SOLAR SYSTEM RESEARCH
USER REQUIREMENTS DOCUMENT

	Name	*Date*	*Signature*
Prepared by :	**Chris Perry** (RAL)	25/05/2002	
Reviewed by :	**Chris Perry** (RAL)	25/05/2002	
Approved by :	**Stefano Beco** (DATAMAT)		
Accepted by :	**Gerhard Kreiner** (ESA) **Pier Giorgio Marchetti** (ESA)		

Distribution List

Company	*Name*	*Number of copies*
ESA	Pier Giorgio Marchetti	e-mail
	Gerhard Kreiner	e-mail
DATAMAT	Stefano Beco	e-mail
	Fiorella Lamberti	e-mail
ASPI	Veronique Perrin	e-mail
CSSI	Pierre Vielcanet	e-mail
RAL	Christopher Perry	e-mail
QIN	Fan Lei	e-mail
	David Rodgers	e-mail
SSL	Conrad Morris	e-mail

Document Status Sheet

Issue	*Date*	*Details*
1.0	18/04/2002	First Issue
1.1	10/05/2002	Issue after URR
1.2	25/05/2002	Some additional minor updates and re-wordings suggested by the ESA technical representative.

Document Change Record

Issue	*RID No.*	*Details*
1.1	UR-SYS-SB-001	Page 1: “Accepted by:” field added and other fields re-organised to take account of this.
1.1	UR-SYS-SB-002	Page 2: Distribution list updated.
1.1	UR-EOA-PGM-004	Page 10: OGSA added to acronym list. Page 11: OGSA web address added to the list of reference documents.
1.1	UR-EOA-PGM-005	Page 11: Statement added concerning applicability of requirements with reference to ongoing advances in GRID technology.
1.1	UR-SSR-SB-001	Sections 4 and 5 combined. Various references in other sections updated accordingly.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

The aim of SpaceGRID in the Solar System Research (SSR) domain is to remove the barriers that currently limit the efficient analysis of distributed data by providing a framework for uniform access to, and manipulation of, local and distributed data sources. The SpaceGRID should not only endeavour to enable research that could not be done before. It should also facilitate analyses that up to now have been very difficult, but not impossible, by reason of data or information inaccessibility, or have only been possible by a chosen few by reason of resource limitations. The SSR part of SpaceGRID is, therefore, not only about handling large amounts of data or providing large scale high performance computing, but is equally about facilitating the handling of small but complex and diverse amounts of data. In the short-term the federation of existing SSR resources will result in improved and faster science return. Longer-term SpaceGRID will encourage the standardisation of design, promoting re-use of analysis systems, long-term security of data and reduced overall cost.

1.1 Purpose of the Document

Task 5 of [SPCGRD-SOW] called on the SpaceGRID consortium to establish the requirements, assess the benefits and evaluate the implementation options for ESA development of a collaborative data analysis system for the Solar System Research community. The purpose of this document is to establish the user requirements for this system.

1.2 Definitions and Acronyms

1.2.1 Acronyms

ACE	Advanced Composition Explorer
API	Application Programming Interface
CDAWeb	Combined Data Analysis Web
CDDP	Centre de Donnees de la Physique des Plasmas
CDF	Common Data Format
CD-ROM	Compact Disc – Read Only Memory
CME	Coronal Mass Ejection
CoI	Co-Investigator
CSDS	Cluster Science Data System
DBMS	Data Base Management System
DITDOS	Distributed Inventory Tracking and Data Ordering System
EISCAT	European Incoherent Scatter Radar
ESA	European Space Agency
ESTEC	ESA Technology Centre
EU	European Union
FITS	Flexible Image Transport System
FTP	File Transfer Protocol

GSFC	Goddard Space Flight Centre
HPC	High Performance Computing
IDL	Interactive Data Language
ISIS	Integrated Software for Imagers and Spectrometers
ISTP	International Solar Terrestrial Physics
LHC	Large Hadron Collider
LLBL	Low Latitude Boundary Layer
NAIF	Navigation and Ancillary Information Facility
NASA	National Aeronautics and Space Administration
OGSA	Open Grid Services Architecture
OODT	Object Oriented Data Technology
PDS	Planetary Data System
PROP	Proposal
PSBL	Plasma Sheet Boundary Layer
RAL	Rutherford Appleton Laboratory
RPIF	Regional Planetary Image Facilities
SCT	Special Conditions of Tender
SMART-1	Small Missions for Advanced Research in Technology-1
SOHO	Solar Heliospheric Observatory
SOW	Statement Of Work
SP	Small Projects
SPARC	Space Physics & Aeronomy Research Collaboratory
SPCGRD	SpaceGRID
SPICE	Spacecraft Planet Instrument “C-matrix” Events
SPP	Solar Planetary Physics
SSL	Science Systems Limited
SSR	Solar System Research
STP	Solar Terrestrial Physics
STPDF	STP Data Facility
TRACE	Transition Region and Coronal Explorer
UARC	Upper Atmosphere Research Collaboratory
UK	United Kingdom
US/USA	United States
WP	Work Package
WWW	World Wide Web
XDF	eXtensible Data Format
XML	eXtensible Markup Language

1.3 REFERENCES

1.3.1 Applicable Documents

[SPCGRD-SCT] AO/1-3863/01/I-LG, Invitation To Tender – Study of GRIDS and Collaborative Environments for Space Application – Appendix 3 – Special Conditions of Tender

[SPCGRD -SOW] AO/1-3863/01/I-LG, Invitation To Tender – Study of GRIDS and Collaborative Environments for Space Application – Appendix 1 - Statement of Work

[SPCGRD-PROP] P-A876/USA/PT-656-01, Is. 1.1, Proposal for "Study of GRIDS and Collaborative Environments for Space Application (SpaceGRID)"

[PSS5-SP] BSSC(96)2 Issue 1, Guide to applying the ESA software engineering standards to small software projects

1.3.2 Reference Documents

[GRID] I. Foster and C. Kesselman, Eds., The GRID Blueprint for a New Computing Infrastructure, Morgan Kaufmann, 1999

[OGSA] http://www.globus.org/ogsa

1.4 OVERVIEW OF THE DOCUMENT

The structure of this document is as follows.

- Section 2 provides an overview of the study logic and methodology that was used for requirements identification and specification.
- Section 3 describes the Solar System Research (SSR) domain to which the study activities have been addressed. In particular, this section provides an overview of the different SSR sub-domains, focusing on current facilities and applications and providing an example case of how GRID technology might be used.
- Section 4 provides the results of the analysis carried out to define the requirements. The section then provides a general description of the requirements for the SSR application of SpaceGRID. The section also lists the specific requirements identified in the course of the study. A relationship table and a set of orthogonal qualifiers have been assigned to assist with the requirement categorisation.

Where possible the development and definition of the User Requirements has not been based on assumptions about the architecture and services provided by the current GRID technology. However, we do note that due to the parallel evolution of the technology (e.g. [OGSA]) the requirements may not fully reflect the "technology state-of-the-art".

2. APPROACH TO REQUIREMENTS DEFINITION

This section outlines the method that has been used to generate a consolidated set of user requirements for use in definition of the SSR portion of SpaceGRID, and for the selection of particular technologies and aspects of the architecture to be investigated within the prototype development. Figure 2- 1 shows the requirements definition phase in the context of the overall SSR study activity.

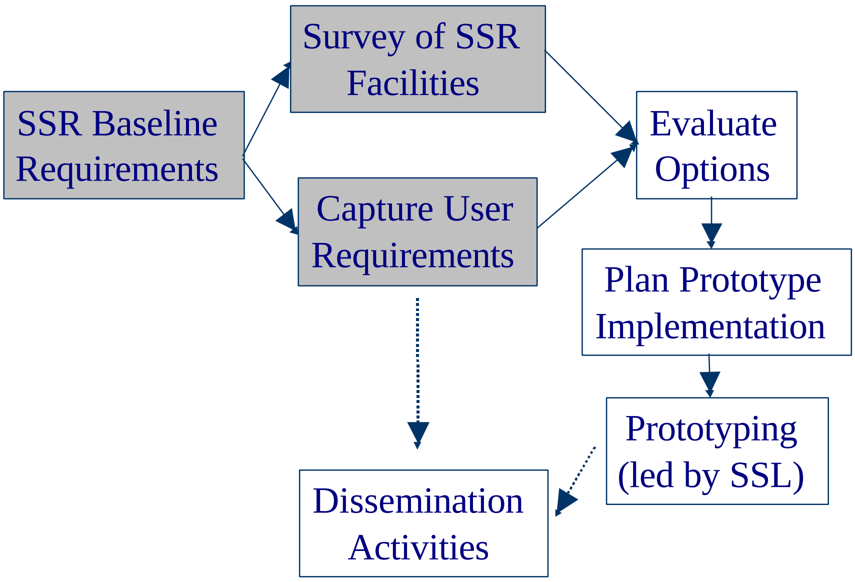

Figure 2- 1: SpaceGRID SSR study activities

2.1 STUDY LOGIC

The requirements analysis was initiated by domain experts examining the disciplines covered by Solar System Research (SSR) and identifying areas where there is already (or planned to be) collaborative working and where, from a science perspective, there is a need for such activities. This provided the necessary technical base to constrain and focus the development of the detailed requirements. This knowledge formed the basis of a consultation with the user community to establish what a SpaceGRID should deliver to provide practical, strategic, benefits to the SSR community and ESA. The detailed requirements described in Section 4 provide the input into the development and evaluation of the "optimistic" implementation options for the collaborative analysis environment described in SGD-SSR-RAL-SSTN-0005 and for the particular areas to be developed within the “realistic” scenario. This “realistic” implementation will be used as the basis for the SSR prototyping application work package (WP4300). The results of the prototyping work shall be used to reassess both the requirements described in this document and the proposed implementation strategy for a full system.

For the purpose of the requirement capture, the community has been split into three sub-domains corresponding to different science communities within Solar System Research. These research areas are distinct not only in the science undertaken but also in their data handling strategies, data formats, analysis systems, tools and environments (as described in Section 3). The sub-domains are:-

1. Planetary Research - Planetary formation, evolution, properties and Solar Planetary Physics
2. Solar Research - The study of the formation, evolution and properties of the Sun
3. Solar Terrestrial Physics - Near Earth space plasma physics, understanding the transport of mass, energy and momentum from the Sun to the Earth.

2.2 METHODOLOGY

The local domain experts for each of the three areas drew up a draft set of requirements and provided a preliminary assessment of the scope for the SpaceGRID system for SSR. Due to the breadth of the SSR field it was considered important to undertake this preliminary assessment prior to the detailed consultation with the community.

