ESA SpaceGrid

Demonstration of Grids in the Solar System Research Domain

Lessons Learnt

Chris Perry, RAL, c.h.perry@rl.ac.uk

(A.P. McDermott, S. Dunkin, C.D. Pike, P.J. Richards, R. Stamper, B.Stewart at RAL and C. Morris & E. Purvis at SciSys)

SpaceGRID Final Presentation, ESRIN 21st May 2003

Overview

- **Introduction**
- **Solar System Research** - Why GRIDs?
- **Community Consultation** - User Requirements
- **Service GRID** - Technology and Standards
- **Solar System Research Demonstrator**
  - Overview
  - Design and Development (see next talk by C. Morris)
  - Data Federations & Local Query Service
  - Interface
  - Key Features (user perspective)
- **Lessons Learnt & Conclusions**

Solar System Research Domain Introduction

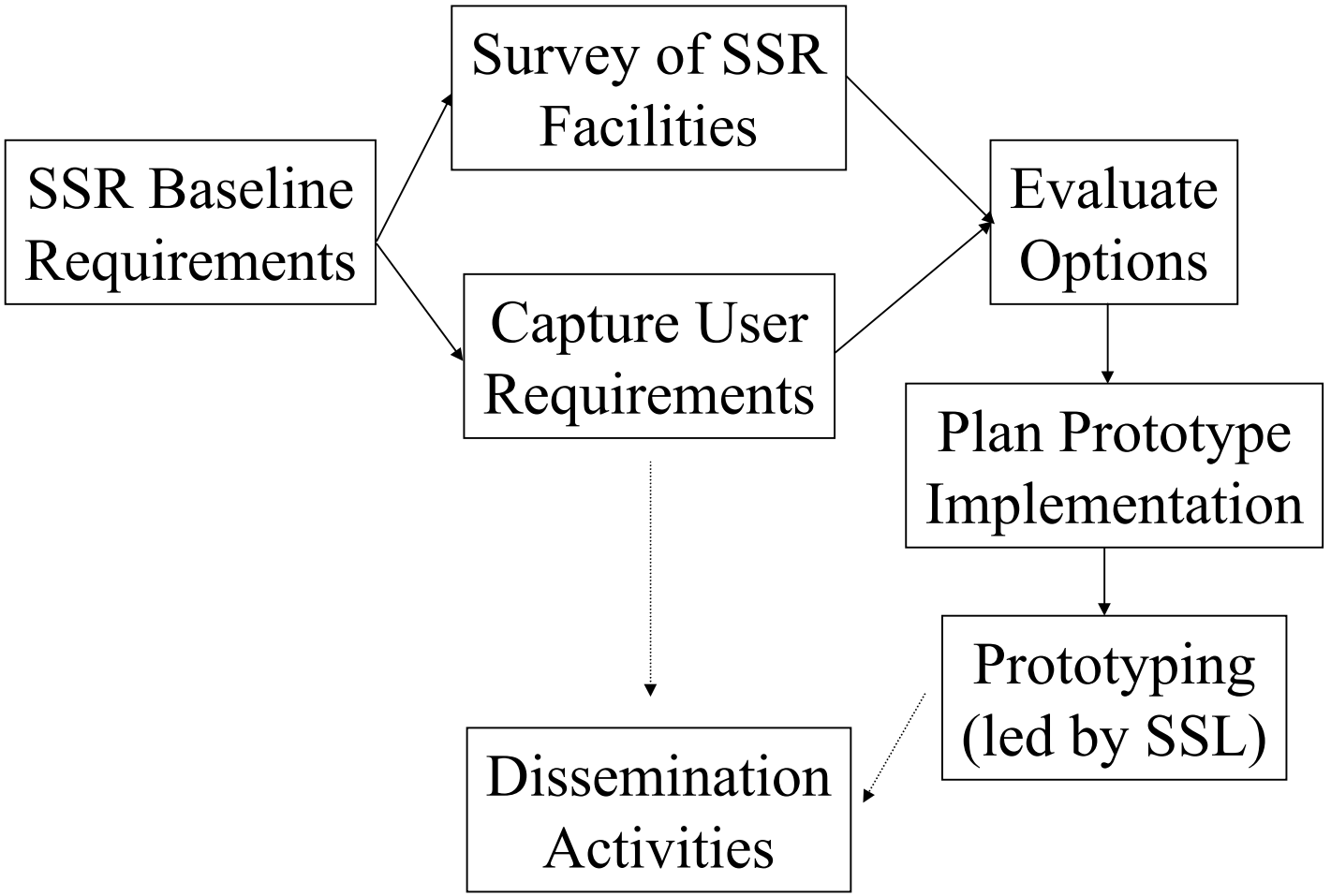
• Solar System Research (SSR) is one of the strands of the ESA funded SpaceGRID study.

• The SpaceGRID study consortium has been led by Datamat. RAL provide SSR domain expertise and SciSys led the prototype implementation.

• The main goal has been to assess how GRID technology can serve user requirements in the SSR domain.

• Demonstrate proof of concept through prototyping and assess what this means in terms of an ESA-wide (and common) GRID infrastructure.

Solar System Research Activities



CCLRC Rutherford Appleton Laboratory

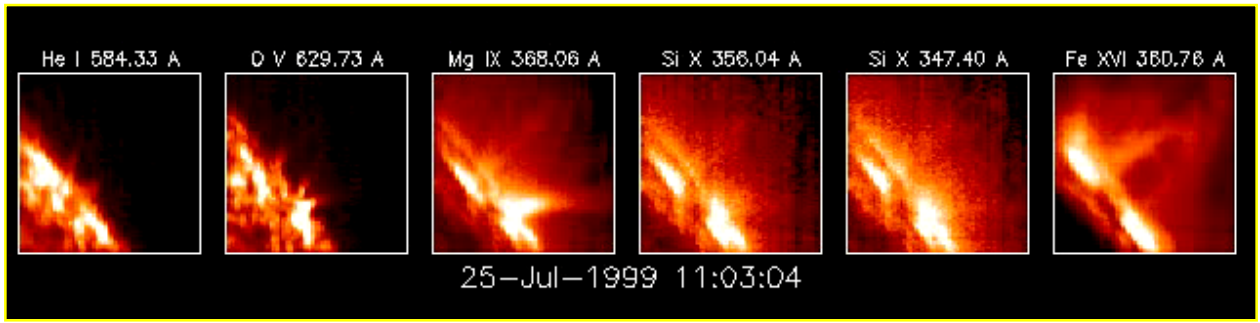
Solar System Research Domain
Why GRIDs?

