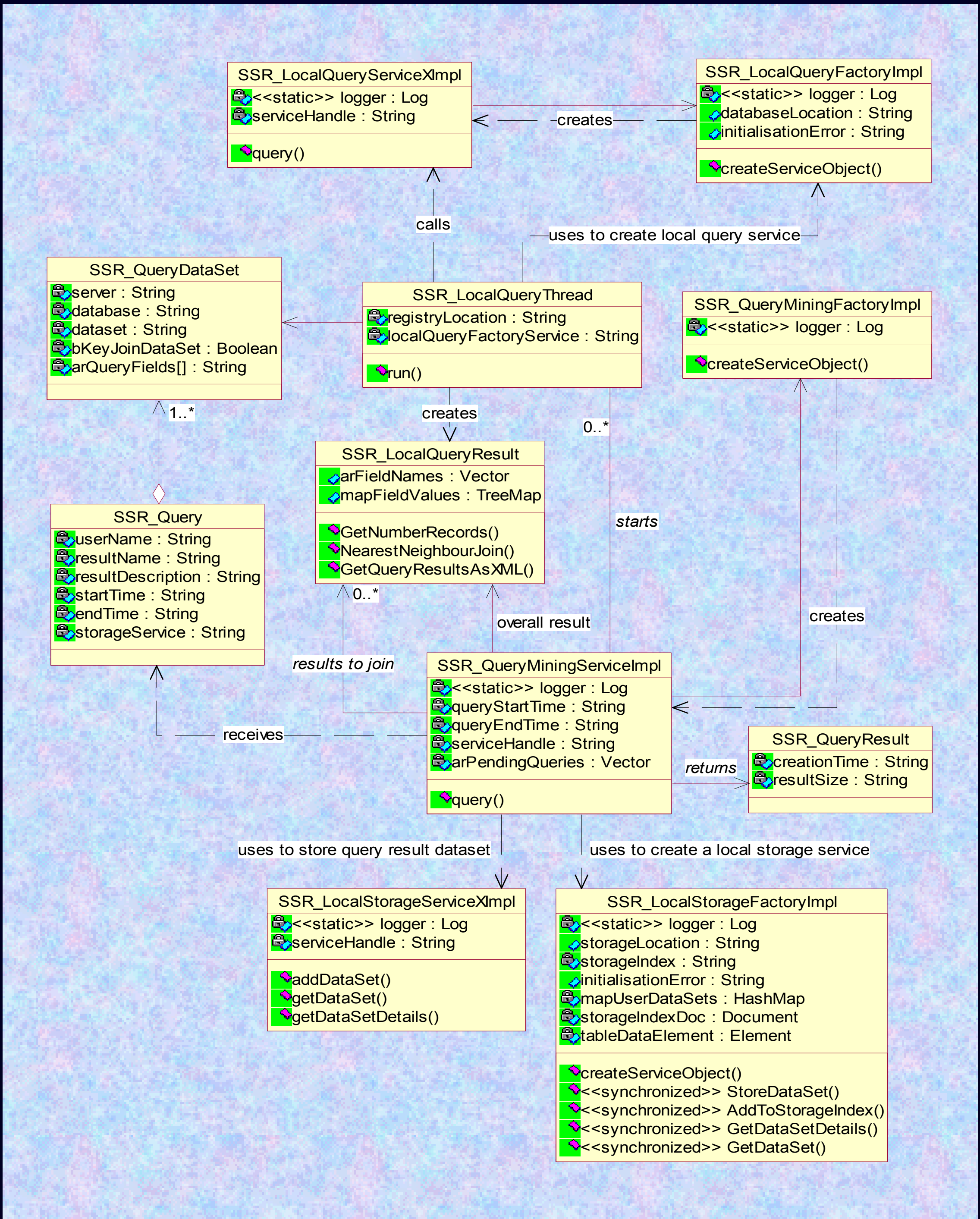
Solar – Summary image catalogues for the EIT and LASCO instruments on SOHO. Event catalogues covering LASCO CME, HXT Flare, BATSE flare and Solar Energetic Proton events.

WDC – A set of geomagnetic and other STP related indices including AE, Dst, F10.7 flux and OMNI parameters.

