Coordination Action for the integration of Solar System Infrastructures and Science

Project No.: 261618
Call: FP7-INFRA-2010-2

Use Cases defining the Required Capabilities

Version 1.2

Title:	**Use Cases defining the Required Capabilities**
Document No.:	CASSIS-UCL-D3-005-d1 *Deliverable:* **3.1.2**
Date:	20-August-2013
Editor:	R.D. Bentley (UCL-MSSL)
Contributors:	D. Derghmans (ROB), K. Steed (KU Leuven)
Distribution:	Project

CAPACITIES

e-infrastructure

Revision History

Version	Date	Released by	Detail
0.5	28-Mar-2012	R.D. Bentley	First Draft
0.9	12-Jun-2012	R.D. Bentley	First Release – for discussion with community
1.0	23-Jul-2012	R.D. Bentley	
1.1	20-Jun-2013	R.D. Bentley	Submitted version
1.2	20-Jun-2013	R.D. Bentley	With corrections

Contents:

Overview

The Coordination Action for the integration of Solar System Infrastructure and Science (CASSIS) is a FP7 project that builds on work undertaken in other projects – primarily HELIO, Europlanet RI and SOTERIA. The objective of CASSIS is to finds ways to facilitate science within the Solar System by improving the interoperability of metadata and services through the use of standards.

This document contains a set of use cases defining the capabilities required to support cross-disciplinary research in heliophysics and intersecting disciplines. The use cases are designed to flush out issues that need to be addresses in the steps we are taking to improve interoperability.

Introduction

The desire to study science problems that span disciplinary boundaries is hindered by the difficulty of using the datasets together. Phenomena that are interrelated cannot be easily compared because the communities involved have evolved independently over decades without any consideration of the need for interoperability. The capabilities share less common terms or parameters than they could if a decision were made to try to standardize parts of the metadata associated with files, derived information and services. With technology no longer being a restriction to sharing data, the important thing is to understand how to make things more interoperable.

In this document we examine the types of Solar System science that might be facilitated if it were easier to use the data and services from the different communities. We have tried to identify science Use Cases that span disciplinary boundaries and help highlight what capabilities will work together and what needs to be changed in order to improve interoperability.

The Use Cases that have been developed in consultation with the user community. In the discussions, the users were not asked to limit their ideas by taking in account constraints such as available resources, implied complexity, or broader usefulness to the community but to consider how they would like to work. For this reason, the requirements enumerated in the document should not be taken as a fixed set of agreed goals for any one project but rather as a framework in which a number of projects could collaborate.

Through the Use Cases we derive a set of requirements that respond to the needs of the "end-users" or consumers of the system. These are the users that will need to request information and data and use software or other resources from the broad set of resources that for a shared research environment.

This document represents a snapshot of where we currently are and as we continue to work with the community it will continue to evolve. The requirements that will emerge from this study have implications on Deliverable D2.1 and D2.2 in that there will be an implied need to examine services and metadata from other communities. They will also affect D2.3 (Recommendations).

Objective of the Use Cases

The objective of CASSIS is to finds ways to permit scientists to work more efficiently over a wider rage of disciplines in the manner in which they are currently accustomed to in areas that that are familiar with. This objective should be realized by facilitating interoperability through the use of current, enhanced and new standards.

Our primary method to determine what functional, architectural, and scientific capabilities are needed in order to address the research topics presented by the user was to collect a series of scientific Use Cases. These describe the processes used by scientists in trying to identify events or phenomena of interest in the collected data stores and then select and retrieve the data related to these events.

The goal of the deliverable is not to exhaustively document the breadth of research topics within Solar System Science, rather, we intend to extract from these use cases the general methods used by researchers, allowing us to better understand the logical sequence of events that takes place during the target identification and data selection phases of a research project.

In order to address a generic problem of Solar System science, the Use Case collection takes into account the observations needed to build a picture in multi-dimensional parameter space. We have a reasonable understanding of what this means in heliophysics[1] but need to talk with groups from other communities to gain an understanding in their domains.

[1] In heliophysics this would mean that is the dataset requested by the users can be related to a broad range of spatial (x,y,z or Q,F,r) and temporal (t) coordinates, different temperature, energy and density regions, and various measures of polarization.

The Use Cases Gathered

We have gathered use cases in discussion with people inside and outside of the project with the aim of covering a broad range of problems in Solar System science. In this way, we have collected a list of different scientific goals and research methodologies that form a sufficiently representative set with which to start to define CASSIS user and science requirements.

The Use Cases that were gathered are summarized in Table 1 and detailed in Appendix 1. The coverage of the collected use cases in terms of data source and dimensions, instrument types, spectral and spatial regions and Solar System science topics are given in Appendix 2.

Many of the use cases describe the scientific selection processes as presently undertaken. They address existing datasets and capabilities albeit with difficulties due to the need to personally interact with multiple data centres and to "manually" combine datasets.

