

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

C.H. Perry¹, A.P. McDermott¹, P.J. Richards¹, R. Stamper¹, C[.] Morris², E. Purvis² and S. Beco³

 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.

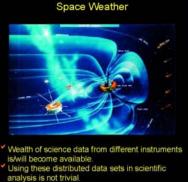




aero-thermodynamic engineering analyses will benefit from multi-disciplinary and collaborative engineering. Radiation Transport Monte-Carlo simulation of high energy particle interactions inside a spacecraft system components or detector with large numbers (>10⁷) of source particles are candidates for 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 (>109). Strong coupling between cells. Challenge the development of independent subtasks for parallel processing.

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



- Sing these valuations are an 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

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 !!



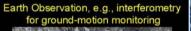
A study funded by the General Studies Programme, © ESA, 2001 Contact-point: Pier.Giorgio.Marchetti@esa.int

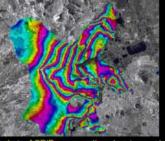


Ö 000/11/26 08:06 Access to multiple distributed data archives. Online data manipulation, analysis and visualisation of complex data sets. Interoperability between legacy systems and heterogeneous databases.

Contact: Gerhard Kreiner@esa.int

Collaborative environment (e.g. sharing of algorithms for analysis and visualisation)





Analysis of GRID-aware earth and environmental onitoring 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.



2 - C. S.





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, multidimensional 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 <u>catalogues</u> of <u>multiple</u> distributed data archives
- The ability to query the <u>data</u> across <u>multiple</u> 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.

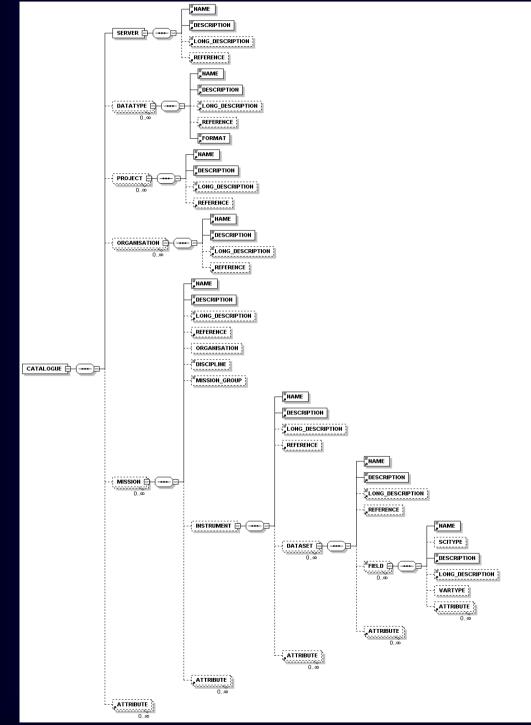
• Much work on this is currently underway within the various Virtual Organisation (VO) projects. The Space Physics Archive Search & Exchange (SPASE) group have a splinter meeting here at EGS (Thursday, 16:30, Gallieni 1) to present work on a data dictionary for space physics. We would encourage interested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

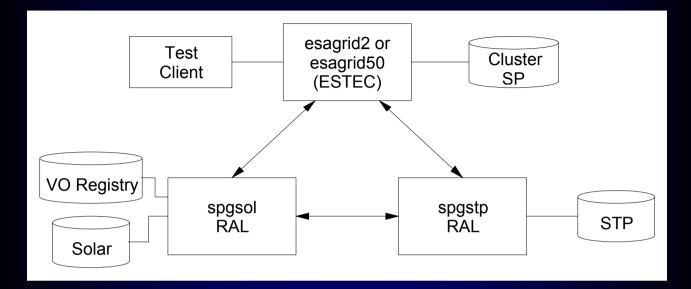




ESA SpaceGRID Study - Lessons Learnt



Solar System Research Prototype Configuration



SSR persistent factory services provided at each hub:-

- LocalCatalogueFactory –
 <u>Access to local metadata information at one site</u>
- 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	9 1	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.