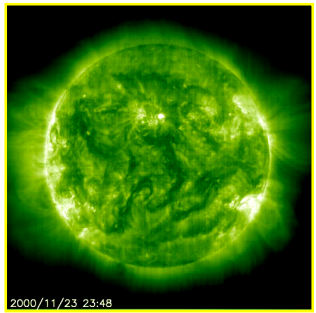
- Solar System Research is a complex multi-disciplinary science
  - Solar Physics, Solar Terrestrial Physics, Planetary

- 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
- Researchers need to combine and manipulate multiple data sets
  - This is where a Grid technology, standardization and the resultant collaborative working environment can help

Solar System Research Domain
Why GRIDs?

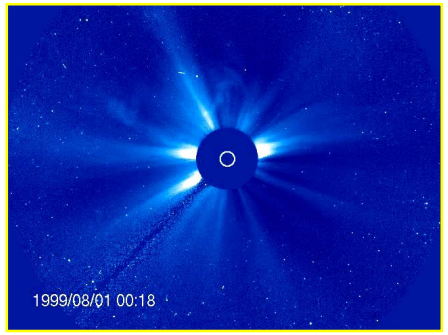
- **Solar Physics**
  - Complex spatial and temporal variations -> helioseismology, CME and flare events.
  - Source of the mass, energy and momentum -> input to STP
  - Main ESA mission is SOHO but complementary data from other sources such as YOHKOH, TRACE and ground based observations.
  - EU project, European Grid of Solar Observations (EGSO), aimed at addressing some of these issues. See presentation at this meeting.

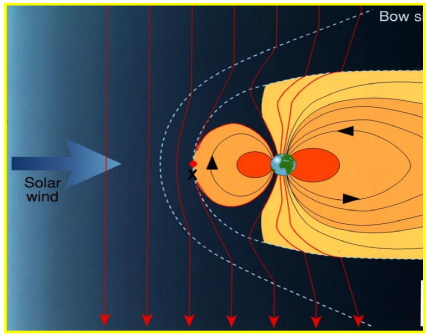


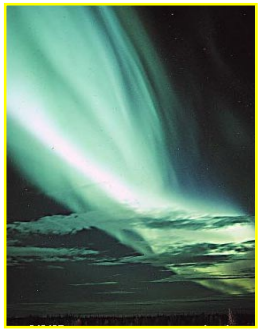


Solar System Research Domain
Why GRIDs?

- **Solar Terrestrial Physics**
  - Complex spatial and temporal variations -> storms, sub-storms, reconnection
  - Main ESA mission is Cluster (also joint China/ESA Double Star). But many other support data sets needed (e.g. ACE, groundbased)
  - Data processing is often responsibility of PI
    - Heterogeneous data handling systems
    - Data assimilation and visualization particularly difficult
  - Use of coupled models

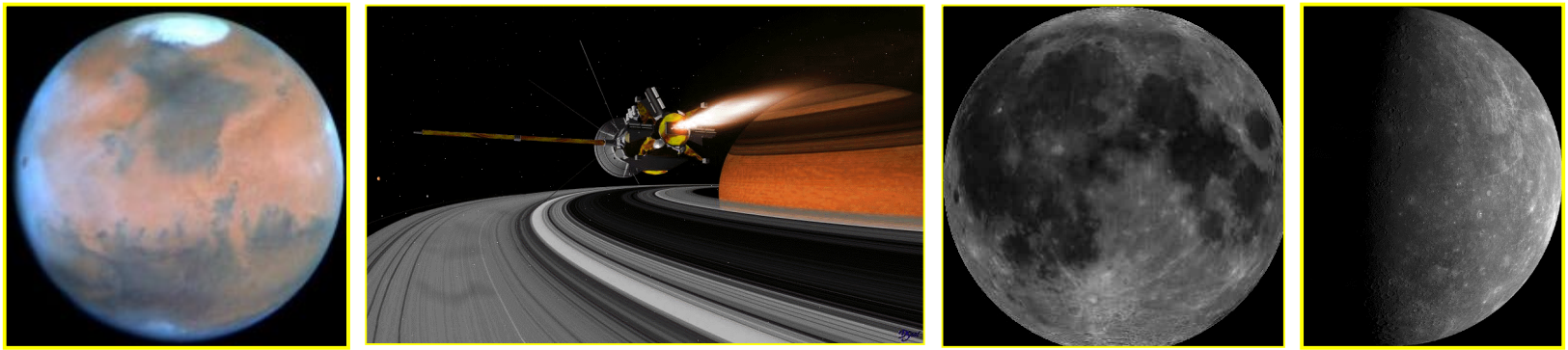






Solar System Research Domain
Why GRIDs?

- **Planetary**
  - Not much current European data (Giotto, Ulysses)
  - But… ESA has an active future planetary program including:-
    - Mars Express, Cassini/Titan, SMART-1 (Moon), Venus Express and Bepi Colombo
  - ESA are in the process of setting up a Planetary Archive based on the NASA PDS standards
  - Comparative planetology, magnetospheres and solar system evolution are high profile, cross domain topics that will benefit from inputs from other areas of SSR but also from other domains such as Earth Observation.



Solar System Research User Requirements

- Community consultation via online questionnaire
- Results and discussions held at MIST, EGS and at project meetings
- 113 responses from 20 countries
- 55% Solar, 36% STP and 9% Planetary
- Broad spread of interests including:
  - planetary interiors & surfaces, planetary and terrestrial atmospheres, ionosphere, aurora, magnetosphere, solar wind, solar terrestrial, interaction, solar flares, corona, solar atmosphere and helioseismology.
- Most people who responded are working on data analysis though significant minorities involved in modelling, operations and software development.