We have also developed use cases that describe processes that are not practically feasible within the current situation of disconnected and incompatible datasets – some of these relate to the planned Solar Orbiter mission. The new Use Cases may require new approaches to managing and analysing data scattered around the world.

The basic ideas of several of the use cases have come from the CASSIS Team. These are then expanded in discussion with specialists from the relevant communities through one-on-one meetings and Community Coordination Meetings. The ideas represent a starting point for the discussion and in some cases the original ideas will be replaced by use cases thought to be more appropriate.

No.	Title	Author
1	Comparison of Coronal and In-situ Solar Wind Diagnostics	G. Poletto
2	Link remote-sensed and in-situ observations in the heliosphere	D. Berghmans & K. Steed
3	Find occurrences when sites are connected by heliospheric propagation	D. Berghmans & K. Steed
4	Relate occurrence of planetary aurorae with solar activity	R.D. Bentley & G Branduardi-Raymont
5	Using observations from L5 to predict terrestrial Space Weather	R.D. Bentley & A. Opitz
6	Relate terrestrial ionospheric conditions with other observations	D. Kataria & R.D. Bentley
7	Variation of the Cosmic Ray flux	R.D. Bentley
8	Influence of space weather effects on Aeronomy studies	D. Berghmans
9	Effect of solar activity on upper atmospheric layers and weather	[R.D. Bentley]
10	Effect of long-term solar variability on the Earth's climate	[R.D. Bentley]
11	Effect of solar activity on terrestrial climate variability	[R.D. Bentley]

Table 1 Scientific Use Cases Titles

Analysis of the Requirements

The Use Cases that have been collected have been examined to provide requirements that could help formulate the architectural design of a collaborative environment, the user interface and the search engine as well as indications of new types of metadata that may be needed. However, many members of the CASSIS Team have been involved in developing infrastructures that span disciplinary boundaries in the past and from this experience we know that gathering use cases does not provide a complete picture of what is required and certainly yields little about what is needed at the system level.

In part, the reason for this is that people find it difficult to think outside of the box; their suggestions are constrained by what they know to be possible and believe might be possible. It is also because most users posses little understanding of what lies inside the services that they use – nor should they have to – and the consequence is that they are unable to provide certain types of requirements.

The experience of the Team provides an insight of what is required and based of this we have drawn up an underlying set of general requirements – described in the section *"Underlying General Requirements"* – that can help us understand how the overall architectural design, user interface and search capabilities need to be formulated to facilitate interoperability over several domains.

The use cases that were gathered have been analysed in order to define the techniques of identification and data selection process most commonly applied. From this we have tried to define a common process – described in the section *"Generalized Interaction Sequence"* – in order to provide a better understanding of how capabilities need to interact. We also have a preliminary understanding of the generalized metadata requirements for such a system.

In essence, our experience gives us the general picture of what is needed while the science use cases that have been gathered push the boundaries and help identify capabilities that may differ in some way from this that have been provided in the past.

Underlying General Requirements

Based on experience gained from working on previous projects, we know that use cases do not always flush all that is relevant to developing a system. Collectively we have a good feel for what is required and in this section we describe other, underlying requirements that need to be considered.

As we have talked with different communities we have started to understand the scale of the differences as we move away from heliophysics. HELIO and SOTERIA are considering very similar things – HELIO looks at the effects of the Sun on the Solar System whereas SOTERIA considers things on the Sun that could affects the Earth. Europlanet RI works in a different type of regime – the research infrastructure it is constructing has to support several disciplines under the umbrella of planetary science.

In the Earth sciences similar discuplines are treated much more independently although they use of an ISO standard for data helps integration. However, considering how these communities might be incorporated into a as collaborative environment ensures that we take a very generalized approach to interfaces and other standards and not hard wire things to thinking within the three projects.

Several issues raised here are related to other Deliverables but we discussed them briefly here in order to try to pull things together.

Issues relate to Data

Accessibility of the data

Not all the data identified in the Use Cases, or that might be used, are accessible. In addition, even where the data are accessible, the nature of the access is an issue with some providers being much more capable than others.

For data that are on-line, *an intelligent (resource rich) data provider is most useful in terms of interoperability*. In this case, the provider basically gives the user what they need in response to a simple query without the user needing to know much about the instrument concerned; included in this category are virtual observatories (VO).

Where data are only available through resource poor providers – that is through a simple data archive – the *archive should be well simply structured following a hierarchical tree based on date*; the granularity of the branches and time interval contained in a file should depend on the amount of data being supplied.

There may be a case for centralizing small datasets into an archive in order to facilitate access. Doing this may require less effort than trying to provide adequate network access and would also simplify the use of the data, however, this may imply the need for a service that would handle various issues including the ability to upload and “register” the dataset. It is also possible to funnel access through a VO and let it provide virtual centralization of the data.

It *should always be easy to identify a data file* and the *file name should be sufficiently unique that the file is able to exist without conflict outside of the environment that it is normally stored in* by the instrument team. The name should in some way identify the instrument and observatory that the data relate to, together with an indication of the date and type of the observation.