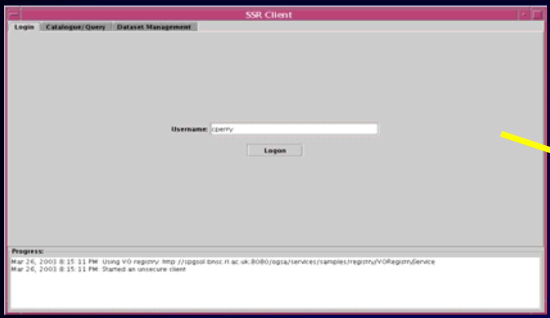
Solar System Research Query Related Sequence Diagram



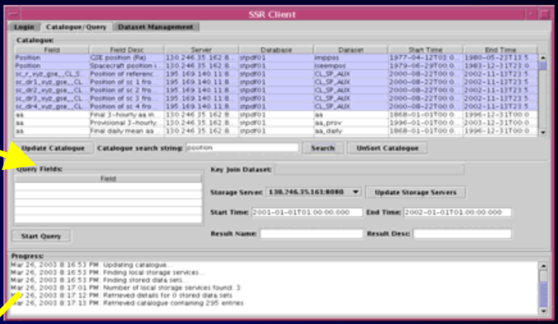
Solar System Research Query Related Class Diagram



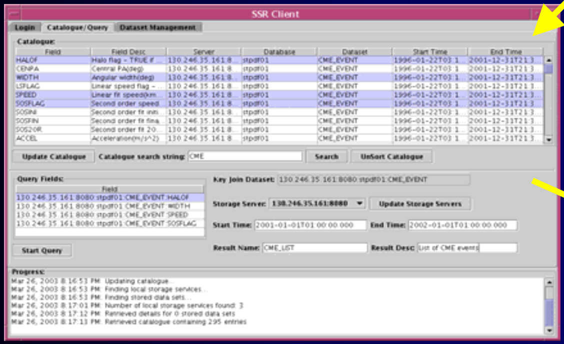
Solar System Research Prototype System (Java Interface)



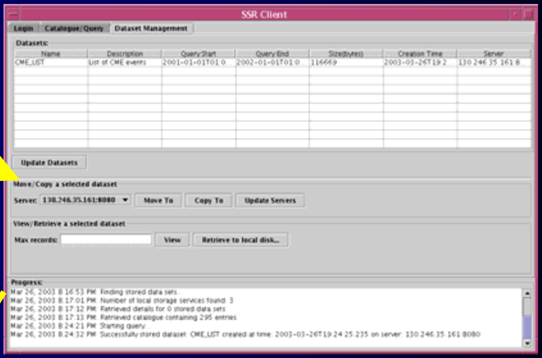
Single sign-on to all SSR services (just one password)!



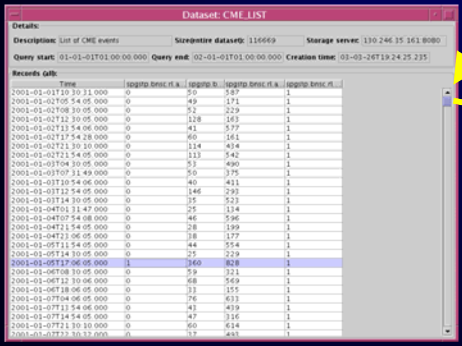
Search for available parameters from all registered archives



Select parameters. Query is split and sent to host archives. Results are returned and joined onto a common timeline.

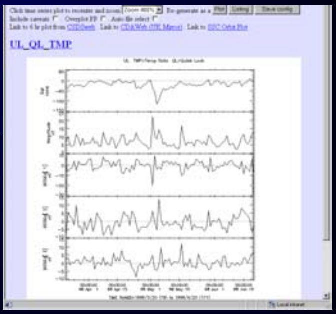


Results placed on Grid storage element. User can move or copy data between storage elements.



Tabular view & download data to local machine.

Under development:- service to plot results.



Solar System Research
Key Features of Prototype

- Single sign on. User does not have to login separately to each resource. This requires some level of coordination between participating archives.

- Resource discovery is via a central VORegistry. Services can be added at which point they become visible to the virtual organisation.

- Queries are automatically split by dataset and directed to the appropriate archive.

- Query results are consolidated into a single table (currently using a simple nearest neighbour join).

- Query results are stored on storage element within the virtual organisation for subsequent retrieval or further processing

Solar System Research Lessons Learnt

• The “Middleware” technology is a rapidly evolving target and still has a steep learning curve.

• Web/Grid services provide a standardised replacement for current CGI based mechanisms.

• Using this technology it has been possible to build a distributed systems that demonstrates both the concepts and identify the potential bottle necks.

• Basic integration with legacy archive systems can be straightforward and can run in parallel with the existing archive access mechanisms.

• These sorts of distributed systems open new opportunities for scientists working with legacy archives.

•Standardisation of the metadata is crucial and requires international coordination e.g. SPASE and other VO work.

Solar System Research
Future Work

- Future development of the core services will remove some of the existing bottlenecks, improve handling of large files and extend the query capabilities.

- Addition of simple plotting capabilities to allow preview of the data prior to download.

- Demonstration of how the services can be used within a dedicated science tool. A simple Sun-Earth connection tool is envisaged that will build on the data sets already federated within the prototype.

- Real world application of this technology within a science project. One area under consideration is the interoperability component of the recently approved ESA Cluster Active Archive.