Solar System Research User Requirements

- A few of the difficulties highlighted in ***data access***:-
  - Locating sources of data or models
  - Ascertaining data availability for multiple data sets
  - Limited search capabilities
  - Multiple requests to different archives, interfaces, access control

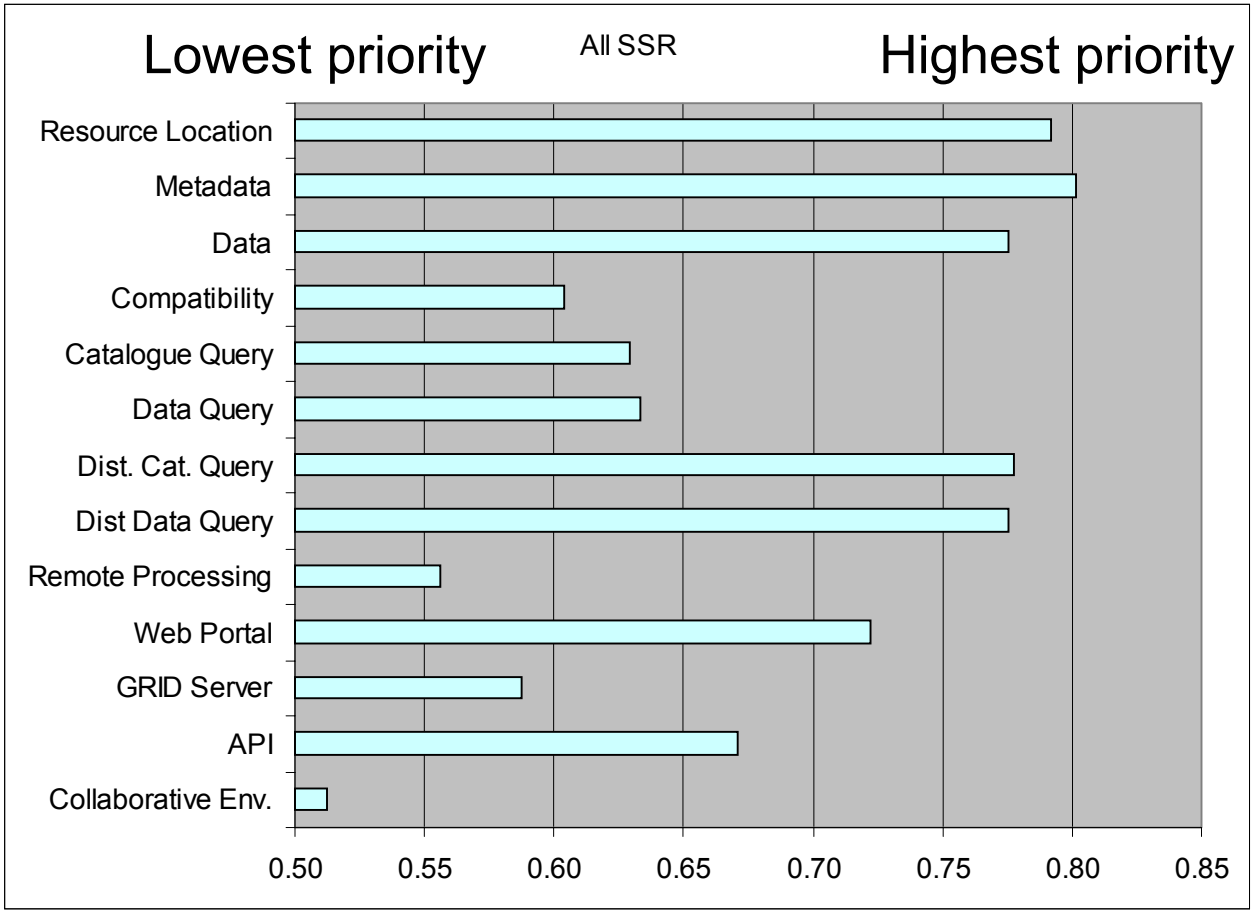
- Main problems ***combining data*** from instrumentation and models:-
  - Lack of standardisation in data specification.
  - Inadequate, incompatible or unreliable metadata.
  - Need for re-sampling temporal or spatial data.
  - Lack of documentation -> incorrect use of available data.

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
Key User Requirements

- User prioritisation of key requirements



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 implementation of OGSI was selected.

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 such as those co-ordinated under IVOA and SPASE.

- SPASE - Space Physcis Archive Search and Exchange
  - Working group involving international Space Physics data centres GSFC (ISTP), JPL (PDS), APL (LWS), SWRI in the US, CDPP and RAL in Europe.
  - Developing standards (e.g. a data model and metadata dictionary) for the Space Physics Research domain.
  - Standards provide the basis for system interoperability and are critical to the success of e-science/Grids.

CCLRC Rutherford Appleton Laboratory

Solar System Research Metadata Standards

- IVOA - International Virtual Observatory Alliance
  - Alliance of astronomical VO projects, initially AVO, AstroGrid and NVO. Now many other partners.
  - Developing standards for Astronomy domain but much of the work is generic and could be applied across domains.

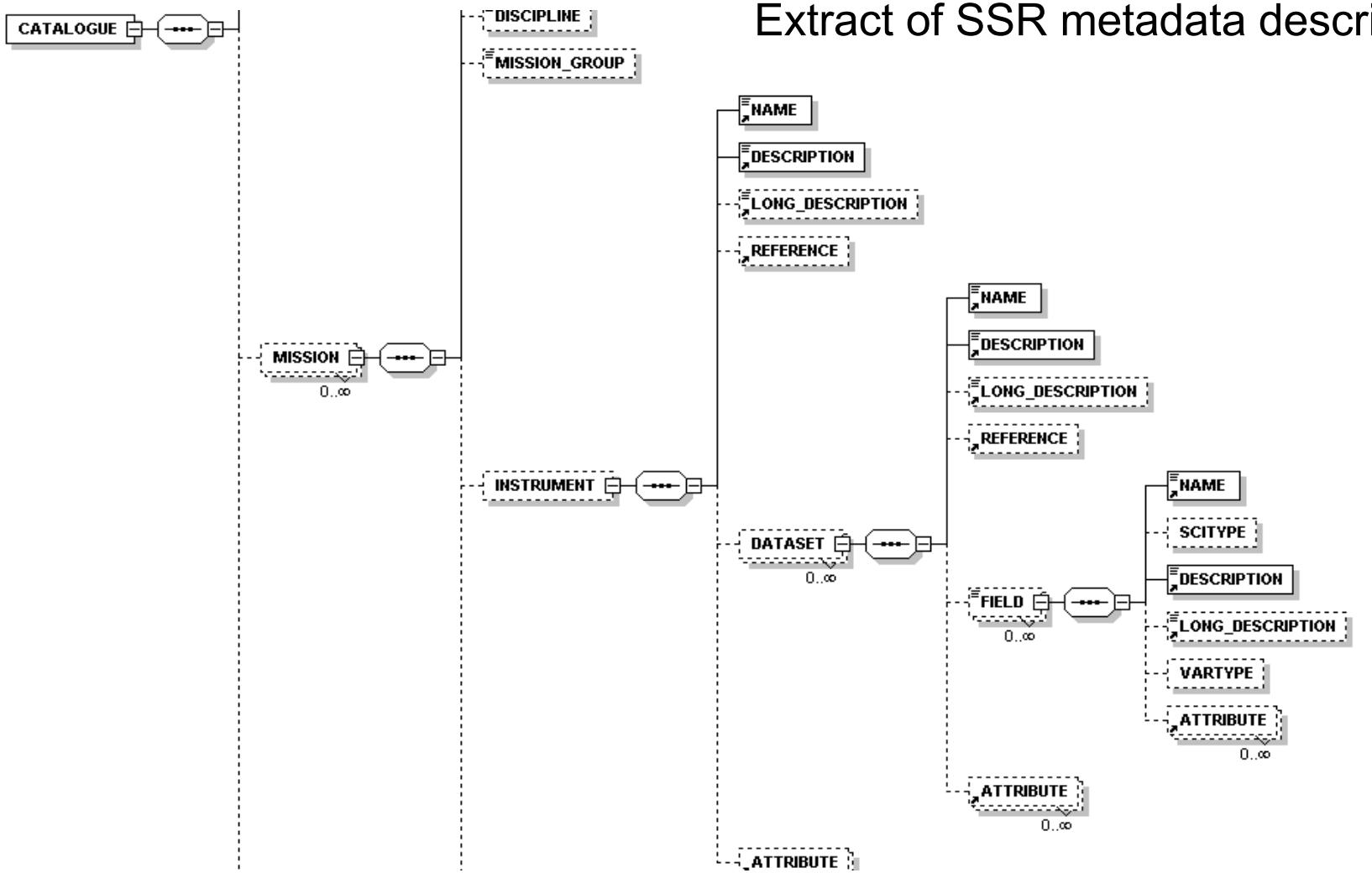
- Outputs from these groups not available in time for SpaceGRID work.