For complex data sets *it desirable that the data remain connected to the site where the expertise about the instrument resides*. That said, it is important that the documentation describing the meaning and use of the data are sufficient to allow the data to continue to be used after the operating phase on an instrument has ended.

The need to provide high quality, intelligent access is important for cross-disciplinary studies where the user may not know a lot about the data and needs as much help as possible.

In some cases the quality of access could be improved if the providers are willing to follow some simple rules. *Guidelines on improving access to data archives are discussed in other CASSIS Deliverables.*

Data contents and quality

The quality of the metadata associated with an observation is very important if the desire is to facilitate access.

As far as possible, the *data should stand by itself*. It should be as complete and accurate as possible with *everything explicitly described and with minimal requirements for externally defined information and no assumed knowledge*. This is particularly important for users that are not familiar with the data and therefore do not necessarily know that they need to look for additional information.

Since it is desirable to determine whether a data file is of interest before it is copied, bit is extremely beneficial if an instrument/observatory produces a high quality *catalogue of the observations*; it is desirable that this should include some sort of *quality flag* that provides an additional criteria on which to base selection.

Also, as part of the process of ensuring that information are recorded as fully as possible, the *observations should be supported by engineering logs and other similar types of information in an electronic format*.

There are issues about the types of file used by the different communities and what this implies to the ability describe the data. It should not be necessary for all communities to adopt the same type of file – some files are better suited to certain types of data than others – but it is possible to improve the uniformity of the file metadata. In particular, there is a need for the use of *standardized parameters and terms and an integrated set of coordinates* in all types of data and data products.

Note that these issues are discussed in more detail in CASSIS Deliverable D2.2.

Availability of data products & metadata

As the volumes of data have grown it has become increasingly important that that the user can determine as much as possible about what is in the data from data products and metadata. This should ensure that it is possible to make an informed selection of the data without needing to access the data themselves.

Products of this type – and this includes what are normally called quick-look data – are extremely important for researchers that are not familiar with particular data sets, presenting them with things that are easier to assimilate. The *generation of quick-look products should* *be encouraged* but it is desirable that they conform a set of standards that makes them easier to use.

Many data products have been designed to meet particular needs and are generated as part of the instrument data system. However, other products may be required to satisfy specific criteria that cannot be determined a priori and these products therefore need be created on the fly; in such cases, steps need to be taken to ensure they can be generated as rapidly and easily as possible.

Level of processing

The different communities handle their data differently and provide products processed to different levels. Some communities always provide their data as processed products while others provide "raw" data that needs to be processed.

If the user is required to process the data, then there is an implied level of understanding of what the data means. Since this is not always the case, processing pipelines can be useful in levelling the playing field and can be particularly useful if the user is not familiar with the data.

- Traditionally the solar community used to distribute un-calibrated data together with the calibration routines – the user then extracted and calibrated the subset of the observations that they needed. This approach is very flexible but there are issues related to *provenance* of the calibration data and routines that have not been properly addressed.
- For some recent solar instruments and observatories the paradigm has started to shift. Data volumes have grown to the extent that it is no longer practical to just distribute the requested dataset files and the data system needs to be able to extract and calibrate the subset that is required by the user. The first solar mission to have this requirement is NASA's Solar Dynamic Observatory (SDO but others[2] are in the pipeline. For these observatories the data will there be delivered to a higher level of processing that was previously the norm.
- In space physics the data are almost always reduce to products, sometimes combining observations from instruments together. Many of the data products are available at several processing levels.
- The planetary data are more complex because the planetary community is really a cluster of sub-communities, analogous to an extent to the geosciences. Europlanet RI has recently developed a Data Model to try to unify the different types of science. Again many data products are available at several processing levels.
- The geoscience community involves very large datasets and products very important. We are still learning about this community.

Differences in data policy

An issue related to the accessibility of the data is the data access policy of an instrument or observatory. There are considerable differences from instrument to instrument and between the communities with regard to this with some instruments adopting an open data policy and others being much more restrictive; in some cases it depends on who you know.

[2] The Advanced Technology Solar Telescope (ATST) that is currently being built in Hawaii and the proposed (ground-based) high-resolution European Solar Telescope (EST) will both add to the torrent of data and have similar requirements.

Since the mid 1990's data within the solar physics community has for the most part adopted a policy of open access. Planetary missions do not generally have open data policies although there are variations between mission and between the different agencies (NASA, ESA, etc.).

Some purists argue that VOs should only provide access to publically available data, however this could constrain the ability to do science. While *all groups should encouraged to make their data open*, it should recognized that there may be cases where this does not occur and the VO should still try make provisions in order to improve accessibility.

One of the most legitimate cases is where there are restrictions on access for an interval after the observations were made in order to give the instrument team the opportunity to analyse the data; this can often be circumvented if the person wanting to use the data works with the instrument team (which can be beneficial to both sides). Other cases where the observer just makes their observations and leaves benefits almost no one.