An understanding of the community's requirements for SSR collaborative data analysis system was an essential input to the system requirement definition activity. Using the baseline requirements, the domain experts surveyed the SSR community. Use was made of existing lines of communication such as direct links with key players in the field, community e-mail lists and relevant project meetings. The primary source of information was a questionnaire used to gather inputs from as broad a spectrum of SSR users and data providers as possible.

The questionnaire (see Annex 1) was hosted on a web server running at RAL and the community first notified via E-mail on 15/01/2002. The E-mail distribution list was a combination of Solar and STP lists based on existing lists for these groups. At this time a list for the planetary community was not available and so the project scientists on Bepi-Colombo, Mars-Express and SMART-1, and the maintainer of the UK planetary forum were invited to forward the notification to their lists. A reminder was sent to the same list on 01/02/2002 and a final reminder on 01/03/2002 to an extended list including e-mail addresses of between one and two hundred planetary scientists working on current ESA missions. In addition to the community questionnaire, related discussions were held with other interested parties including the co-ordinator of the EU European GRID of Solar Observations, an international group led by NASA/GSFC working on the development of standards for space physics queries, the Cluster Science Data System Implementation working group and the UK AstroGrid project.

The results of the consultations and survey were collated and analysed (Section 4) to extract the key requirements, (Section 4.3). Where possible, the requirements have been categorised on the basis of an agreed set of qualifiers. For example, a distinction has been made between the general infrastructure requirements that will be common in many areas (and addressed at least in part by other studies such as the EU DataGRID project) and specific requirements for SSR applications. The [SPCGRD-SOW] placed particular emphasis on the considerations for existing and future ESA science programme missions and this was taken into account when formulating the requirements. However, in order to maximise the science return from these missions, interoperability with data from non-ESA source was also considered.

3. SOLAR SYSTEM RESEARCH DOMAIN

Solar System Research (SSR) is a broad, multi-disciplinary science involving the study of a complex 3-d environment with interactions and phenomena that occur over a variety of temporal and spatial scales. In comparison with the Particle Physics and Astronomy domains, which up until recently have been the main drivers in the application of [GRID] technology, the SSR domain is less focused on measurements of objects and more on the process and evolution of systems. As a result, SSR measurements are often based on time series measurements rather than the more object-oriented observations of other domains. This added complexity implies differences in the storage, cataloguing and data access requirements for SSR. Therefore, unlike the 'LHC challenge' of the High-Energy Physics community and Astronomy's 'Virtual Observatory', the problem in SSR is not the handling of enormous data volumes. Typical archive volumes in the SSR domain are of the order of a few terabytes, or less. Instead the challenge for SSR comes from the variety and complexity of the data and the need to combine products from a number of sources in order to carry out scientific research. For example, in Solar physics SOHO data may be used in conjunction with data from other missions such as Yohkoh, TRACE or ground-based observatories. In Planetary studies SMART-1 data will need to be compared to data from previous lunar missions such as Apollo, Clementine and Lunar Prospector. In STP the problem is particularly acute due to the large number of multi-point in-situ measurements that are required to constrain the context of an event.

A collaborative analysis environment offers the potential to streamline the tasks of retrieving, analysing and sharing distributed data products thus improving the efficiency of the sorts of analyses that are already carried out today. In addition, the provision of distributed computational and querying facilities will open up new avenues for event identification, statistical surveys and inter-disciplinary investigations. Many of the existing SSR data resources (for example those for SOHO and Cluster) are self-contained and have been developed independently using different underlying software infrastructure, cataloguing systems, access controls and data formats. They are also tied to mission funding with a limited duration. Facilities such as the NASA NSSDC and Combined Data Analysis Web (CDAWeb) have led the way in the promotion of easy access to a broad range of data. While they provide the ability to access and display multiple data sets, the data is all held locally and there is no support for complex operations such as searching or manipulation of the data. A large number of SSR data sets (particularly in STP) are provided by individual instrument teams and dedicated to their own data. This is because of the need for a detailed knowledge of the instrument operation in order to extract the best quality data. In general these data resources are funded either nationally or via the local institution and have a limited lifetime, after which the ability to access the data and technical support can decline rapidly. A potential benefit of SpaceGRID will be to ease the eventual migration of these legacy data to a secure long-term archive before they become inaccessible. It will be necessary to accommodate these legacy systems as well as supporting future developments. Data and Metadata format standardisation across the whole of SSR is not a realistic solution and so alternative approaches such as data translation will be required.

Consideration must also be given to the theoretical modelling activities of the community. Models underpin the interpretation and understanding of the observations and facilitating their combined use should lead to major advances.

The three areas covered by SSR represent distinct research communities that to a large extent work independently from each other. The following sections provide brief descriptions of each of these sub-domains.

3.1 PLANETARY

Planetary activities break down into several areas studying the formation and evolution of planets:-

- Interiors
- Surfaces
- Atmospheres
- Magnetospheres (including SPP, Solar Planetary Physics).

Some of these will also require access to laboratory data (i.e. analysis of asteroids/meteorites), however, laboratory data is usually the result of individual groups and has not traditionally included the involvement of the national space agencies.

The limited number of planetary data sets means that most scientists combine data sets from either a single mission (e.g. data from different wavelength ranges), or using future datasets alongside existing data (mostly from US missions). It is important to note that SpaceGRID will be of limited use if there is no way to either identify or preferably incorporate these existing datasets. As part of this exercise of combining data sets it will be important to have a common coordinate system to make it easier to locate a specific area in a variety of datasets. For planetary surfaces, areas are often better known as names (i.e. Utopia Planitia). Therefore a facility to convert names into coordinate locations would simplify the selection of data from different missions. Ideally this would be combined with a facility to know what kind of data exists for a particular area of research (i.e. "I want to know what magnetic data exists for the terrestrial planets").

3.1.1 Planetary Data and Facilities

For planetary data the main facility is the Planetary Data System (PDS) in the US. The data held by the PDS are distributed between a set of seven discipline specific nodes that are staffed by specialists in the extraction, use and interpretation of the data. The PDS defines a set of standards that are intended to assist in the long-term accessibility of the data with a particular emphasis on documentation required to access the data. The data itself may be stored as either text or binary files and the PDS imposes very few restrictions on the physical or logical organization of the files. The data will always have an associated PDS label, written in the Object Description Language, that provides a detailed description of the observations through the use of a very comprehensive data dictionary. Spacecraft orbit, attitude and instrument pointing information is normally provided in the form of a SPICE (Spacecraft Planet Instrument “C-matrix” Events) kernel produced by the Navigation and Ancillary Information Facility (NAIF) PDS node. Together with the F77 and C toolkits supplied by NAIF, this provides a standard way of handling the ancillary observation geometry data required both for planning and data interpretation. The PDS uses, or is developing, a number of tools to support access to the distributed data held within their nodes, these include the DITDOS (Distributed Inventory Tracking and Data Ordering System) that provides data location facilities and OODT (Object Orient Data Technology) for data interoperability.

In addition to PDS, NASA has "Regional Planetary Image Facilities" (RPIF's) whose job it is to disseminate US planetary data to those people who want to use it (from professional researchers to the media and public). These include a half a dozen centres located around Europe.

There are currently very few European based planetary datasets available. The main historical data set is from the 1986 ESA/Giotto encounter with Halley’s comet (data available on a set of International Halley Watch CD-ROMs). However, this lack of data is set to change due to ESA’s active planetary programme over the next decade, which includes:-

- BepiColombo (Mercury, ~2009)
- Cassini/Huygens (Saturn/Titan, ~2004)
- Mars Express (Mars, ~2004)
- Rosetta (comet 46 P/Wirtanen, ~2011)
- SMART-1 (Moon, ~2004)

To support these missions ESA is currently in the process of setting up a European planetary archive to be co-located with the ESA planetary operations centre at ESTEC. The ESA planetary archive is following the PDS standards so the data should be well described, of good quality, and relatively easy to use.

3.1.2 Planetary Software

The ISIS (Integrated Software for Imagers and Spectrometers) software developed by the US Geological Survey is quite widely used for the analysis of spectral data (particularly in the US). The PDS also provides IDL procedures for assisting in the handling of PDS labelled data and tools, such as MapMaker to assist with image tiling/mosaicing. However, as with the other SSR sub-domains, a lot of the planetary data analysis is undertaken using specialised software developed within individual groups.

3.1.3 Planetary example case

This section provides a simple example of how SpaceGRID might be used within the planetary sub-domain.

> The South Pole-Aitken (SPA) Basin is the least known lunar terrain. Clementine laser altimeter measurements showed it to be 12 km deep. At 2500 km diameter it is also the largest impact feature in the Solar System. The albedo map shows that the floor of the basin is markedly darker than the highlands surrounding it. Both iron and titanium concentration maps show enhanced values associated with the basin floor materials. These results suggest that the floor of the SPA is made up of relatively mafic rocks this could either be a consequence of volcanic resurfacing or the major impact may have stripped off the feldspathic crust, exposing the mafic upper mantle of the Moon. Elemental mapping of the SPA is required to enable us to choose between these possibilities and would be a major step forward in constraining formation theories.

SpaceGRID could used in this analysis to identify and collect together the required datasets for subsequent analysis. For example, “Obtain lunar elemental and mineralogical composition data for the South Polar Aitken basin”. The steps that would be undertaken by SpaceGRID to such a request might be:-

1. Use database of Lunar regions to convert “South Polar Aitken Basin” to lunar latitude and longitude
2. Identify possible sources of requested data (SMART-1 & Lunar Prospector)
3. Locate source for SMART-1 data (ESTEC PDS based Planetary Archive)
4. Identify the instrument that returns elemental composition information (D-CIXS)
5. Identify and extract all data from the mission that was taken from that location
6. Re-format data, standardise units and co-ordinate system

7. Repeat steps 3) to 6) for Lunar Prospector data obtained from PDS archive in the US.
8. Deliver product to the user for comparative study using specialised local tools.

3.2 SOLAR

Solar activities break down into several areas studying the processes driving the sun:-

- Solar wind
- Solar flares
- Corona and Coronal Mass Ejections (CMEs)
- Solar Atmosphere
- Helioseismology

The main issues within the solar area are the standardisation of metadata to encompass all datasets and the creation of comprehensive catalogues of the data. This will address the main problems of identifying observations associated with a particular type of event amongst the heterogeneous and widely distributed sets of data. Users will want to be able to identify events based on time, location, wavelength, detected feature or some combination of these. It is very important that users are able to judge the quality and applicability of the data (e.g. via a thumbnail image or plot) before downloading the files, and should not need to know the details of the instrument to be able to do this. Some pre-processing of the data might be necessary prior to selecting the data for download. For example the user may request a movie be made and displayed in order to judge the time evolution of a feature. Other added-value data products (e.g. velocity maps) might also be available to help in the selection process. A lot can be done remotely from the user i.e. local to the data store. This implies that archives should also be centres of processing power in order to avoid excessive network traffic. For example, the user might request an overlay display of any pair of datasets, or a movie from several datasets, or they may request that an EIT image be displayed and that the CDS spectra at each location clicked on in the EIT image are displayed. Finally when the user is happy with the data selection, the data, either raw or processed is downloaded to the user's own site or is pipeline processed at a suitable site. Processed data could include not only fully cleaned and calibrated data, but also added-value products like velocity maps or line ratio maps, or anything user-up linked software could provide.