- 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. Metadata support is not yet ideally suited to handle some aspects of the Space Physics metadata. http://cdsweb.u-strasbg.fr/doc/VOTable/

Solar System Research Technology and Standards

Extract of SSR metadata description



Solar System Research
Prototype Implementation Plan

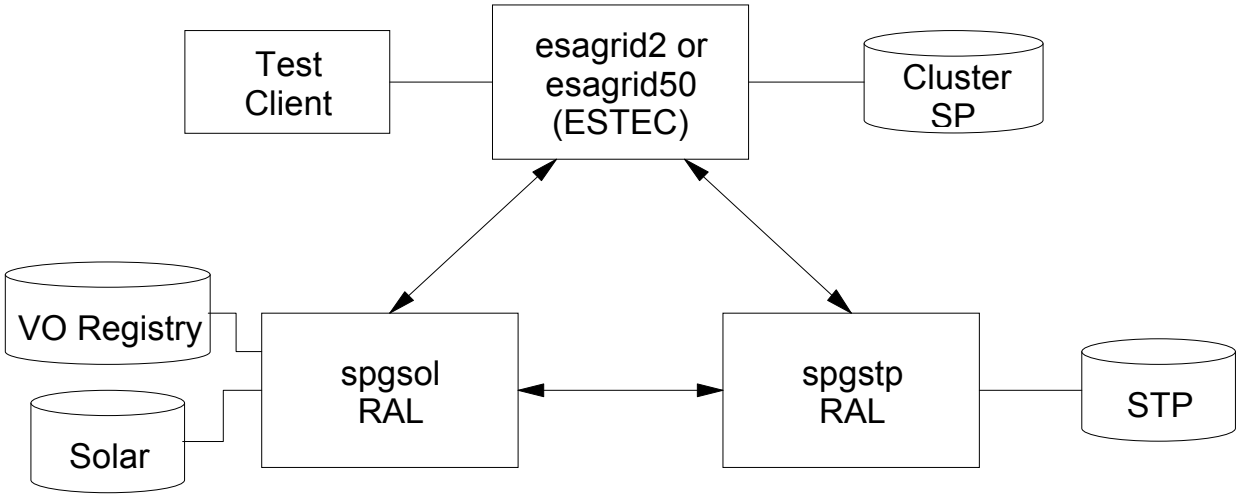
- The aim of the prototyping was to address the key user requirements through the application of Grid technology
- It would involve the development of a new infrastructure based on the OGSI technology rather that Grid enabling an existing application
- The main functions to be provided were:-
  - Login to the system
  - Generate, execute and monitor distributed queries
  - Apply a limited set of data manipulation functions
  - Preview the data and metadata of queries
  - Download the results from queries

Solar System Research Prototype Implementation Plan

• Due to the relatively short timescale (~6 months) for the design and development of the prototype a number of constraints were agreed.

- Only a small number of data sets would be federated.
- No ability for user upload of data or code would be provided
- The full functionality of the query and processing systems would not be implemented (only a sub-set of the functions and operators would be provided, and would only work with simple parameters). The query function would be limited to the selection of a time range and a list of parameters to be returned.
- Distributed queries and processing would not be optimised.
- Data preview would be limited to simple time series data types (scalars and vectors) displayed in a simple tabular view.
- There would only be a basic implementation of the portal application.

Solar System Research Prototype Implementation Plan



- The proposed configuration would be installed on three nodes to allow distributed access to be adequately tested.

- See next talk for details of the service design and implementation.