HELIO has developed techniques that make it possible to undertake searches including data that have not been released for one reason or another. The technique basically involves the provider harvesting information about the observations from an archive and making the observing catalogue available, either from the archive site or through a VO. Using this approach the user knows that potentially useful observations were made and who should be contacted in order to gain access.

Tools and Capabilities

The users from the different communities work in very different analysis environments and one of the challenges for interoperability is to establish an environment that all feel comfortable in. Being able to do this is important for users that are not familiar with demands of the different domains and with the data and any processing that is required.

One of the keys to this is *making any capabilities available as services*. With a modular approach it is much easier to tailor the environment to the needs of the user, combining the services to provide the desired capability; the ability to call services from scripting languages and workflows is also desirable. An advantage of the service driven approach is that the internal working of the service can be completely hidden from the user and all they need to know about is the functionality of the capability.

In some cases, useful tools already exist within projects or communities – these might include searching, processing, plotting, etc. *Finding ways on integrating these with a more general collaborative environment* would reap enormous benefits but this requires some level of agreement between the communities on common standards for access to the services (and data resources). Using *interfaces based on recommendations from standards bodies* such as the IVOA and IPDA is a step in the right direction in this respect.

There is a particular need of a *capability to do processing on demand*. There is an increasing requirement for this in order to generate quick-look products for (combinations of) specific data sets and time intervals or generate event lists according to some new criteria. These products are required to help the user understand what they are looking at and to allow them to make informed choices of what observations they need in order to address their science objective. Doing the processing on demand is necessary because it is not possible to produce all that might be required *a priori*.

More *general data processing requirements can also be implemented as services*. Most data need some form of extraction and processing to provide products that are easier to use; in an environment where the user is trying to combine information from many different domains then the need for this is greater that ever. Pipeline processing of selected datasets, including extraction and calibration need to be developed; doing so in a way that allows the capability to be reused could add to reap benefits down the line.

One of the biggest challenges is how to provide integrated access to the different types of data from the domains. While in the future we hope that it will be possible to increase the interoperability of the datasets by adopting a standardized set of terms and increasing the amount of common metadata, this will always have limits of what can be achieved and there are many issues that need to be addressed now. Semantic mapping tools and ontologies are needed that are based on data models that span the domains. Some progress has been made on this in HELIO but much more is required.

Note that these issues are discussed in more detail in Deliverable D2.1.

Requirements from the Use Cases

From the Use Cases that have been collected we have extracted several requirements that could help formulate the architectural design of a collaborative environment, the user interface and the search engine as well as indications of new types of metadata that may be needed.

The requirements are those that respond to the needs of the "end-users" or consumers of the system. These are the users that will need to request information and data and use software or other resources from the broad set of resources that for a shared research environment.

At the opposite end of the supply chain the resource providers may be equally diverse, but with a very different list of requirements. Therefore we have opted to make a separate consultation with the community of data or resource providers (which group may overlap extensively with the end-users) to be prepared and presented in a separate document. The discussion with the data providers will need to be constrained by what is reasonable and practical and hence cannot be included as part of this document.

The Use Cases on the possible effect on the Earth's weather and climate are some of the most challenging highlight issues related to the availability of suitable data. Reducing the information in the solar and weather and climate domains presents problems. Also, before (all) data became generally more accessible in the mid 1990's it could be difficult to find data that can be used to undertake some studies.

Trying to ensure that future datasets are more interoperable may be the most important step in ensuring that this type of comparison of data from different domains is even possible.

Generalized Interaction Sequence

The analysis of the use cases indicates that in general the search and assessment of data has to be performed in a multi-step format that allows the user to interrupt the process at any level and to loop back to previous stages if required; the user should also be allowed to perform exploratory data analysis externally and incorporate the results if desired.

The query of observational parameters could be organized in two steps:

1) The general data selection, with the purpose of finding those data that fulfil very general parameters, and
2) A refined data selection, which allows the user to set a larger number of constraints for a more detailed characterization of the observations to be selected.

At any stage, if the query does not produce matching results, sufficient information should be provided so that the user can understand which requested parameters were not satisfied within the available observations; ideally feedback should be provided on how much the parameter ranges need to be extended in order to find some positive results.

This approach allows the user to get a feel for the data available and should minimize the number of null returns from searches that have too stringent criteria at the outset. For instance, a user may ideally require very high spectral resolution, but in the absence of this the user may wish to be informed of any moderate spectral resolution data that are available.

For the experienced user, or one that is aware of what observations are available, provision should be made to start a search directly from the refined data selection stage.