3.2.1 Solar Data and Facilities

The main European solar dataset currently in use is from the joint ESA/NASA SOHO mission however these are not located at an ESA institute. The SOHO archives in US and Europe (UK, France, Italy) generally hold the data on robotic tape stores. The archives do not provide the very latest data. Interfaces are web-based, with selection of files by parameter (date, position, wavelength etc) with little or no information for user on the characteristics of data to assist with making a sensible selection. Transfer of the selected data is via ftp or hard medium. Data files are all FITS but contents differ (e.g. some have single observation per file, some multiple images per file or even multiple files per image). Summary data images and samples of latest data are available from the web.

Other data sets and archives that are used in conjunction with SOHO include:-

- TRACE – Archives in US and UK. Own data format for files – similar to Yohkoh.

- YOHKOH – Archives in Japan, US and UK. Yohkoh has its own data format for files. One file is produced per orbit and interpreted by an *index* file. There is a web-based archive interface that can be used to select on time and instrument with web download of tar file. Typical response time for download is a few minutes.
- GOES – X-ray monitor, available as plots or data lists from NOAA web.
- Many ground-based observatories providing data that are usually accessed by ftp. The best known ones create FITS files with appropriate metadata.

There are also a number of legacy data sets from SMM, BATSE, OSO-7 (Space Data Archive Center in US) and Hinotori (NAOJ, Japan) that are used for science analysis.

In addition to the missions listed above that are aimed at remote sensing of the near-Sun environment, there is also the ESA/Ulysses mission that has the specialised task of investigating the high latitude solar wind and interplanetary magnetic field. Data is accessible either on CD-ROM or at the ESTEC Ulysses data centre via a web-based interface. Ulysses provides some overlap between the Solar and Planetary sub-domains. In a similar way, missions like Wind, ACE and the particle instrument on SOHO provide up-stream solar wind measurements that are of interest to both the Solar and STP sub-domains.

Eventually data from all the sources mentioned above will be required to form the available resources within any large scale SSR GRID. However, a representative subset of instrument data to include imaging and spectral instruments, for example SOHO/MDI, EIT and CDS together with Yohkoh, TRACE and some ground-based data, would form a sufficiently diverse set with which to define metadata and test prototype techniques.

The European Grid of Solar Observations, supported through the European Commission, is a partner initiative to the SpaceGRID and will be attempting to address many of these issues.

3.2.2 Solar Software

The international solar physics community has standardised on a comprehensive (over missions and instruments) library of data reduction and analysis software called SolarSoft. The software is predominantly written in Interactive Data Language (IDL). This has the drawback that it requires the commercial IDL environment in order to run but the significant advantage that it is platform independent. The SolarSoft library is maintained in the US, but contributed to worldwide with mirror sites being updated daily to ensure that the software tree remains up-to-date.

3.2.3 Solar example case

This section provides a simple example of how SpaceGRID might be used within the solar sub-domain.

> Identify the brightest 100 X-ray bright points seen in SOHO spectroscopic data, which also have Yohkoh SXT and MDI co-spatial and co-temporal data. Overlay magnetogram contours on monochromatic images and derive emission measures from the spectra.
>
> SpaceGRID tools required:
>
> 1. Specify data characteristics (eg wavelength or solar feature) and/or instrument set-up (e.g. filter or exposure time)

2. Search single or multi-instrument catalogues or observing logs for data matching the criteria.
3. Remotely inspect sample of data in order to check or refine (1)
4. Calibrate the individual datasets
5. Produce images and overlays according to catalogued pointing information
6. Analyse line intensities

In developing and prototyping a SpaceGRID architecture it will be necessary to assess a number of options. Should the calibrated data be cached locally to the user for further analysis or should just the processed contour plots be downloaded? Should the user's own software be up linked to the GRID to perform the spectral analysis or should the spectral data, being a much smaller volume, be sent to the user?

3.3 Solar Terrestrial Physics (STP)

Solar Terrestrial Physics (STP) area covers a broad range of activities involved in studying the interaction of the solar wind with the Earth's magnetosphere and the resulting evolution of particles and fields in near-Earth space and the effect on the ground:-

- Ionosphere
- Aurora
- Plasmasphere
- Magnetosphere dynamics (storms and sub-storms)
- Magnetosphere boundary layers (LLBL, Magnetopause, PSBL)
- Bow shock
- Solar wind interaction

The main issues within the STP area are over data location, access and standardisation. Data location is made difficult by the large number of in-situ and ground based measurements that are available. In order to build up a complete picture of the scientific process under investigation many of these point measurements may need to be combined. Although there are a few major STP archives (CDPP in France, GSFC in the US), individual instrument teams provide much of the data and as a result there is little standardisation in the access interfaces or formats of delivered data. High quality data is often not available online, instead requiring instrument teams to run specialised software and apply local knowledge of the instrument. There is little format standardisation except at the level of survey products where the ISTP (International STP) has standardised on the NASA Common Data Format (CDF) and a comprehensive set of metadata that allows automated functions such as co-ordinate transformation to be applied in a robust manner. For the Cluster mission, a flatfile ASCII exchange format has been developed by Queen Mary College, University of London and adopted by the Cluster data archiving group. It is too early to tell at this stage how widely it will be adopted.

3.3.1 STP Data and Facilities

The main European space based STP dataset is from the ESA Cluster mission. Simple parameters are available at spin (prime) resolution from each spacecraft, and one-minute (summary) resolution from a single spacecraft, from the Cluster Science Data System (CSDS). The spin resolution data is only accessible by Cluster CoIs who must provide a user name and password to gain access. Both prime and summary data are stored in CDF. The CSDS is based on a distributed architecture consisting of national data centres that are responsible for the generation of data from their instruments. The data is then exchange around the centres (Austria, China, ESTEC, France, Germany, Hungry, Sweden, UK and US) so that they all maintain a local mirror of the entire data set. High-resolution and publication quality data is only accessible by making requests directly to the instrument teams. Data can be requested as CDF, ASCII or a preview plot. Plans for a long-term Cluster archive are still to be decided. It is expected that the Chinese/ESA Double Star mission will re-use much of the infrastructure developed for Cluster.

There are a significant number of European ground-based datasets. EISCAT (European Incoherent Scatter Radar) produces data that can be request from national representatives. Other ground-based data including magnetometers, all sky cameras, ionosondes etc. also available at individual facility web sites. Geomagnetic activity indices are available from the World Data Centres for STP in the UK or World Data Centre for geomagnetism in Japan.

Within Europe the main national STP archive is the CDPP (France) that supports more than seventy archived data sets and is actively involved in ensuring the long-term availability of these data. The CDPP is primarily concerned with supporting the French contribution to missions but increasingly involved in related international data sets and on development of tools to enhance data access and interoperability. Other legacy archives, such as those from AMPTE UKS & IRM, and Equator-S are co-located with the Cluster data centres in Germany and the UK.

Work on data interoperability within ISTP has been led by developments at NASA/GSFC. Survey level data from all ISTP missions are available from the NASA, Coordinated Data Analysis Web (CDAWeb) and this system is increasingly being used for higher resolution products. These are provided in CDF or can be converted on the fly to ASCII or preview plots. The main CDAWeb site is located at NASA/GSFC but there are partial mirrors in the UK and Germany. The Satellite Situation Centre (SSC), also located at GSFC, is the main online resource for spacecraft orbit information and provides tools to help identify spacecraft conjunctions with each other and with particular magnetospheric regions.

3.3.2 STP Software

Due to the complexity and variety of STP data there are no standard applications that are in widespread use across the whole domain. General tools are available that provide basic analysis and display of ISTP compliant CDF files; these include QSAS (QMW – Science Analysis System) developed for Cluster and the KPVT (Key Parameter Visualisation Tool). However they are somewhat limited, not least by the lack of high quality data in CDF format. Other data processing systems in use by parts of the community include ISDAT (Interactive Science Data Analysis Tool) produced by IRF-U and used by the Cluster wave community and SDDAS (Southwest Data Display and Analysis System) developed by SWRI and used by the Cluster PEACE team. Both of these systems provide access to distributed data sets held by compatible servers. However, it remains the case that most research activities involve the development of specialised data processing software to combine, analyse and visualise the products used within a single analysis.

3.3.3 STP example case

This section provides a simple example of how SpaceGRID might be used within the STP sub-domain.

A solar-terrestrial physicist studying the morphology of of the Earth magnetotail during the onset of geomagnetic storms might want to identify in-situ observations of the magnetic field in the magnetotail during an interval around multiple geomagnetic storm onsets. For this, data from any spacecraft would be acceptable but the spacecraft would have to be in the right region of geospace. The definition of the magnetotail would necessarily be approximate because the size and shape of the magnetosphere are significantly altered by the magnetic storms themselves. The definition of a geomagnetic storm would involve reference to a secondary data source e.g. a storm might be counted as an interval when the Dst index dip below -100nT.

The steps that SpaceGRID would need to take:-
1. Identify a location for the Dst index and retrieve data for the period of interest (note there may be more than one site holding the same data).
2. Process the Dst dataset to identify the times when the index dipped below -100nT.
3. Identify a list of suitable candidate spacecraft that have orbits that take them inside the magnetosphere and have a magnetometer on board.

BEGIN Iteration over spacecraft

1. Locate an on-line source of data for that spacecraft and request data for a time window around each magnetic storm. This will probably require multiple requests, registration, and authentication might also be necessary, even for data with no access restrictions.
2. Process the retrieved spacecraft data to select those observations when the spacecraft was in the magnetotail. It is assumed that region information is included within the metadata otherwise a further query on spacecraft position will be required.
3. Convert the position data to a defined coordinate system and the magnetic field data to specific units. The appropriate coordinate system will depend on the application, but in this example it would be Geocentric Solar Magnetospheric (GSM), and the magnetic field would be expressed in nanoTeslas (nT).

END iteration over spacecraft

In this case we have only considered data acquisition. How the data should be presented would be highly dependent on the analysis to be done, for example an event-based study or a statistical survey. A generic display tool for generating time series stack plots would be useful for an event-based study, but frequently the most useful output of the system would be simply a collection of time-ordered datasets, or a single composite dataset, with the data from multiple sources presented in a common set of coordinates and units.

4. SPECIFIC REQUIREMENTS

This chapter deals with the results and analysis of the user consultation exercise and the identification of the specific user level requirements. In addition a number of general issues that emerged from the requirements definition exercise are highlighted (see section 4.2).