CCLRC Rutherford Appleton Laboratory

Solar System Research Test Data Federations

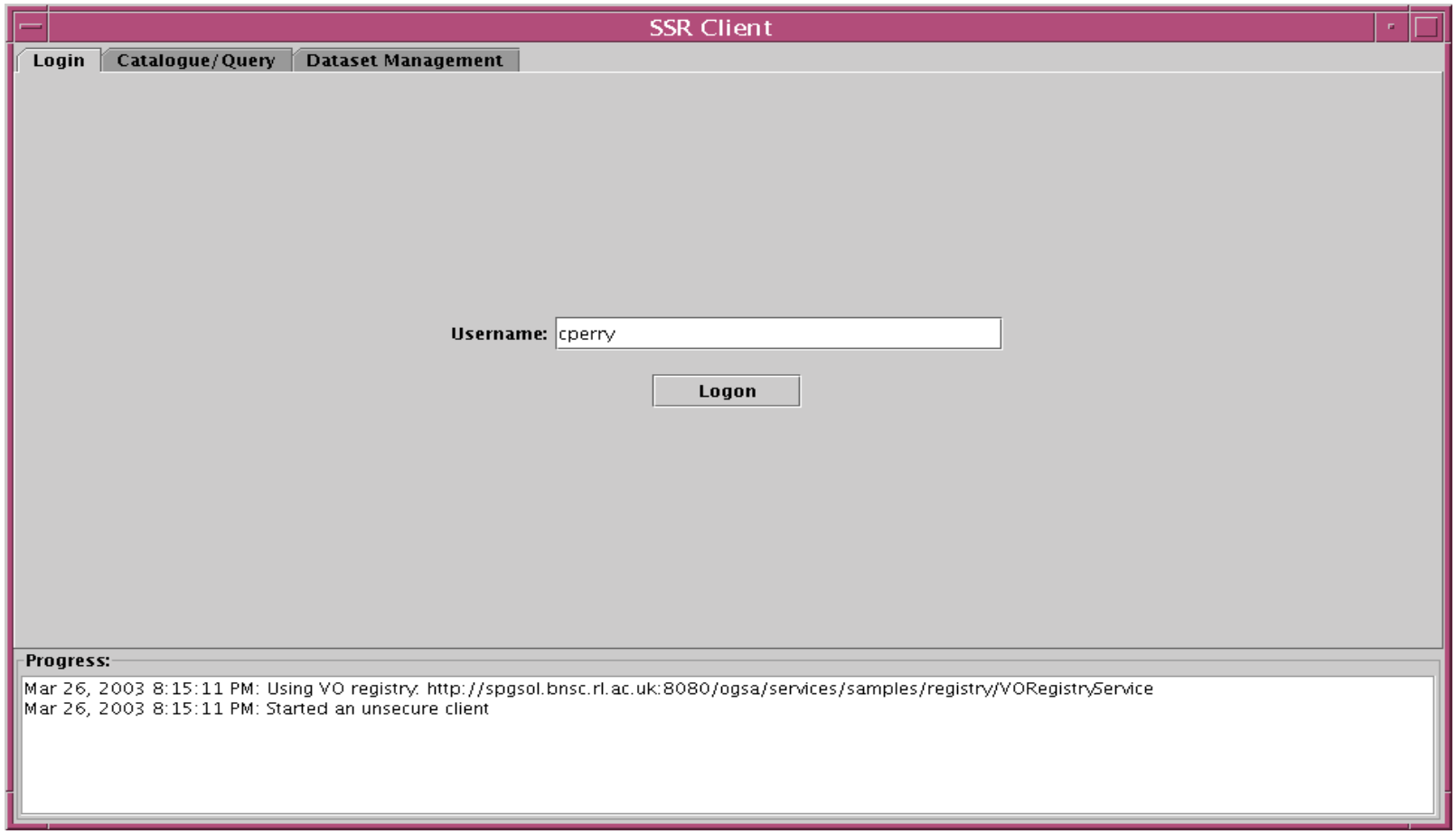
Data Federation	Fields (entries)	Server:Port
Cluster Summary Parameter Data	91	195.169.140.11:8080 (esagrid2)
Solar	52	130.246.35.161:8080 (spgsol)
World Data Centre	134	130.246.35.162:8080 (spgstp)

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.

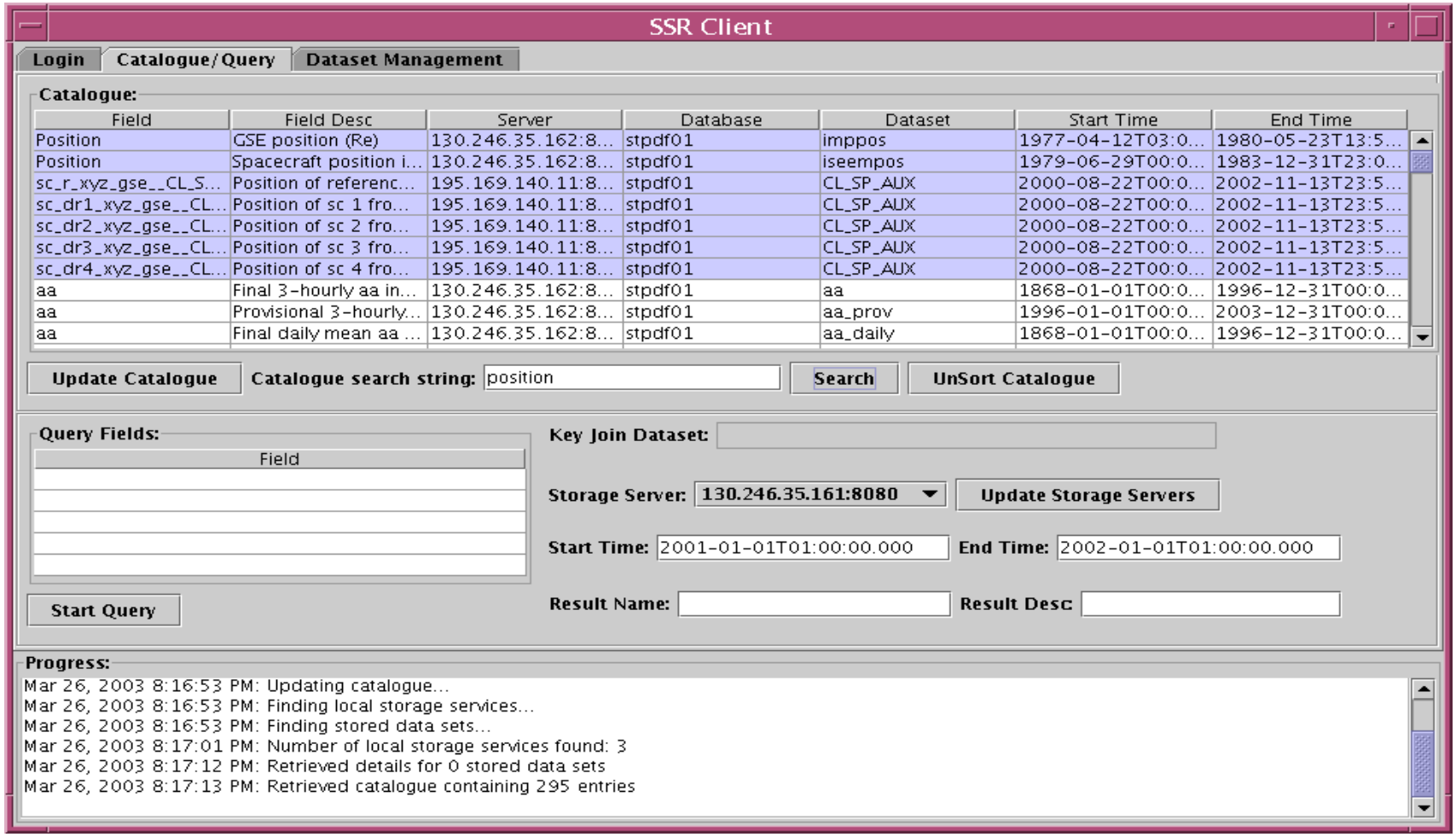
WDC – A set of geomagnetic and other STP related indices including AE, Dst, F10.7 flux and OMNI parameters.