A possible step sequence is the following:

1. Initial search parameters determined from either/and:
   a. Values or a range of values of the fundamental physical parameters
      i. Date/time or time interval
      ii. Location, pointing coordinates (instrument, satellite), field-of-view
      iii. Observation type
   b. Values or a range of values determined from catalogues of features, objects, or transient events
2. If appropriate, a definition of how different sets of observations might be related; this may require some type of modelling to determine the relationship for a given search
3. Initial assessment of the response to the multi-instrument search based on 1 and 2 to determine general coverage of the observations
4. Selection of instruments based on their characteristics, location and operational status
5. Initial selection of datasets based on match to chosen scientific objective
6. Final selection of datasets; data quality and need for processing could be factors
7. Extraction, calibration and retrieval of required observations

Summary of Requirements

A summary of the requirements is given below:

No.	Description	Code
GN01	All types of data and data files should be supported; read routines should be created in conjunction with those with expertise in the data.	
GN02	As far as possible the data should remain connected to the site where the expertise about the instrument resides	
GN03	All groups should encouraged to make their data open access	
GN04	Catalogues describing data that have restricted access (but could be of interest) should be provided so that the observations can be included in the search	
GN05	Encourage providers to follow simple rules for naming data *[It should always be easy to identify a data file; file name should be sufficiently unique that the file is able to exist without conflict outside of the environment that it is normally stored in]*	
GN06	Encourage providers to follow simple rules for storing data. *[Archive should be well simply structured following a hierarchical tree based on date]*	
GN07	Encourage use of standards for data to ensure maximum interoperability *[Standardized terms and an integrated set of coordinates]*	
GN08	Data should stand by themselves; everything should be explicitly described and there should be a minimal requirement for external information and no assumed knowledge	
GN09	Instruments and observatories should be encouraged to produce high quality catalogue of the observations; this should include some sort of quality flag	
GN10	Observations should be supported by engineering logs and other similar types of information in an electronic format	
GN11	Data that are provided should be processed to as high a level as possible	
GN12	The capability to extract and calibrate a subset of each dataset should be provided; such pipelines should be well documented.	
GN13	The capability to generate quick-look data on the fly should be provided. The generation of quick-look products should be encouraged,	
GN14	High quality, intelligent access should be provided to support cross-disciplinary studies where the user may not know a lot about the data and needs as much help as possible	
GN15	Make capabilities available as services that use standard interfaces *[Standard interfaces based on recommendations from standards bodies]*	
GN16	Guidelines need to be created to help integrate resources into a more general collaborative environment	
GN17	Tools need to be provided to connect the metadata associated with the different domains; these should be based on semantic mapping techniques and ontologies.	

Summary and Discussion

This document reports on a study that is still developing.

Our conclusions are preliminary but provide a focus for the process of consulting the community in order to fill out the use cases and flesh put the requirements

One of the main objectives is to determine how to improve interoperability by identifying techniques that will help bridge the different ways of doing things. This needs to be done in a way that all sides feel comfortable in using data from other domains.

Appendices

Details of the Science Use Cases

No.	Title	Author
1	Comparison of Coronal and In-situ Solar Wind Diagnostics	G. Poletto
2	Link remote-sensed and in-situ observations in the heliosphere	D. Berghmans & K. Steed
3	Find occurrences when sites are connected by heliospheric propagation	D. Berghmans & K. Steed
4	Relate occurrence of planetary aurorae with solar activity	R.D. Bentley & G Branduardi-Raymont
5	Using observations from L5 to predict terrestrial Space Weather	R.D. Bentley & A. Opitz
6	Relate terrestrial ionospheric conditions with other observations	D. Kataria & R.D. Bentley
7	Variation of the Cosmic Ray flux	R.D. Bentley
8	Influence of space weather effects on Aeronomy studies	D. Berghmans
9	Effect of solar activity on upper atmospheric layers and weather	[R.D. Bentley]
10	Effect of long-term solar variability on the Earth's climate	[R.D. Bentley]
11	Effect of solar activity on terrestrial climate variability	[R.D. Bentley]

Same table as earlier – make sure the contents are cross-copied

1) Comparison of Coronal and In-situ Solar Wind Diagnostics

G. Poletto (INAF)

Objective: Comparison of solar wind diagnostics using coronal and *in-situ* plasma parameters. This implies the identification of the source region of high/slow wind streams.
Data Sets: Magnetograms. Disk images in XUV. Off-limb images in the visible. Spectra at coronal levels in the UV range. Integrated disk intensities in H Lyman-alpha, Lyman-beta, OVI doublet lines. In situ data: density, temperature, elemental composition, magnetic field.
Data Selection Steps: Identify, on the basis of the Ulysses orbit, SOHO-Sun-Ulysses quadrature dates. Retrieve coronal data along the radial to Ulysses. Identify time-delayed *in-situ* data that correspond to coronal data.
Data Processing during the selection phase: Preview coronal and *in-situ* data to discard data of low quality; Identify "special" time intervals when isolated events might have occurred (e.g., CMEs); co-align disk and coronal images.
Data Processing for Selected Data: Retrieve calibrated data.

Possible variant:
Compare the solar wind time variation at different points in the Solar System.
Long-lived features on the solar surface causing regions of slow and fast solar wind can be observed in locations that are separated in longitude – i.e. that are not connected by the Parker Spiral. For example, co-rotating interaction regions (CIRs) can be observed for days or longer??