Comments resulting in requirements are *italicised* in the text and include a user requirements identifier of the form *UR-SSR-tt-xx*.

The requirements are numbered according to the following scheme:-

UR-SSR-**ty-num** (**qual**, …)

Where **ty** is the type of requirement according to the following categories, and num is the sequential number of the requirement within that category.

- FN: functional
- PR: performance
- RS: resource
- DS: distribution
- AC: access
- DC: design constraints

A tabulated list of requirements and an index providing the page numbers where a particular requirement is referenced are provided in sections 4.3 and 4.4.

4.1 Results and Analysis of Consultation

In this section we present the results of the consultation with the Solar System research community. The primary inputs for this exercise were the responses to the questionnaire collected over an eight-week period during the first three months of 2002. A copy of the questionnaire and the corresponding guidance notes can be found in Annex 1. The final version of the E-mail distribution list to which the notification was sent is provided in Annex 2. The individual responses to the questionnaire and the spreadsheet used to help analyse the inputs are provided in Annex 3 and Annex 4 respectively.

The questionnaire consisted of six main sections covering:

1) Organisation
2) Knowledge of GRID Technology
3) Domain Specific Inputs
4) Looking for Standardisation
5) Collaborative Environment
6) SpaceGRID Priorities

The analysis presented here follows that same breakdown. The responses have been separated into the three SSR sub-domains for those cases where significant differences have been identified.

4.1.1 Organisation

A total of 113 responses to the questionnaire were received corresponding to a return of about 16% of the final version of the E-mail distribution list used to notify the community. This is a higher than anticipated return and indicates a significant level of interest within the SSR community for improved data interoperability and collaboration facilities. This is particularly encouraging since one of the potential risks identified in the proposal was lack of participation from what is a very diverse domain.

The geographical distribution of responses is shown in Table 4-1. These figures are calculated from the E-mail address of the respondent and are shown both as the actual number of responses and as a percentage of the number (from the same country) in the distribution list. As can be seen, although the largest number of response came from the UK, the level of return was close to the average.

In several cases a single response was submitted on behalf of a group of users. When these are considered the total number of individuals covered by the survey increases to well over three hundred. However, to avoid skewing the results to a single response, due to a few large groups, we have decided not to include these as additional weightings in this analysis.

Domain	*Country*	*Number of Responses*	*% Return*
.uk	*UK*	*38*	*17%*
.edu	*US academic*	*15*	*26%*
.gov	*US government*	*9*	*47%*
.fi	*Finland*	*7*	*19%*
.fr	*France*	*6*	*7%*
.ru	*Russia*	*5*	*63%*
.int	*ESA*	*4*	*31%*
.at	*Austria*	*3*	*50%*
.de	*Germany*	*3*	*5%*
.mil	*US DOD*	*3*	*150%*
.se	*Sweden*	*3*	*12%*
.be, .ca, .ch, .com, .cz, .dk, .gr, .hu, .in, .it, .mx, .pl, .sk, .za	*Various*	*1 or 2 per address*	*11%*
Total		*113*	*16%*

Table 4-1: Geographical Distribution of Responses

The SpaceGRID design should ensure good accessibility, by the SSR community, to the system's services (e.g. through a distributed architecture, use of data caches, mirrors & replicas, and good links onto national and international research networks) (UR-SSR-FN-001).

The breakdown between the different scientific sub-domains within SSR is shown in Table 4-2 as a percentage of the total number of responses. The Solar and STP areas are well represented while the number of responses received from the planetary community is comparatively low. We believe that this lower response can be attributed to the lack of a good planetary E-mail distribution list for the initial notification (see Section 2.1) and to the limited number of European planetary data sets that are currently available. As a result we are aware that the analysis of the planetary responses is not fully representative of that part of the community.

SSR Area	***% of total responses***
Solar	55%
STP	36%
Planetary	9%

Table 4-2: Distribution of responses by sub-domain

The core SpaceGRID SSR infrastructure shall be general such that it can be used to support research activities within all three areas of SSR as well as inter-domain and interdisciplinary studies (UR-SSR-FN-002).

To help with an understanding of the facilities required, the community was asked to specify the sorts of activities in which they are involved. *Table 4-3* lists the percentages of respondents involved in the six most common categories. Since SSR is predominantly a research domain and the questionnaire was addressed primarily to university groups and research organisations, it is not surprising that the main activities highlighted by the community were data analysis (including data dissemination), theoretical modelling and related software development. What was less expected is the relatively high proportion, over forty- percent, of users who indicated involvement in operations activities. *SpaceGRID should have the capability to handle science planning and operations products such as inputs from models, orbit predictions and planning events (UR-SSR-FN-003).* The availability of such products, via a common interface, will assist the subsequent analysis of the collected data (e.g. by allowing comparison of planned instrument mode changes with features in the data).

Activity	***% of users involved***
Data Analysis	96%
Data Provision	52%
Theoretical Modelling	49%
Operations	42%
Software Development	40%
Engineering	26%
Others	6%

Table 4-3: Community involvement in SSR activities.

4.1.2 GRID Technology

To assess how widespread knowledge about GRID technology was within the domain, the community was asked to indicate their level of involvement in current or planned GRID related initiatives. The majority of users hadn't previously heard about GRIDs or were not sure what the technology could offer (see *Table 4-4*). Those who knew something about it had mostly found out through the media, conference talks, scientific publications, or as a result of calls for ideas from their national e-science programmes.

GRID Involvement	
Involved in current GRID activities	5%
Know something about it but not currently using it	36%
Heard of GRIDs, but now sure what it all means	38%
Not heard about GRIDs before	21%

Table 4-4: Community knowledge and involvement in GRID activities

Of those actively using GRIDs about half are involved in High Performance Computing developments, such as a Globus 2 based computing Cluster supported by the Nordic GRID committee and the UK White Rose HPC GRID. The remainder working on distributed data systems such as the Active Region Monitor (ARM) for solar physics in the US and the AstoGRID project in the UK.

The system shall support both novice and expert users (UR-SSR-FN-004).

4.1.3 Domain Specific Inputs

This section deals with information about data use, access and processing that is specific to the Solar System Research domain.

4.1.3.1 Data Use

The community were asked to indicate their main use of SSR data. The results (*Table 4-5*) are consistent with the related activity breakdown described in section 4.1.1, with the main uses of data being for research and to a lesser extent in support of operations activities. A broad range of topics within the research areas were identified covering planetary interiors & surfaces, planetary and terrestrial atmospheres, ionosphere, aurora, magnetosphere, solar wind, solar terrestrial interaction, solar flares, corona, solar atmosphere and helioseismology (a full list can be found in the results spreadsheet in Annex 4).

Data Use	
Research	94%
Operations	18%
Other	2%
Commercial	0%

Table 4-5: Breakdown of SSR data usage.

An additional item of interest is that none of the respondents indicated the use of SSR for commercial purposes. As previously mentioned this is unsurprising given the nature of the domain. Although this is not the only consideration, it does have implications for the level of data security and overall robustness when compared to a system that must handle commercially sensitive information. *The system shall not be required to handle commercially sensitive data (UR-SSR-DC-006).*

To further understand the communities' use of existing SSR resources, each of the sub-domains was split into a set of data set types allowing classification based on the sort of observation. These types represent different sub-disciplines, which often have their own data formats and analysis tools.

The breakdown of the solar data types used by the community is shown in Table 4-6. Data usage is dominated by the space-based remote sensing observatories such as SOHO, TRACE, Yohkoh and GOES (X-ray). Data from ground-based observatories and in-situ measurements (for example from Ulysses) were also identified but at a much lower level.

Solar Data Type	
Remote Sensing (space)	70%
Remote Sensing (ground)	18%
In-situ	11%
Other	1%

Table 4-6: Data types used within the solar sub-domain.

The corresponding breakdown for STP data types is shown in Table 4-7. In this case the data from in-situ measurements such as from Cluster, Polar, Wind, ACE, Geotail and numerous other sources (see Annex 4) represent the largest fraction of data usage. However, ground-based data sets including EISCAT, SuperDARN and ground-based magnetometer chains also widely used.

STP Data Type	
In-situ	46%
Ground Based	34%
Remote Sensing (space)	11%
Geomagnetic Indices	9%
Other	1%

Table 4-7: Data types used within the STP sub-domain.

Finally, the breakdown of Planetary data types is shown in Table 4-8. The small number of returns from the planetary community limits the conclusions that can be drawn from this sub-domain. Data sets currently in use are mostly from US missions and include Clementine, Lunar Prospector, Mars global surveyor, Mars pathfinder, Pioneer, Voyager, Ulysses, Galileo, Cassini. With the exception of Ulysses, which is providing data on the interplanetary medium, there are no ESA planetary missions currently returning data. However this will change over the next decade due to an active

European programme including Cassini/Hygens, Smart-1, Mars-Express, Rossetta and Bepi-Colombo.

Planetary Data Type	
Remote Sensing	47%
In-situ (space)	28%
Ground Based	18%
Laboratory	7%
Lander	6%
Other	5%

Table 4-8: Data types used within the planetary sub-domain.

SpaceGRID shall have the ability to handle the basic data types and associated metadata used by the community (UR-SSR-FN-005). These include scalars (e.g. density), vector quantities (e.g. fields, positions and velocities), tensors (e.g. electron pressure tensors), 1d arrays (e.g. spectra), 2d arrays (e.g. images and velocity distributions), and 3d arrays (e.g. multi-spectral image cubes and 3d particle distribution measurements) (UR-SSR-DC-007). *SpaceGRID shall have the ability to handle time-series of each of the supported data types (UR-SSR-FN-006).*

The community were asked to list particular difficulties associated with accessing the data sets that they require in undertaking their research. The most common general problems identified and the corresponding SpaceGRID requirements are:-

- Difficulty in locating the source for a particular data set. *SpaceGRID shall provide a centralised facility for locating data from amongst a set of distributed data resources (UR-SSR-FN-007)*
- Ascertaining data availability. *SpaceGRID shall allow the availability of supported data sets to be checked based on a set of query conditions (UR-SSR-FN-008).*
- Data search and selection tools are not very sophisticated. *See section 4.1.3.3*.
- Data not available online (e.g. only survey plots or low quality data provided). *SpaceGRID should provide a simple and secure method for the distribution of user supplied products (UR-SSR-FN-009).*
- Access restricted. Need to make specific request to instrument team for the data and then not always readily available. *See UR-SSR-FN-009.*
- Numerous separate requests needed in order to obtain all required data sets. *SpaceGRID shall support distributed data requests (UR-SSR-FN-010).*
- Separate passwords required for each service. *SpaceGRID should provide access to multiple data services via a single sign-on (UR-SSR-FN-011).*
- Data delivered in native format that may then require post processing before it can be used. *See below.*
- Inability to preview data sets prior to download. *SpaceGRID should allow supported products to be previewed prior to download (UR-SSR-FN-012).*
- Non-availability of documentation to support the data sets and calibrations. *SpaceGRID should support the distribution of data set documentation (UR-SSR-FN-013).*

- Slow network access that sometimes limits the size of files that can be downloaded via the web. *See UR-SSR-FN-001*.