Solar System Research Example



Single sign-on to all SSR services (just one password)!
See next talk for further discussion on security.

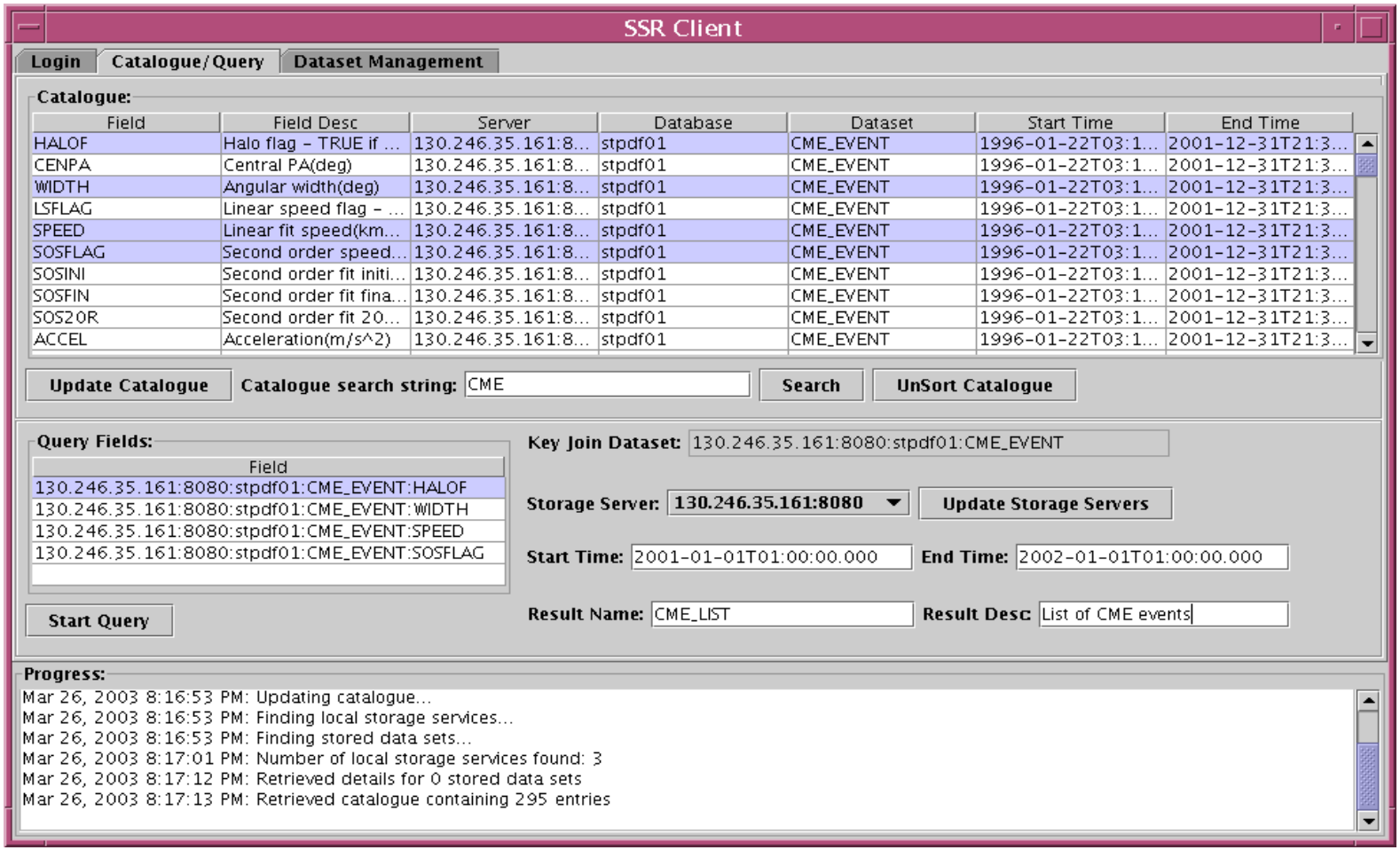
Solar System Research Example



Search for available parameters from all registered archives

CCLRC Rutherford Appleton Laboratory

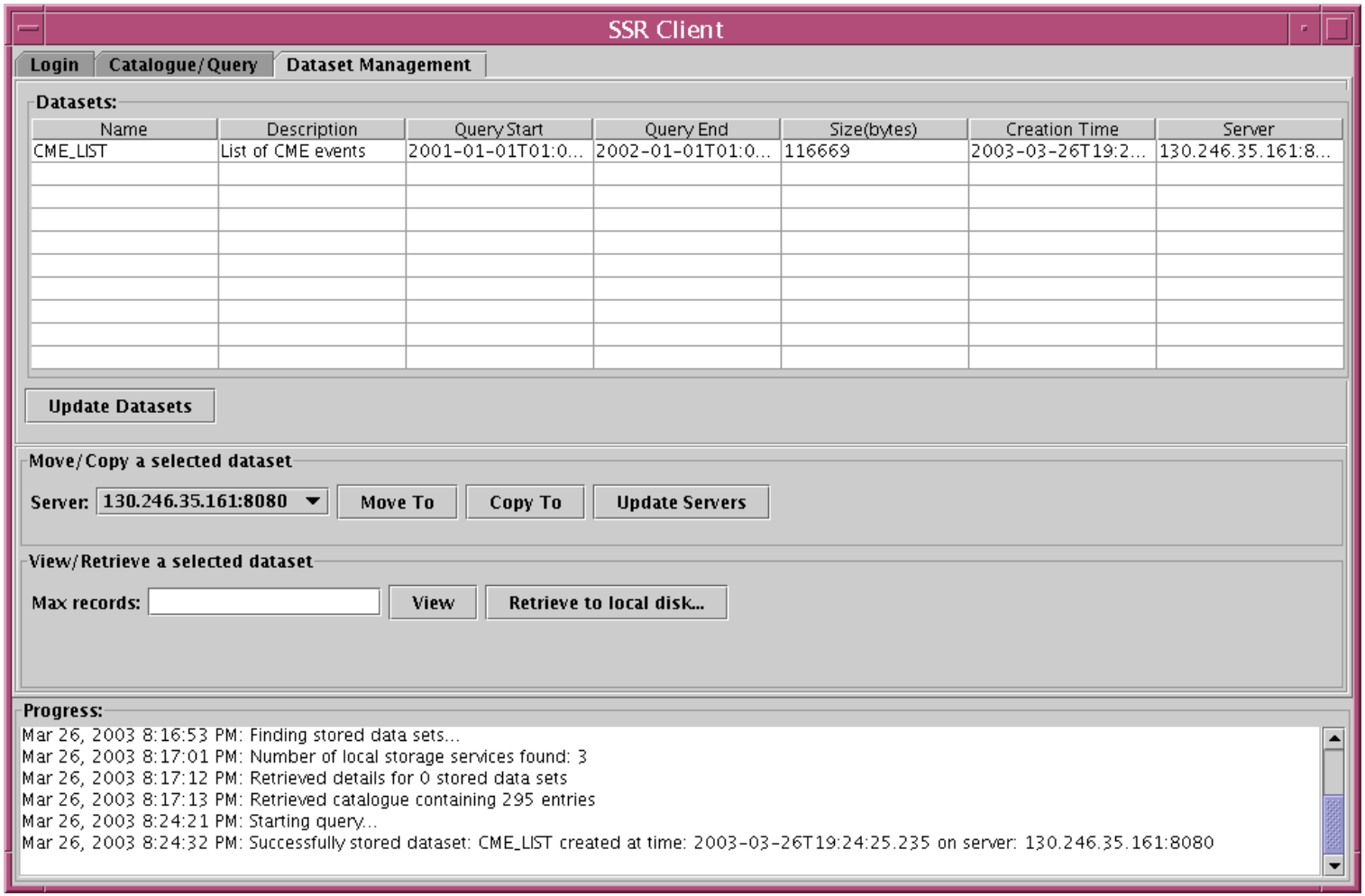
Solar System Research Example



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

SpaceGRID

Solar System Research Example



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

Solar System Research Example

Dataset: CME_LIST

Details:

Description: List of CME events **Size(entire dataset):** 116669 **Storage server:** 130.246.35.161:8080

Query start: 01-01-01T01:00:00.000 **Query end:** 02-01-01T01:00:00.000 **Creation time:** 03-03-26T19:24:25.235

Records (all):

Time	spgstp.bnsc.rl.a...	spgstp.b...	spgstp.bnsc.rl.a...	spgstp.bnsc.rl....
2001-01-01T10:30:31.000	0	50	587	1
2001-01-02T05:54:05.000	0	49	171	1
2001-01-02T08:30:05.000	0	52	229	1
2001-01-02T12:30:05.000	0	128	163	1
2001-01-02T13:54:06.000	0	41	577	1
2001-01-02T17:54:28.000	0	60	161	1
2001-01-02T21:30:10.000	0	114	434	1