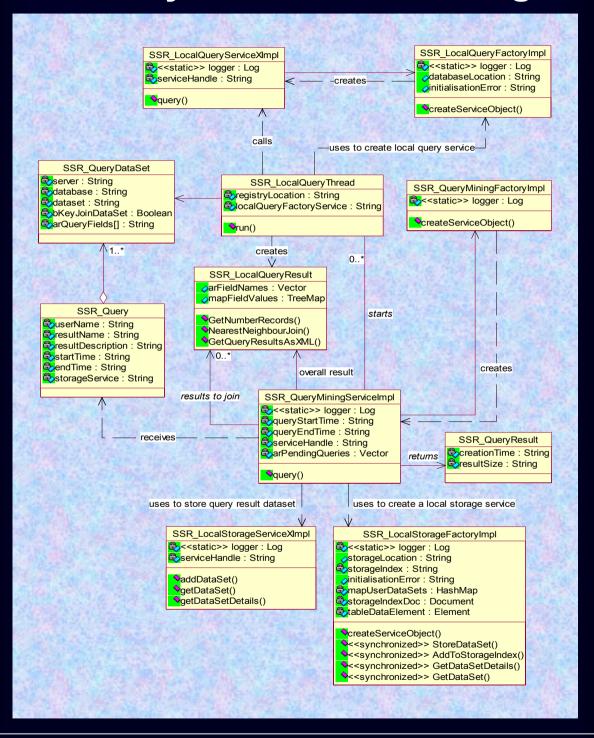




<u>Client</u>	SSR_Registry Helper	Hub1::SSR_Query MiningServiceImpl	Hub1::SSR_Local QueryServiceXImpl	Hub2::SSR_Local QueryServiceXImpl	Hub1::SSR_Local StorageServiceXImpl
1. fact	orvAddress = aets	Service("QueryMining	FactoryService", storage	Server, false, registryL	ocation, logger)
12 1. 14	2. create	and the second			
1	Sale Dr		Section 2	-Walter	ALC: PARTY
3. que	eryResult = query(queryDetails)	Sector Sector	199	ALC: NO.
1 0			Care In the	and the second	The second
4. fact	oryAddress = gets	Service(databaseNam	e + "/LocalQueryFactory	/Service", databaseSer	ver, true, registryLocation, I
			* [for e	each dataset to be que	ried]
		5.1	create	Contraction of the second	Section Section
	S. S. C. K. M	100		100 B	State of the
6 m		5.2 lc	calQueryResult1 = quer	y(datasetName, startTi	me, endTime, queryFields)
1.14				19 19 19	
14 Bus	1945 227	5.3	destroy	Was Sus	AND THE REAL
	12 Bar		5.4 create		CLOBING ST
	The later				Test in the set
		5.5 loc	alQueryResult2 = query	(datasetName, startTim	ne, endTime, queryFields)
		ter al	A State of the state		
	222		5.6 destroy		ALL PARTY AND ALL PARTY
	a de partos	10-1	A DE LESS		
		6. loca	QueryResult1.NearestN	leighbourJoin(localQuer	yResult2, logger)
1.1	Site free	<	Section and		5.000
1.5		7. quer	ResultDataset = localC	ueryResult1.GetQuery	ResultsAsXML(queryInfo)
	THE BASS	<	THE PART		CACE AND AND
	8. factoryAddre	ess = getService("Loo	alStorageFactoryService	e", storageServer, true,	registryLocation, logger)
19		and the states	9. 0	create	(*************************************
		Constant and and	A LONG WE W	CARGE AND AND A	
	10. queryRe	sult = addDataSet(us	erName, resultName, re	esultDesc, startTime, e	ndTime, queryResultDatase
	Alt Sta		11. 0	lestroy	Y
	Contraction of		12/ C 16 19		
	12. destroy		A STATISTICS		T.
1.0	10 10		SE 102 1	243 53	a man a
L L	A STAT		ALTER STATE		CHARLES AND THE SA
					10 20 20.00
1		in the second second	18	1	



Solar System Research Query Related Class Diagram

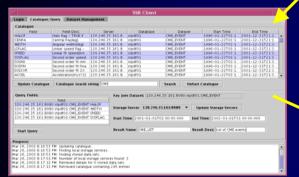




Solar System Research Prototype System (Java Interface)



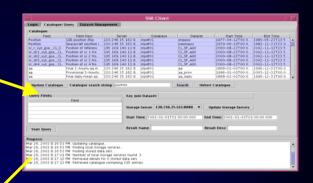
Single sign-on to all SSR services (just one password)!



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

F		Datase	t: CME_LIST		· []
Details:					
Description: List of CME event	s	Size@ntire	dataset): 11666;	Storage server: 13	0 246 15 161 8080
Query start 01-01-01T01:0	0:00.000 Query	end: 02-0	1-01T01.00:00.0	Creation time: 03-03-2	ST19.24.25.235
Records (all):					
Time	spasto breici	taul seene:	b ipato bric ri	a. spgstp.bnsc.rt	
2001-01-01710 30 31 000	0	50	587	1	
2001-01-02705 54 05 000	0	49	171	1	
2001-01-02108 30 05 000	0	52	229	1	
2001-01-02712 30 05 000	0	128	163	1	
2001-01-02713 54 06 000	0	41	5.77	1	
2001-01-02717 54 28 000	0	60	16.1	1	
2001-01-02721 30 10 000	0	224	434	1	
2001-01-0272154 05 000	0	113	5.42	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	2.93	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	
2001-01-04T2154 05 000	0	28	199	1	
2001-01-04T23 06 05 000	0	2.6	177	1	
2001-01-0571154 05 000	0	44	554	1	
2001-01-05714 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	611	1	
2001-01-07T13 54 06 000	0	43	439	1	
2001-01-07T14 54 05 000	0	47	316	1	
2001-01-07721 30 10 000	0	60	614	1	
2001-01-07722 30 32 000	0	27	491	1	

Tabular view & download data to local machine.

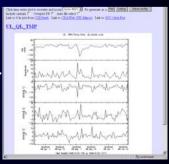


Search for available parameters from all registered archives

-			SSR Client			· .
Login Catalogue	Query Dataset Ma	nagement				
Datasets						
Name	Description	Query 2tert	Query End	Size(bytes)	Creation Time	Server
CMELLIST	Ust of CME events	2001-01-01701.0	2002-01-01701-0	116669	2003-03-26719 2	130 246 35 161 8
Update Datasets						
Move/Copy a select	ted dataset					
Server: 130.246.3	.161:8080 · Ma	we To Copy To	Update Servers			
View/Retrieve a sel	ected dataset					
Max records:		View Retrieve	to local disk			
Mar 26, 2003 8 17.0) PM Finding stored d 1 PM Number of local 2 PM Retrieved details	storage services found.	3			
Mar 26, 2003 8 17:1 Mar 26, 2003 8 24:2	3 PM. Retrieved catalog 1 PM. Starting query	pue containing 295 entri		1-26719/24 25.235	on server: 130.246.35-1	162 8080

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

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.