Combining data is a key factor within the SSR domain. Users were asked to identify the types of data that they combine when carrying out their research. Almost all the respondents made use of multiple data sets from the same mission and from other missions within the same sub-domain. As expected the numbers drop considerably as we move away from the core research area so that less than a quarter of the community are participating in research that involves more than one of the SSR sub-domains and only 7% are involved in interdisciplinary activities with other research domains.

Combining Data	
From the same mission	86%
From missions in same area	86%
With other activities within SSR	23%
Interdisciplinary (other domains)	7%

Table 4-9: Scope of inter-working with different data sets.

As with data accessibility, the community were asked to identify particular difficulties that had been encountered when combining different data. The most common general problems were:-

- Heterogeneous data and metadata formats. *The system shall allow standardization in the description of delivered data, i.e. metadata descriptions (UR-SSR-FN-014). The system should allow standardization in the delivery format of data from different sources (UR-SSR-FN-015). The structured nature of XML and its extensibility make it a strong candidate to address part of this problem.*
- Data provided in different units. *For supported data sets, the system shall have the ability to use the metadata information to ensure consistent use of units (UR-SSR-FN-016).*
- Inadequate, incompatible or unreliable metadata (e.g. lack of standard keywords).
- Different timescales in time series data requiring re-sampling (interpolation) of the data. *SpaceGRID shall support the manipulation of time series data including sub-sampling, smoothing and re-gridding on to a common time line (UR-SSR-FN-017)*.
- Different spatial scales and orientation in image data require re-sampling of the data. *The system shall support the manipulation of array data including, re-sampling, cropping and geometric transformation (UR-SSR-FN-018).*
- Problems with co-alignment of remote sampling data.
- Different co-ordinate systems or other reference information. *For data sets including the necessary meta-data information, SpaceGRID should support domain specific co-ordinate transformations (UR-SSR-FN-019).*
- Mapping issues (e.g. magnetic conjugacy between ionosphere and magnetosphere) when combining in-situ and remote sampling data.
- Lack of documentation on data sets, instrument characteristics and caveats information making correct use of unfamiliar data difficult.

In this section we have concentrated on data resources. As we saw from *Table 4-1* almost half the community are involved in theoretical modeling activities. These tend to be specialized tasks using dedicated resources and requiring detailed knowledge in order to configure and run the models. However, the model output can be treated in the same way as any other data resource in this context with the same issues over access and compatibility.

4.1.3.2 Data Source

A major challenge for SpaceGRID within the SSR domain is to improve existing resource location, accessibility and compatibility (see the main problem areas listed in the previous section). *To address these issues the SpaceGRID infrastructure should be linked either directly or via gateway interfaces to a number of existing data resources (UR-SSR-FN-020). The system shall provide simple mechanisms to allow additional resources to be easily added in the future (UR-SSR-FN-021).*

To identify the main data resources that are currently in use, the community were asked to list the online facilities that they most regularly use. Table 4-10 provides a selection of fifty of online archives reported in this way.

Solar System Research Data Archive	URL
ACE Science Data Center (US)	http://www.srl.caltech.edu/ACE/ASC/
AstroWeb: Solar Telescopes (France)	http://cdsweb.u-strasbg.fr/astroweb/solar.html
BASS 2000 (France)	http://bass2000.bagn.obs-mip.fr/New2001/pagef_ang.html
Centre de Données de la Physique des Plasmas (France)	http://cdpp.cesr.fr/english/index.html
Cluster Ground Based Data Centre (UK)	http://wdcc1.bnsc.rl.ac.uk/cgi-bin/gbdc/summary.pl
Cluster Science Data System, 8 data centres (1 US, 1 China), I mirror site at ESTEC/ESA (ESA)	http://sci2.estec.esa.nl/cluster/csds/csds.html
Coordinated Data Analysis Web (US, UK & Germany)	http://cdaweb.gsfc.nasa.gov/
EISCAT data base (UK)	http://www.eiscat.rl.ac.uk/
Greenland and STARE ground magnetograms servers (DK/FI)	http://www.geo.fmi.fi/PLASMA/RADAR/STARE/
HAO/MLSO coronagraph (US)	http://www.hao.ucar.edu/
HESSI Data Centre (US)	http://hesperia.gsfc.nasa.gov/hessidatacenter/
IMAGE data center (US)	http://150.144.211.77/image/image_main.html
IMAGE magnetometers (Finland)	http://www.geo.fmi.fi/image/
Imaging Riometer for Ionospheric Studies (UK)	http://www.dcs.lancs.ac.uk/iono/summary/
Institute of geophysics and Planetary Physics (US)	http://www.igpp.ucla.edu/data.cfm
ISAS Data Archive and Transmission System (Japan)	http://www.darts.isas.ac.jp/
JSOC Cluster Info (UK)	http://jsoc1.bnsc.rl.ac.uk/index.html
Los Alamos National Lab (US)	http://leadbelly.lanl.gov/lanl_ep_data/
Magnetometers - Ionospheric Radars- Allsky Cameras Large Experiment (Finland)	http://www.geo.fmi.fi/MIRACLE/
Mees Solar Observatory (US)	http://www.solar.ifa.hawaii.edu/mees.html
Multi-Experiment Data Operation Centre for SOHO (France)	http://www.medoc-ias.u-psud.fr/
National Geophysics Data Centre (US)	http://www.ngdc.noaa.gov/ngdc.html
Nobeyama Radioheliograph (Japan)	http://solar.nro.nao.ac.jp/norh/

NSO Digital Library (US)	http://www.nso.noao.edu/diglib/
NSSDC archive (US)	http://nssdc.gsfc.nasa.gov/
Oersted Science Data Center (Denmark)	http://web.dmi.dk/projects/oersted/SDC/
Planetary Data System (US)	http://pds.jpl.nasa.gov/
SOHO ARCHIVE (France)	http://www.medoc-ias.u-psud.fr/archive/
SOHO ARCHIVE (UK)	http://trace.solararchive.rl.ac.uk/soho/
SOHO ARCHIVE (US)	http://sohowww.nascom.nasa.gov/catalog
SOHO ARCHIVE SOI/MDI (US)	http://soi.stanford.edu/
SOHO ARCHIVE SUMER (Germany)	http://www.linmpi.mpg.de/english/projekte/sumer/FILE/SumerEntryPage.html
SOHO CATALOG CDS (UK)	http://solg2.bnsc.rl.ac.uk/atlas/search.shtml
SOHO CATALOG EIT (US)	http://umbra.nascom.nasa.gov/eit/eit-catalog.html
SOHO CATALOG LASCO (US)	http://lasco-www.nrl.navy.mil/database.html
SOHO CDS Database (UK)	http://solg2.bnsc.rl.ac.uk/
Solar Data Analysis Center, SDAC, (US)	http://umbra.gsfc.nasa.gov/sdac.html
Solar UK Research Facility, SURF (UK)	http://surfwww.mssl.ucl.ac.uk/surf/
Space Environment Centre, NOAA (US)	http://www.sec.noaa.gov/
SPENVIS, Space Environment Information System, (Belgium)	http://www.spenvis.oma.be/spenvis/
SPIDR (US & Russia)	http://spidr.ngdc.noaa.gov/ (http://clust1.wdcb.ru/spidr/)
SSC Satellite Situation Centre (US)	http://sscweb.gsfc.nasa.gov/
TRACE Data Archive (German mirror)	http://trace.kis.uni-freiburg.de/
TRACE Data Archive (UK mirror)	http://trace.solararchive.rl.ac.uk/soho/
UK CDHF for STP/CDC (UK)	http://cdhf.bnsc.rl.ac.uk:8080/STP_CDHF/home.htm
Ulysses Data System (ESA)	http://helio.estec.esa.nl/ulysses/archive/
US Geological Survey (US)	http://www.usgs.gov/
WDC-C1 (UK)	http://www.wdc.rl.ac.uk/
World Data Center for Geomagnetism (Japan)	http://swdcdb.kugi.kyoto-u.ac.jp/
YOHKOH Data Archive (UK)	http://surfwww.mssl.ucl.ac.uk/surf/

Table 4-10: A selection of fifty online archives used by the SSR community.

The list highlights the large number of heterogeneous sources that are in use by the community very few are under Agency control and many are located outside Europe. In addition nearly a third of respondents indicated that they had data sets that they would like to make available to other users if this were easy to do so. *SpaceGRID should provide a simple mechanism to allow users and small groups to distribute their data via the Grid (see UR-SSR-FN-021)*. However, consideration of issues such as quality, validation, metadata and documentation will also be required and are likely to limit the ease with which high quality data can be made available.

Table 4-12 lists the six different methods used for locating particular data sets and the percentage of the community that use each one.

Data Location	***% using this method***
Use domain list of links	68%
Ask colleague	59%
Search by mission	54%
Search by Instrument	51%
Search by data type	31%
Journal reference	13%

Table 4-12: Methods used for locating SSR data.

The large number and widely distributed nature of SSR data resources makes location of a required data set very difficult. Domain specific links pages are the main source of information however the information from these sources is difficult to maintain and often out of date. For example one of the main STP links resources, the Magnetospheric OnLine Directory (MOLD) was withdrawn in October 2001 for this reason. *SpaceGRID shall provided improved facilities for locating online sources of data based on a general query (see UR-SSR-FN-07).*

Table 4-13 lists the types of access control used by facilities providing access to SSR data. None of the online accessible data requires payment although significant fractions within all areas of the domain (and the majority in STP) are protected by access control. *SpaceGRID shall provide an access control system that will prevent un-authorised access to SpaceGRID resources (UR-SSR-AC-001).* Provision of access to commercial data is not required (see UR-SSR-DC-006).

Data Access	***Planetary***	***Solar***	***STP***	***All***
Public (anonymous)	64%	61%	49%	58%
Public (with access control)	14%	28%	23%	26%
Proprietary	22%	11%	28%	16%
Commercial	0%	0%	0%	0%

Table 4-13: Access control on SSR data.

4.1.3.3 Data Request

This section deals with the process of requesting data from a remote resource including the sorts of queries that are currently applied and that would be desirable. It also examines the means for submission of a request and delivery of the results.

Current facilities are usually limited in the options that they provide to constrain a data request. Table 4-14 provides a breakdown of the most common types of constraint for each of the sub-domains and for the domain as a whole. For both solar and STP the selection is usually based on time while for planetary data requests there is a much more uniform spread. The most common alternative to time or location was the use of a study number or catalogue reference identifier.