2001-01-02T21:54:05.000	0	113	542	1
2001-01-03T04:30:05.000	0	53	490	1
2001-01-03T07:31:49.000	0	50	375	1
2001-01-03T10:54:06.000	0	40	411	1
2001-01-03T12:54:05.000	0	146	293	1
2001-01-03T14:30:05.000	0	35	523	1
2001-01-04T01:31:47.000	0	25	134	1
2001-01-04T07:54:08.000	0	46	596	1
2001-01-04T21:54:05.000	0	28	199	1
2001-01-04T23:06:05.000	0	38	177	1
2001-01-05T11:54:05.000	0	44	554	1
2001-01-05T14:30:05.000	0	25	229	1
2001-01-05T17:06:05.000	1	360	828	1
2001-01-06T08:30:05.000	0	59	321	1
2001-01-06T12:30:06.000	0	68	569	1
2001-01-06T18:06:05.000	0	33	155	1
2001-01-07T04:06:05.000	0	76	633	1
2001-01-07T13:54:06.000	0	43	439	1
2001-01-07T14:54:05.000	0	47	316	1
2001-01-07T21:30:10.000	0	60	614	1
2001-01-07T22:30:32.000	0	37	493	1

Results can be browsed via the client and then downloaded for more detailed examination.

Solar System Research Example

Work is underway on the integration of Starlink Tableview software to provide additional client side manipulation capabilities.

In this example the SSR prototype has been used to create a combined data set of observed CME events from SOHO and the geomagnetic activity index (Dst).