2) Link remote-sensed and in-situ observations in the heliosphere

D. Berghmans (ROB) and K. Steed (KU Leuven)

Objective: Link remote-sensed observations of the Sun with in-situ data are other points in the Solar System. Context Example: Link Solar Orbiter remote-sensed and in-situ data
Data Sets: Ephemeris information for observing spacecraft (in relation to the Sun) Observations of conditions on and near the solar surface (magnetic field, euv images) Details of solar events and/or events at other locations Prep of event catalogues (dimming areas)
***Data Selection Steps**:* 1) Use propagation model to determine travel times and direction of motion Alternative method(s): assuming constant radial outflow, assuming Parker-spiral propagation, etc Or 2) Use in-situ time series data to determine arrival times
Data Processing during the selection phase: Models need to be run on each event. Time series data of in situ data may needs to be displayed in standard coordinates (if events not detected *a prori*) to allow identification of magnetic cloud boundaries Backward projection based on in-situ velocity to give time window for CME lift-off
What output is required and how is it used? What format should the output be in? In situ data interpreted (transformed) in local coordinate system (standard variance) to allow more detailed interpretation of user location wrt magnetic structure (Calibrated?) EUV images to determine extent of dimming Objective it to determine evolution of the structure
What is making things more difficult than it needs to be? Is there any assumed knowledge? Assumptions of type of structure (force-free flux rope) [matching returned remote and in-situ data from Solar Orbiter; pointing, fov for remote; relative location of spacecraft -> tool?] Interpretation of remote sensed data for different radial distances (plate scale)

Issues:

3) Find occurrences when sites are connected by heliospheric propagation

D. Berghmans (ROB) and K. Steed (KU Leuven)

Objective: Finding occurrences when spacecraft X (Solar Orbiter, Solar Probe, ACE, etc.) that measures the solar wind is exactly (to within given accuracy) between the Sun and another spacecraft Y (e.g. Mars Express).   Context Example: Prediction of space weather (e.g. CME arrival) at other planets
Data Sets: Ephemeris information for spacecraft X and Y (in relation to the Sun); including operational availability of observatories and instruments
***Data Selection Steps**:*   Need to use some sort of propagation model to determine travel times and direction of motion; ability to select models   Predict times when observatories suitable aligned (tool; campaign -> data selection rationale) for in-situ; relative location of missions for supporting remote-sensed observations
Data Processing during the selection phase: Examine signatures in the in-situ data   Examine supporting data
What output is required and how is it used? What format should the output be in?   Tools that give orbital information for heliosphere and close to planets   Similar to selection tool for alignment, etc.   Ability to compare model output
What is making things more difficult than it needs to be? Is there any assumed knowledge?   Access to data difficult (planetary data, WIND)   Lack of "engineering data"; quality of data checking; data gaps

Issues:

4) Relate occurrence of planetary aurorae with solar activity

R.D. Bentley (UCL) and G. Branduardi-Raymont (UCL)

Objective: Examine the relationship between auroral activity on the gas giants and other non-terrestrial planets with solar activity and variations in the solar wind. Where the propagation paths link the planets to observatories at other points in the solar system, also examine the signature at those points. Predict interesting times for future observations from propagations. Need to work forwards and backwards.
Data Sets: List of remote-sensed observations of planetary aurorae from observatories such as HST (optical, UV and EUV), XMM and Chandra (X-rays) etc. Radio observations of planetary environments from ground- and space-based observatories Remote-sensed observations of the solar surface In-situ solar wind observations in the heliosphere and planetary environments
Data Selection Steps: Working backwards: Look for occurrences of aurorae in catalogue or signature of onset in radio observations and use propagation model to identify related solar activity Working forwards: Use propagation model to predict when solar activity will cause effects on planetary environment
Data Processing during the selection phase: Use of propagation tool
What output is required and how is it used? What format should the output be in? Solar wind parameters (speed, density, mag. File, ?) vs. time at various distances from the Sun (e.g. at Jupiter and Saturn). To go further than Saturn is not straight forward with codes like mSWiM.
What is making things more difficult than it needs to be? Is there any assumed knowledge? Lack of a catalogue that combines observations of planetary aurorae[3] that have been identified within the general observations made by suitable instrument/observatories.

Notes:

1. The magnetospheres of the gas giants are very large and encompass planetary moons. There are also additional processes going on that complicate the issue.
2. Images may show the occurrence of aurorae but unless they are at high cadence the timing will only be tentative; radio observations likely to be more precise.