Query Constraint	***Planetary***	***Solar***	***STP***	***ALL***
Time	36%	79%	85%	77%
Other	28%	14%	13%	14%
Location	36%	7%	2%	9%

Table 4-14: Query constraint usage

The community were asked to identify any additional selection criteria that would be beneficial to their research activities. The responses resulted in the following user requirements:-

- *SpaceGRID shall allow full catalogue queries on multiple fields (e.g. time and position and wavelength) (UR-SSR-FN-022).*
- *SpacegRID shall allow cross-catalogue searches (e.g. the ability to identify times where all the requested datasets are available or meet some criteria such as being in the same region) (UR-SSR-FN-023).*
- *SpaceGRID shall support searches using data parameters to identify particular periods that meet some criteria and then only return data from those periods (e.g. when the interplanetary magnetic field is southward) (UR-SSR-FN-024).*
- *SpaceGRID should support free text search on metadata and documentation (UR-SSR-FN-025).*
- *SpaceGRID may support fuzzy queries and pattern recognition (e.g. all spectra that contain at least 60% of a specified set of lines) (UR-SSR-FN-026).*

A breakdown of the mechanisms used for requesting data are listed in Table 4-15. Online web based requests is the most readily accessible and also the preferred method for the majority of users. *SpaceGRID shall provide a web-based mechanism for requesting data (see UR-SSR-FN-032).*

Request Method	
Web based	77%
Ask colleague	13%
Other	7%
Automated E-mail	3%

Table 4-15: Preferred data request mechanisms

A similar breakdown for the data delivery mechanisms is shown in Table 4-16. Online web based retrieval remains the most popular method although the ability to login and fetch data via ftp was also popular due to the ease with which multiple files can be transferred in a single operation, a feature that is often not provided via web interfaces. SpaceGRID will require an offline delivery mechanism to support the output from complex queries that may take a considerable time to execute. *SpaceGRID shall support web-based and offline (e.g. ftp) mechanisms for data delivery (UR-SSR-FN-027).*

Delivery Method	
Web based	57%
Ftp (get from server)	28%
Other	9%
Ftp (put from server)	4%
E-mail	2%

Table 4-16: Preferred data delivery mechanisms

The number of requests made per year varied considerably between ten and tens of thousands with a typical number being several hundred.

Data volumes per request also covered a large range from much less than a megabyte to several gigabytes, the average size being in the range 10 to 100 MB. *A SpaceGRID node should maintain at least 50 GB of storage for caching of temporary results and workspace (UR-SSR-RS-001).*

When asked if it would be beneficial if some or all of the processing could be done either at the archive or on a remote data processing resource prior to the data delivery; 63% thought that this would help their research. *SpaceGRID should allow pre-processing of requested data prior to download (UR-SSR-FN-028).*

Most users wanted data to be delivered online “while-you-wait” but would accept longer delays, up to a few days, in the case of large or complex requests.

4.1.3.4 Data Processing

The hardware and operating system(s) configurations reported by the community consisted mainly of Unix workstations (a variety of makes), Linux and Windows PC’s and one or two Mac and VMS users. The Interactive Data Language (IDL) produced by RSI and to a lesser extent Matlab produced by MathWorks are widely used throughout the domain for data processing, analysis and visualisation.

The core SpaceGRID services shall be designed to operate under Linux and Sun/Solaris (UR-SSR-DC-001).

The [SPACEGRID-SOW] requires that the middleware used within the core SpaceGRID services be open source (UR-SSR-DC-002).

The usage of the main standard data formats for local storage is listed by sub-domains and as an overall figure in Table 4-17. For each format the percentages indicate the fraction of users within the specified area who use that format. Of these standard formats, FITS is most popular with the solar community and ASCII with the planetary and STP communities. The NASA CDF format is mainly used by the STP community, the figure under the planetary community is misleading since it is the due to a respondents who are involved in both planetary and STP research. Most planetary data originates from the US Planetary Data System and so conforms to the PDS standard. This specifies a label that contains a rich set of metadata fully describing the data set and making use of a comprehensive set of keywords as defined in the PDS data dictionary. The data itself may be stored as either ASCII or binary and may require specialised software to extract it.

Local Storage Format	***Planetary***	***Solar***	***STP***	***All***

ASCII	73%	31%	87%	50%
FITS	0%	86%	17%	55%
CDF	36%	13%	40%	19%

Table 4-17: Standard formats used for local data storage

The corresponding results for data exchange are shown in Table 4-18. These show the same trend as the local storage figures. The numbers are generally lower indicating less data exchange than download. The most noticeable difference is the exchange of solar data in FITS format.

Exchange Format	***Planetary***	***Solar***	***STP***	***All***
ASCII	56%	24%	64%	35%
FITS	0%	35%	13%	22%
CDF	18%	10%	23%	12%

Table 4-18: Standard formats used for data exchange

The respondents did not identified any other what could be considered standard data formats that are in widespread use by the community. Of the other formats reported, IDL save-sets are widely used as a quick and simple means for storing and retrieving complex data sets processed within IDL. Other formats listed by a small number of users included image files, and a variety of specialised (often binary) formats often developed in conjunction with particular instrument or mission teams data processing systems.

The ASCII, CDF and FITS formats are sufficiently widely used within the community that all three should be supported for reading and writing by the SpaceGRID system (UR-SSR-DC-003). ASCII is actually a representation rather than a single data format. In this situation we are referring to tabular data such as fixed format or delimiter separated fields that can readily be accessed generic read routines.

In most cases the retrieved data has to undergo some form of manipulation before it can be used with local data analysis applications. The two most common forms of post processing are

- Reformatting to allow the data to be read into an analysis program
- Calibration and cleaning of the data.

The most commonly reported community supported toolkits and processing applications are:-

- Planetary: ISIS
- Solar: SolarSoft (IDL based), IRAF
- STP: IDFS, ISDAT, OVT, PAPCO, QSAS

In addition, there are a huge number of specialised applications being used and developed in support of particular instruments, missions or as part of research collaborations. With the exception of SolarSoft, which is very widely used within the solar community, there is little standardisation of applications even within the sub-domains.

4.1.4 Looking for Standardisation

The community were asked to identify the aspects of data analysis and research that are currently hindered by lack of standardisation amongst the various datasets, their metadata (catalogues) or datafile formats? The responses echoed the problem areas identified in section 4.1.3.1. In particular:-

- Too much time is spent trying to locate data and find periods/events when all the data sets necessary for a particular analysis are available.
- Too much time is spent on getting data into compatible formats where it can be combined with other data sets. These activities include format translation and re-sampling data onto a common timeline so that derived products can be generated.
- Too much time is spent on data that has incomplete, incompatible or incorrect metadata.
- The lack of efficient database management and data mining solutions means that too much time is spent on the development of ad-hoc analysis software (duplication of effort).

The community were also asked to identify the greatest impediment to working with multi-instrument datasets. The most common areas of difficulty are:-

- The learning curve for using unfamiliar data and their associated software is very steep making them difficult to use correctly without assistance from the instrument team. When using a large number of datasets this can be very time consuming.
- Lack of user-friendly software, user guides ("cookbooks") and calibration documentation.
- Restricted access.
- The same data formatting, metadata and re-sampling issues described above.

4.1.5 Collaborative Environment

The SSR community reported a very high degree of collaboration, with 88% of the respondents reporting an active involvement with other institutions. However, the size of these collaborations is generally quite small typically involving a core of two to four individuals with each being responsible for the provision and interpretation of a particular data set or model. The primary mode of communication for these research activities is E-mail, phone and direct contact at meetings, workshops and conferences.

Distribution of processed data products is mostly achieved through the use of local web sites and ftp servers. Configuration control of distributed products is informal, based on e-mail notification or personal contact to inform of updates.

When asked if any use had been made of concurrent collaborative communications methods (see Table 4-19) only telephone-conferencing was identified as a technique that a significant minority of the community had tried. This is also the only technique that is being used regularly.

Communication methods	
Tele-conference	46%
Video-conference	10%
Net meeting	4%
AccessGRID	0%

Table 4-19: Concurrent collaborative communication methods

The relatively low use of the newer technologies can be explained by a number of related factors:

- The numbers involved in collaborations are small and therefore direct contact via phone and e-mail remains an efficient means of communication.
- The community does not have the necessary analysis tools to support efficient concurrent collaborations. A lot of processing is undertaken using specialist code that requires expert knowledge to configure, modify and run in order to obtain the required results.
- Video and net meeting facilities are not yet common at the desktop. Although many institutes have access to these facilities, they must be booked and involve a level of overhead that can be prohibitive for small collaborations.

However, a significant proportion of the community, 78%, felt that simultaneous access to large data sets would be beneficial to their research activities.

4.1.6 Baseline Requirements Priorities

The community were asked to prioritise a set of general GRID based functional requirements that had been identified by the domain specialist as part of the development of a baseline. These represent the core requirements for the system.

- *The system shall provided improved facilities for locating online sources of data based on a general query (see UR-SSR-FN-007).* This will simplify the process of locating distributed resources via knowledge based search and query tools.
- *The system shall allow standardization in the description of delivered data, i.e. metadata descriptions (see UR-SSR-FN-014).* This standardization of data descriptions will assist in the production of uniform responses to queries and the handling of data within analysis and visualization applications.
- *The system should allow standardization in the delivery format of data from different sources (see UR-SSR-FN-015).* The standardization of data delivery formats will reduce the post processing required to allow the actual use the data. As with the previous requirement it will simplify the development of analysis software. It is not realistic for SpaceGRID to specify a single delivery format for use by all domains. A possible solution is to use a standard internal format and translators to convert between this and the most common formats used within each sub-domain.
- *The system should maintain good compatibility between itself and other related GRID initiatives being developed at both national and international level (UR-SSR-FN-029).* This is to avoid duplication of effort and to maximize the resources that are available through the system. For example compatibility with EGSO should provide a uniform interface into a large-scale federation of space and ground-based solar observations.