Data can be sorted on any column to assess different relationships

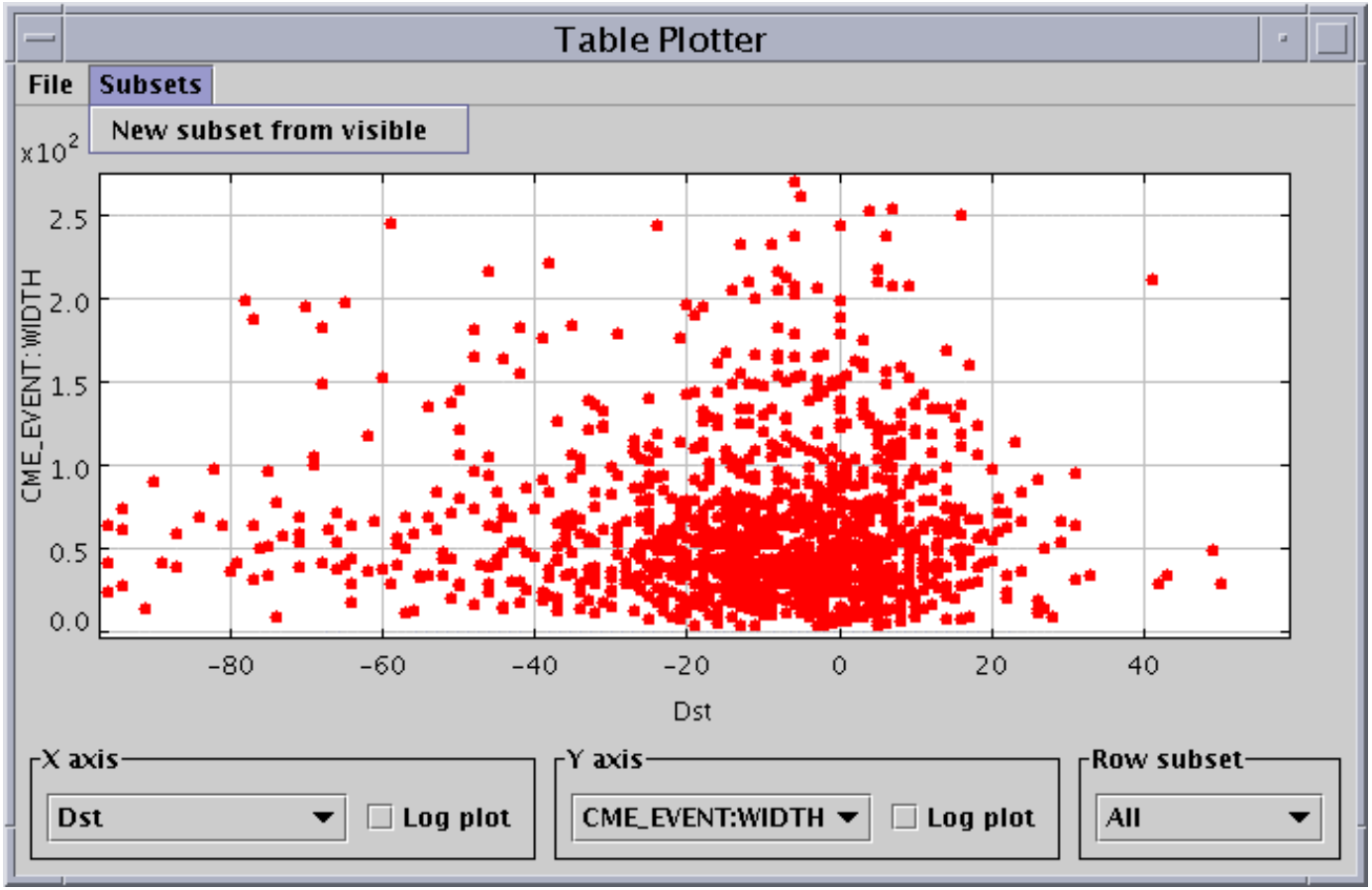
File Windows Subsets

	Time	CME_EVENT:CENPA	CME_EVENT:WIDTH	CME_EVENT:SPEED	Dst
1	2001-01-01T10:30:31.00	259	50	587	
2	2001-01-02T05:54:05.00	222	49	171	
3	2001-01-02T08:30:05.00	261	52	229	
4	2001-01-02T12:30:05.00	226	128	163	
5	2001-01-02T13:54:06.00	59	41	577	
6	2001-01-02T17:54:28.00	152	60	161	21
7	2001-01-02T21:30:10.00	71	114	434	23
8	2001-01-02T21:54:05.00	284	113	542	23
9	2001-01-03T04:30:05.00	192	53	490	-29
10	2001-01-03T07:31:49.00	269	50	375	-14
11	2001-01-03T10:54:06.00	270	40	411	-15
12	2001-01-03T12:54:05.00	36	146	293	-3
13	2001-01-03T14:30:05.00	273	35	523	1
14	2001-01-04T01:31:47.00	74	25	134	3

Delete
New column
Sort up
Sort down

Solar System Research Example

Simple scatter x-y plots can also be produced to examine the dependence of two parameters. The cursor can be used to zoom in on a region and then these events can be selected as a subset.



Solar System Research Example

Expressions can be used either to specify row selection criteria or to define the contents of a new column.

	Time	CME_EVENT:CENPA	CME_EVENT:WIDTH	CME_EVENT:SPEED	Dst	new
1016	2001-10-04T03:06:05.00	114	64	144	-81	89.21096
1007	2001-10-01T23:06:05.00	135	41	338	-79	-336.67768
1006	2001-10-01T22:06:05.00	56	64	153	-77	130.54267
1129	2001-10-29T06:06:06.00	127	50	251	-76	58.32213
325	2001-04-02T07:27:20.00	83	57	603	-73	150.47269
1154	2001-11-01T14:30:05.00	115	195	1053	-70	-343.07772
1011	2001-10-03T01:31:53.00	88	41	302	-68	301.81073
1222	2001-11-25T23:06:54.00	104	182	1574	-68	-1490.37024
404	2001-04-23T10:29:10.00	122	37	556	-66	-481.92250
964	2001-09-23T16:54:05.00	131	64	299	-64	174.67843
1042	2001-10-09T14:06:07.00	66	44	523	-64	-522.81561
1133	2001-10-29T17:54:05.00	45	28	386	-64	202.77428

Table columns

File Columns

Index	Name	Class	Shape	Units	Description	UCD	Expression	$ID	Datatype
0	Index	Long	*		Table row index			$0	
1	Time	String	*					$1	char
2	CME_EVENT:CENPA	Integer	1					$2	int
3	CME_EVENT:WIDTH	Integer	1					$3	int
4	CME_EVENT:SPEED	Integer	1					$4	int
5	Dst	Integer	1					$5	int
6	new	Double	*				$4*cos($2)	$6	

Table Parameters

File

Name	Class	Value
Column count		6
Row count		1376
StartQuery	String	2001-01-01T00:00:00.000
EndQuery	String	2002-01-01T00:00:00.000

Create new column

Column name: new

Column units:

Column description:

Expression: $4*cos($2)

OK Cancel

Solar System Research
Key User 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
Conclusions & Lessons Learnt

- Using GRIDs technology it has been possible to build a distributed systems that demonstrates both the concepts of distributed service based data systems, and identifies some potential bottle necks.
- Web/Grid services provide a standardised replacement for current CGI based mechanisms.
- Basic integration with legacy archive systems can be straightforward and can run in parallel with the existing archive access mechanisms.
- The “Middleware” technology is a rapidly evolving target and still has a steep learning curve.
- Standardisation of the metadata is crucial and requires international coordination e.g. SPASE and other VO work.
- This technology will open up new opportunities for scientists working with legacy archives.

Solar System Research Conclusions & Future Activities

• Need to address some of the technical limitations identified during the current study (see next talk).

• Real world application of this technology within a science project. For example consider use of Grids within aspects of ESA's SSR archives (Cluster and Planetary archives).

• Need a European forum (similar to IVOA and perhaps based on SPASE) to address domain standards and science drivers. Involvement of the science community, national and international facilities.

• Identification and implementation of scientifically valuable tools/applications/visualisation building on top of, and fully exploiting the benefits, of the Grid systems.