---

[3] The planned Auroral Planetary Imaging and Spectroscopy (APIS) system may provide the necessary information – TBC

5) Using observations from L5 to predict terrestrial Space Weather

R.D. Bentley (UCL) and A. Opitz (UPST)

Objective: Examine the viability of using observations from near L5 to predict space weather conditions at Earth.
Data Sets: Remote sensed observation from EUVI, COR and HI In-situ observations from IMPACT/PLASTIC In-situ observations from the near-Earth environment Geophysical observations on/near Earth Catalogues of Earth-directed CMEs seen from STEREO and SOHO, etc.
***Data Selection Steps**:* Determine time interval when STEREO-B spacecraft was within 10 degrees of L5 using the HELIO Instrument Location Service Identify occurrence of Earth-directed CMEs Compare time-shifted in-situ observations at L5 with those on the Sun-Earth line
Data Processing during the selection phase:
What output is required and how is it used? What format should the output be in? Comparison plots of the in-situ observations at both locations; the tool would need to apply the appropriate time shift to the data according to the angular separation of the spacecraft from the Earth and the difference in radial distance.
What is making things more difficult than it needs to be? Is there any assumed knowledge?

Notes:
STEREO-B has been dropping back behind the Earth at 22 degrees per year and was within 10 degrees of L5 (Lagrange point located 60 degrees behind the Earth) for a good part of a year. The spacecraft made observations similar to those that might be made by a mission orbiting L5; at this location the structure of the quiescent solar wind can be observed over 4 days before it is observed near the Earth.

6) Relate terrestrial ionospheric conditions with other observations

D. Kataria (UCL) and R.D. Bentley (UCL)

Objective:
Determine the causes of variations in conditions in the Earth's ionosphere as observed by *Cubesats* in the QB50 constellation.
Data Sets:
In-situ observations from the QB50 spacecraft Remote-sensed observations of the Sun In-situ observations at L1 and from the near-Earth observatories Ground-based data based on observations of GPS
***Data Selection Steps**:*
Approximate location can be determined from a service like the HELIO ILS.
Data Processing during the selection phase:
Combine data from multiple *Cubesat* spacecraft into a single dataset. Ability to compare with other datasets…
What output is required and how is it used? What format should the output be in?
Data from QB50 over-plotted on geographic maps from ground-based observations
What is making things more difficult than it needs to be? Is there any assumed knowledge?
How data will be processed is not defined for the constellation yet; depending on what is decided, we may need a service that is able to combine data from constellations of spacecraft into a single dataset.

Notes:
1. The QB50 constellation of *Cubesat* spacecraft is partly funded by an FP7-SPACE grant – the spacecraft are currently scheduled for launch in first half of 2015. The spacecraft will form a line many thousands of kilometres long after deployment.
2. The comparison can probably not be done in near-real time because of the difficulty of recovering both space- and ground-based observations in a timely manner. The maps of ionospheric conditions that are published in real time are actually an interpolation in time and space; the network of ground stations is space at best and non-existent in some areas and the times are which the data become available to generate the maps are very variable.

7) Variation of the Cosmic Ray flux

R.D. Bentley (UCL)

Objective:
The Earth is bathed in a cosmic ray flux that has two components: a galactic cosmic ray (GCR) background that is modulated by the solar cycle and *short-lived* variations caused by solar activity (SCR). According to the Euratom Directive (96/29), the radiation exposure of flight crew working for European airlines has to be monitored and action taken if the accumulated dose exceeds 6 mSv per annum. The dose is normally determined by calculation based on monthly values of the Heliospheric Potential that provides an estimate of the slowly varying comic ray background, with some account for solar activity. The objective of this Use Case is to determine whether the Heliospheric Potential adequately describes the variations in the cosmic ray flux; increases are caused by particles emitted by energetic solar flare and reductions result for the passage of CMEs (Forbush Decrease).
Data Sets:
Event lists of the occurrence of solar energetic particle (SEP) events and coronal mass ejections (CME) and of ground level events (GLE) and Forbush decreases. Energetic particle monitors, e.g. from GOES or ACE. Ground-based neutron monitors (cf. the Neutron Monitor Data Base, NMDB) Coronagraph data (to provide estimates of the mass of the ejections)
***Data Selection Steps**:*
Determine the occurrence of SEP events; look for the association with CMEs and variations in the cosmic ray level as indicted by the neutron monitors. Dose rates monitored on-board aircraft is available for the times of events. Note: variations monitored on or near the ground could be affected by time of day.
Data Processing during the selection phase:
Ability to plot event data on a variety of time series data
What output is required and how is it used? What format should the output be in?
What is making things more difficult than it needs to be? Is there any assumed knowledge?
Easy access to all relevant data so that plots can be generated. Lack of systematic on-board monitoring of eradiation xposure levels.

Issues:

8) Influence of space weather effects on Aeronomy studies

D. Berghmans (ROB)

Objective:

Aeronomy studies of the higher atmosphere of the Earth can be conducted by solar occultation campaigns by tracking the solar signal through the Earth atmosphere as the Sun "sets" for an observing instrument on-board a satellite. The objective here is to determine whether the observed variability in such a data set corresponds to atmospheric structures, solar variability, or noise from the particle environment of the spacecraft (South Atlantic anomaly, auroral activity).