- *The system shall provide improved facilities for querying the catalogues of a single data archive (see UR-SSR-FN-022).* Most existing archives provide some mechanism for accessing data based on information contained in a set of catalogues. However, this functionality is often limited to picking a particular data set and time range. SpaceGRID should be able to extended catalogue search capabilities, for example by providing the ability to requests periods for which data is held for two or more data sets within the archive.
- *The system shall provide the ability to apply queries on data parameters within a single archive (see UR-SSR-FN-024).* This extends the query capability to allow for a search to be made using the data as well as the catalogues. A request might ask for periods when the solar X-ray flux is above some threshold.
- *The system shall be able to query transparently the catalogues of multiple distributed data archives (UR-SSR-FN-030).* This is a further extension to the catalogue query that automatically handles queries applied across a number of distributed data archives.
- *The system shall be able to query on data parameters across multiple distributed data archives (UR-SSR-FN-031).* This is the most complex form of query and is identified as a separate requirement due to the complexity of efficient implementation of such a query.
- *The system shall allow the user to manipulate and process data remotely prior to download (see UR-SSR-FN-028).* Some of the data sets within the Solar System Research too large for efficient download and processing on standard user equipment. The ability to apply some level of processing at the data source and then just return the results should be much more efficient than downloading the data for local processing. This should be particularly true of large-scale statistical studies that may involve processing of a significant fraction of the total archive.
- *The system shall provide a web portal to access the distributed resources attached to the Grid (UR-SSR-FN-032).* This feature offers simple web based entry points into the SpaceGRID system. The advantage being that access can be provided to a range of facilities without the need for the user to download and install specialized software.
- *A GRID server application shall be provided that allows users to link their own data into SpaceGRID (see UR-SSR-FN-021).* Many SSR missions are PI based with the individual instrument teams being responsible for the reduction and dissemination of their data. A SpaceGRID server application should provide an effective means of federating these data.
- *A software library should be made available to allow programs access to SpaceGRID facilities (UR-SSR-FN-033).* This feature would allow existing tools and analysis software to be GRID-enabled. Such applications would then have access to the data and resources provided by the SpaceGRID infrastructure.
- *The system shall provide an online collaborative working environment – virtual SpaceGRID meeting facility (UR-SSR-FN-034).* A collaborative environment draws together some of the other features described above into a system for concurrent collaborative analysis including features such as the near real time distribution of data and visualization.

The results of the prioritisation for the whole domain are shown in Figure 4- 1. Respondents were asked to provide ratings on a scale of 1 to 5. To allow inter-comparison each of these sets of results was renormalized on a scale between 0 and 1 based on the minimum and maximum values. The average value for each requirement was then calculated and plotted.

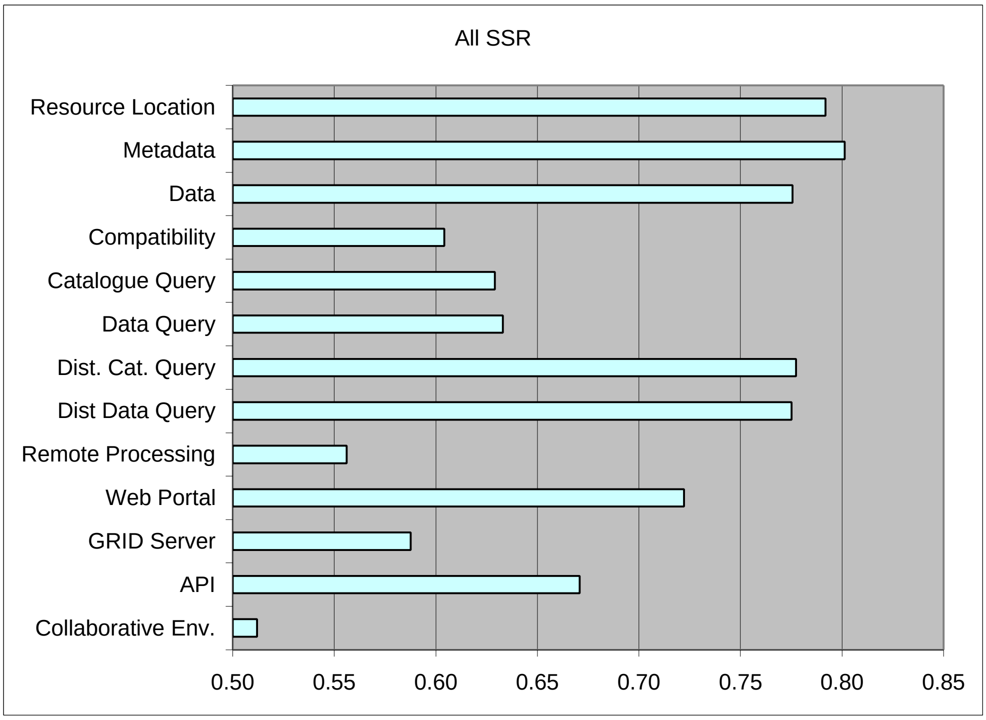

Figure 4- 1: Priority of baseline requirements

The highest rated requirements were those dealing with the identification, distributed querying and standardised delivery of SSR data, which is consistent with problem areas identified by community in earlier sections. The lowest overall rating went to the concurrent collaborative working environment (though there were a few users who identified this as their top priority). As was seen in section 4.1.5 there is currently little experience of this technology within the community despite the availability of facilities like SPARC and UARC. Other factors are the relatively small numbers involved in each study, the large number of disparate studies and that much of the analysis is not well suited to concurrent analysis. We should emphasise that these comments are directed towards the need for a concurrent facility, rather than the more general collaborative environment defined by the other requirements.

The breakdown of the results by sub-domain is given in Figure 4- 2. In general similar results are obtained across the three areas. We do not draw any particular conclusions about the variations in the planetary results due to the small sample. The lower priority from the solar community on data delivery standardisation is consistent with this being less of a problem due to the widespread use of FITS and SolarSoft. By contrast, the solar community identified better catalogue search capabilities as a higher priority and the generation of these catalogues is one of the key activities that EGSO is planning to address. The high STP ratings for a web portal and GRID server are the result of the large number of distributed data sets that are available in this area.

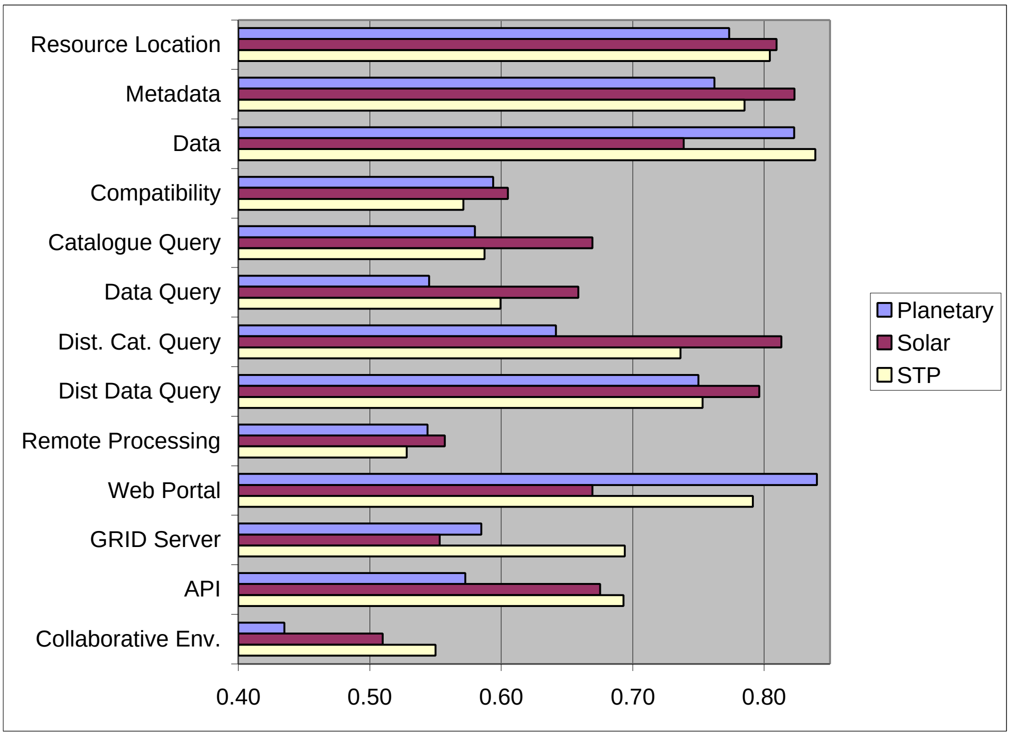

Figure 4- 2: Priority of baseline requirements by SSR sub-domain

The community was also asked to identify any high priority items that they felt were not covered by the baseline requirements. *The main addition was that the system should have the ability to interface to models (UR-SSR-FN-035).* Specifically, that 1) models should be accessible via the GRID, 2) that it be possible to drive the models using data from the GRID, and 3) that the data from the models to be accessible through the same interfaces as for other data resources.

The community were asked to identify what they consider to be the single most important, and therefore highest priority items for SpaceGRID to address in the SSR domain. The various responses can be reduced to two broad functional categories.

- The highest priority, by a long way, is improved facilities for locating, querying, previewing and delivering multiple data sets in a simple and standardised way.
- The provision of high power central processing, data storage and fast network access that can be used to provide online data processing and data mining capabilities.

4.2 General Description

The Solar System Research domain consists of three distinct user communities that have developed independently from one another. Although there are some overlaps between the different sub-domains and with other domains (such as EO) these are not common and the majority of the community are working within their own field. Even within the sub-domains there is a considerable variation in the level of standardisation.

The requirements analysis identified that the GRID related activity that will most benefit the SSR community is the provision of a general infrastructure to support collaborative work through improved facilities for location and data interoperability. This would help to address the current problems of heterogeneous data formats, lack of standard applications and widely distributed data resources. The reasons for concentrating on the development of the infrastructure rather than specific applications are:-

- We need the infrastructure before applications can be developed to take advantage of it. While re-use of work from other initiatives such as the EU Data GRID and AVO project should play a part in this, there are particular SSR issues that need to be addressed.
- The emphasis on SSR is distributed access to a diverse set of existing facilities. This is different to the problems facing some other domains which are technology driven e.g. by very large data volumes.
- A lot of the work in SSR is based on the temporal or spatial evolution of some property. This is different to some other domains which are much more object oriented e.g. observations of an astronomical object or a particular run from the LHC.
- Provision of a general infrastructure in support of the whole of SSR will provide the greatest benefit to the community as epitomized by the CDAWeb system in the US.

Existing ESA infrastructure in this area is currently very limited with most of the current facilities in use by the community being outside the control of the agency. However, there is increasing support for the planetary sub-domain in the form of a centralised ESA planetary archive that will support the large number of upcoming ESA planetary missions. This will represent a major European resource, though timescales for the supported missions put the availability of anything other than test data beyond the scope of the SpaceGRID study.

Although the requirements analysis was focused on data from measurements it was clear from the responses that access to data from models is also needed. Models are important both as a way of extending science analysis and as a tool for operations, e.g. using models of the Martian thermosphere to estimate drag effects on Mars orbiters. Most analyses of data involve relating them to a model of the situation. Driving models with time-stream of real data, and interaction between models of different domains (e.g. solar-wind/magnetosphere) to simulate the real are just two potential areas that could be supported by the SpaceGRID (i.e. via the API, UR-SSR-FN-033).