Data Sets:

Solar radiometer data of the instrument going in occultation
Comparable Solar radiometer of a non-occulted instrument
Spacecraft orbital data
Particle environment data

Data Selection Steps:

Time ranges are identified from the spacecraft orbital data for which the spacecraft gets occulted
For these time ranges, cases are excluded for which

a) the spacecraft is in the South Atlantic anomaly or in the auroral zones
b) strong solar variability occurred

The remaining cases constitute input data for aeronomy studies to determine the transmission of the solar flux through the Earth atmosphere.

Data Processing during the selection phase:

The spacecraft position with respect to the Sun-Earth line has to be evaluated.
The spacecraft position with respect to the geomagnetic field has to be evaluated.
The selected time ranges have to be evaluated for solar variability.

What output is required and how is it used? What format should the output be in?

The output are cases (time-ranges) in which (partially) occulted data is available that is undisturbed by known environmental effects.

What is making things more difficult than it needs to be? Is there any assumed knowledge?

?(maybe this is an too-easy case).

Issues:

9) Effect of solar activity on upper atmospheric layers and weather

[R.D. Bentley (UCL)]

Objective:
Examine the effects of structure in the solar Wind and events on the sun on the upper and middle atmospheric layers of the Earth to determine if they are related to terrestrial weather events.
Data Sets:
Remote-sensed observations of the Sun; related in-situ observations close to Earth Generalized list of solar events that could affect the Earth Generalized indices of conditions in the upper atmospheric conditions Reports of anomalous weather conditions *(the latter two need to be selectable on a regional basis)*
Data Selection Steps*:*
Need to correlate weather and solar events. Looking for the consequences of solar activity may be the starting point. Comparison over many years is needed.
Data Processing during the selection phase:
The data needs to be reduced in a way that facilitates comparison.
What output is required and how is it used? What format should the output be in?
Data reduced to generalized information against time that can be compared.
What is making things more difficult than it needs to be? Is there any assumed knowledge?
Deals with very different communities so a lot of issues.

Notes:
Solar activity represents only a very small variation of the total solar output; even large solar eruptions only release an amount of energy comparable to what the whole Sun emits in a few seconds. Much larger variations in the sunlight received at the Earth's surface are due to the Earth's tilt and its elliptical orbit and many scientists therefore have difficulty accepting that solar activity has a major affect on our weather.

However, there is some complex chemistry going on in the upper layers of the atmosphere that is affected by different types of solar emission and it is thought that this could be important effect. Since there are many other effects in play, it will be difficult to isolate those caused by solar activity; the key may be to identify whether specific regions might be affected and also to *study a larger number of events*. The size of any effects, where and when they occur and whether they have an influence on lower layers of the atmosphere needs investigation.

10) Effect of long-term solar variability on the Earth's climate

[R.D. Bentley (UCL)]

Objective: Examine the effects of long term variations on solar activity (reflected by the solar cycle) and solar output on the Earth's atmosphere, oceans and polar ice cover.
Data Sets: Datasets describing the variations in solar output on the timescales of tens of thousands of years. Datasets describing the climatic conditions on timescales of tens of thousands of years; these need to cover as much of the globe as possible and show general trends.
***Data Selection Steps**:* No real selection; mainly comparison in trends.
Data Processing during the selection phase: Data need to be processed so that comparison of trends is possible.
What output is required and how is it used? What format should the output be in?
What is making things more difficult than it needs to be? Is there any assumed knowledge?

Notes:
This Use Case does not provide a good avenue to pursue the idea of interoperability too far. In principle it would be beneficial for such a study to have "standardized" file formats that could be used but there are very few datasets that represent solar variability and climatic conditions that far in the past and there are few options to extend them; also, in all cases the data are really proxies rather than actual measurements.

As a consequence we are not pursuing this Use Case any further. A new Use Case on *climate variability* has been introduced instead.

11) Effect of solar activity on terrestrial climate variability

[R.D. Bentley (UCL)]

Objective:
Examine the effects variations on solar activity on short-term climate variation that are represented by periodic or intermittent changes related to general weather patterns
Data Sets:
Remote-sensed observations of the Sun; related in-situ observations close to Earth General weather indices – temperature maps, rainfall maps and patterns, etc.
***Data Selection Steps**:*
Need to correlate "observed" apparent variations in climate and solar events. Comparison over many years is needed.
Data Processing during the selection phase:
The data needs to be reduced in a way that facilitates comparison.
What output is required and how is it used? What format should the output be in?
Data reduced to generalized information against time that can be compared.
What is making things more difficult than it needs to be? Is there any assumed knowledge?
Deals with very different communities so a lot of issues

Notes:
1) Solar observations that represent the "Sun as a Star" are required; the detail afforded by the various imagers is greater that is needed but could be used with reduced resolution.

2) Some changes could to be regional. Lockwood (2010) suggests that the recent cold winters in Europe are a consequence of the low level of solar activity and the delayed solar maximum; similar claims are made for the extremely cold weather in Europe in the 17th Century that corresponded to the time of the Maunder Minimum in solar activity. Lockwood claims this is related to how chemistry and circulations patterns combine to affect specific regions.

This Use Case should be easier to study than effects on weather or long term climate changes.