The larger than expected response from the community indicates a clear interest in the potential of what can be delivered by a GRID based infrastructure. However, the three main comments when it came to discussion of the implementation were 1) “keep it simple” so that the facilities can be accessed by a broad spectrum of the community, 2) “easily upgraded” so that it can take on more functions and abilities as it grows, and 3) the needs of the community should drive the technology and implementation rather than the other way around.

4.3 REQUIREMENTS SUMMARY

To assist with mapping of the user requirements to particular GRID technologies and also to help with the identification of common requirements, one or more qualifiers may be provided. This qualification of requirements is orthogonal to the normal groupings described earlier. The values taken by **qual** are as follows:-

- CM - Common (to suggest some requirements as common to more domains)
- DI - data Intensive
- CI - communication intensive (i.e. need for a high speed and/or large bandwidth networks, maybe satellite communications, etc.)
- DP - distributed processing
- HT - high throughput
- LM - LCMPP
- OD - on-demand processing
- CE - collaboration environment
- PR - portal

4.3.1 Functional Requirements

UR-SSR-FN-001	CI, PR	The SpaceGRID design should ensure good accessibility, by the SSR community, to the system's services (e.g. through a distributed architecture, use of data caches, mirrors & replicas, and good links onto national and international research networks).
UR-SSR-FN-002	CE	The core SpaceGRID SSR infrastructure shall be general such that it can be used to support research activities within all three areas of SSR as well as inter-domain and interdisciplinary studies.
UR-SSR-FN-003	DI	SpaceGRID should have the capability to handle science planning and operations products such as inputs from models, orbit predictions and planning events.
UR-SSR-FN-004	PR	The system shall provide services in support of both novice and expert users.
UR-SSR-FN-005	CE	SpaceGRID shall have the ability to handle the basic data types and associated metadata used by the community.
UR-SSR-FN-006	CE	SpaceGRID shall have the ability to handle time-series of each of the supported data types.
UR-SSR-FN-007	CE, DP	SpaceGRID shall provide a centralized facility for locating data from amongst a set of distributed data resources.
UR-SSR-FN-008		SpaceGRID shall allow the availability of supported data sets to be checked based on a set of query conditions.
UR-SSR-FN-009	CE	SpaceGRID should provide a simple and secure method for the distribution of user supplied products.
UR-SSR-FN-010	DP	SpaceGRID shall support distributed data requests.
UR-SSR-FN-011	DP, CM	SpaceGRID should provide access to multiple data services via a single sign-on.

UR-SSR-FN-012	CE	SpaceGRID should allow supported products to be previewed prior to download.
UR-SSR-FN-013	CM	SpaceGRID should support the distribution of data set documentation.
UR-SSR-FN-014		The system shall be able to deliver standardized information describing an online resource or delivered data, i.e. metadata descriptions.
UR-SSR-FN-015	DI	The system should allow standardization in the delivery format of supported data from different sources. Unsupported data formats will be delivered as native files.
UR-SSR-FN-016	CE	For supported data sets, the system shall have the ability to use the metadata information to ensure consistent use of units.
UR-SSR-FN-017	DI	SpaceGRID shall support the manipulation of time series data including sub-sampling, smoothing and re-sampling on to a common time line.
UR-SSR-FN-018	DI	The system shall support the manipulation of array data (e.g. images) including, re-sampling, cropping and geometric transformation.
UR-SSR-FN-019	DI	For supported data sets (i.e. those including the necessary meta-data information), SpaceGRID should support domain specific co-ordinate transformations.
UR-SSR-FN-020	DP	The SpaceGRID infrastructure should be linked either directly or via gateway interfaces to existing data resources.
UR-SSR-FN-021	CE	The system shall provide simple mechanisms and tools to allow additional resources to be easily added in the future.
UR-SSR-FN-022	HT	SpaceGRID shall allow full catalogue queries on multiple fields (e.g. time and position and wavelength).
UR-SSR-FN-023	HT	SpaceGRID shall allow cross-catalogue searches and return a consolidated response (e.g. the ability to identify times where all the requested datasets are available or meet some criteria such as being in the same region).
UR-SSR-FN-024	HT, DI	SpaceGRID shall support searches using data parameters to identify particular periods that meet some criteria and then only return data from those periods (e.g. when the interplanetary magnetic field is southward).
UR-SSR-FN-025		SpaceGRID should support free text search on metadata and documentation.
UR-SSR-FN-026	HT, DI	SpaceGRID may support fuzzy queries and pattern recognition (e.g. all spectra that contain at least 60% of a specified set of lines).
UR-SSR-FN-027	HT	In addition to web-based delivery, SpaceGRID should support an offline delivery mechanism for bulk transfers (for example ftp or scp).
UR-SSR-FN-028	DP, CE	The system shall allow supported data products to be manipulated and processed either separately or as part of a query, using pre-defined and user up-loaded functions, and pipelines constructed from these functions.
UR-SSR-FN-029	CE	The system should maintain good compatibility between itself and other related GRID initiatives being developed at both national and international level.
UR-SSR-FN-030	DP	The system shall be able to query transparently the catalogues of multiple distributed data archives.
UR-SSR-FN-031	DP, HT, CI	The system shall be able to query on data parameters across multiple distributed data archives.
UR-SSR-FN-032	PR	The system shall provide a web portal to access distributed resources from a single web site.
UR-SSR-FN-033	CI	A software library should be provided to allow user applications access to SpaceGRID facilities.

UR-SSR-FN-034	CE	The system shall provide facilities that can be used to promote collaborative working (e.g. bulletin boards, news systems and Wiki-webs) and may include a concurrent collaborative working environment.
UR-SSR-FN-035	HT, DI, LM	The system should have the ability to interface to models.
UR-SSR-FN-036	CM	The system shall allow the user or administrator to monitor and modify (e.g. cancel, change permissions) resources, queries or processing task for which they are responsible.

4.3.2 Performance Requirements

UR-SSR-PR-001	PR	The system should provide feedback of an action within 30s
UR-SSR-PR-002	PR	The system should complete simple, online (e.g. single catalogue, preview) tasks within an average time of 1 minute.
UR-SSR-PR-003	DI, CI	The system should complete complex, offline (e.g. distributed, multi-catalogue and data manipulation) tasks within an average time of 24 hours.
UR-SSR-PR-004	CE	The concurrent collaborative environment shall provide near-real-time access to SpaceGRID resources.

4.3.3 Resource Requirements

UR-SSR-RS-001		Each SpaceGRID node should support at least 50GB of storage for caching of temporary results.

4.3.4 Distribution Requirements

UR-SSR-DS-001	CE	The SpaceGRID server (hub) application shall be made freely available to data providers.
UR-SSR-DS-002	CE	The SpaceGRID API shall be made freely available to application developers.
UR-SSR-DS-003	CE	A SpaceGRID archive interface or gateway interface shall be installed at the archive or associated SpaceGRID hub for each facility linked in to the SpaceGRID infrastructure.

4.3.5 Access Requirements

UR-SSR-AC-001		The system shall provide an access control system that will prevent un-authorised access to SpaceGRID resources.
UR-SSR-AC-002	CM	The system shall provide a centralised user level authentication system.
UR-SSR-AC-003	PR, CE	The system shall support single sign-on and delegated authentication for multi-site and offline activities.
UR-SSR-AC-004	CE	The system shall provide an access control interface to map between SpaceGRID authentication credentials and the local archives authorisation system.
UR-SSR-AC-005	PR	For un-authenticated users, the system shall fallback to anonymous privileges.
UR-SSR-AC-006	CE	The user should be able to define and modify the access rights on an uploaded resource.

4.3.6 Design Constraint Requirements

UR-SSR-DC-001	CM	The core SpaceGRID services (hub and API) shall be designed to operate under Unix/Linux.
UR-SSR-DC-002	CM	The [SPACEGRID-SOW] requires that the middleware used within the core SpaceGRID services be open source.
UR-SSR-DC-003	CE	As a minimum the system should support the following standard formats for data delivery; ASCII, CDF, FITS and XML.
UR-SSR-DC-004	CE	As a minimum the system should support the following standard formats for information delivery; HTML, XML
UR-SSR-DC-005		As a minimum the system should support standard arithmetic, co-ordinate conversion and timeline interpolation functions.
UR-SSR-DC-006		The SSR SpaceGRID shall not be required to handle commercial data
UR-SSR-DC-007	DI	The basic data types referenced in (UR-SSR-FN-005 & UR-SSR-FN-006) should include scalars (e.g. density), vector quantities (e.g. fields, positions and velocities), tensors (e.g. electron pressure tensors), 1d arrays (e.g. spectra), 2d arrays (e.g. images and velocity distributions), and 3d arrays (e.g. multi-spectral image cubes and 3d particle distribution measurements)

4.4 Index of User Requirements

ANNEX 1 - Questionnaire and Guidance Notes

This annex contains a copy of the web-based questionnaire used to identify the current use and future needs of the Solar System Research community. In addition to the questionnaire are the guidance notes that accompanied it.

Note: This annex is contained in a separate PDF file SGD-SSR-RAL-REQ-001-ANNEX-1.pdf

ANNEX 2 – E-mail Distribution List

This annex contains a copy of the Solar System Research E-mail distribution list used to notify the community about the questionnaire. The list was a combination of Solar and STP lists based on existing lists for these groups. At the time of the initial notification (15/01/2002) a list for the planetary community was not available and so the project scientists on Bepi-Colombo, Mars-Express and SMART-1 were invited to forward the questionnaire to their project distribution lists and also the UK Planetary Forum distribution list. A reminder was sent to the same list on 01/02/2002. A final reminder on 01/03/2002 to an extended list including e-mail addresses of between one and two hundred planetary scientists working on current ESA missions. It is this final list that is contained in this annex.

Note: This annex is contained in a separate PDF file SGD-SSR-RAL-REQ-001-ANNEX-2.pdf and is not intended for public release.

ANNEX 3 - Responses to Questionnaire

This annex contains the responses provided by the Solar System Research community to questionnaire described in Annex 1. The responses are in the form of the html formatted page that was returned to the user on submission of their input. In one case (CDPP) the input was received as a completed version of the MS Word version of the questionnaire.

Note: This annex is contained in a separate PDF file SGD-SSR-RAL-REQ-001-ANNEX-3.pdf and is not intended for public release.

This Annex is almost 300 pages in length so we do not recommend printing the entire document. Instead use the spreadsheet version provided in Annex 4 and then use the reference ID to select the associated bookmark in this Annex using a PDF viewer.

ANNEX 4 - Spreadsheet of Questionnaire Responses

This annex contains the output of the spreadsheet of responses used to help analyse the inputs provided by the community. It contains most of the entries from the questionnaire except for the multi-line responses. For the complete inputs please refer to the individual responses provided in Annex 3.

Note: This annex is contained in a separate PDF file SGD-SSR-RAL-REQ-001-ANNEX-4.pdf and is not intended for public release.