A Collaborative Research Environment for Heliophysics

Robert Bentley (UCL-MSSL)
David Berghmans (ROB), Andre Csillaghy

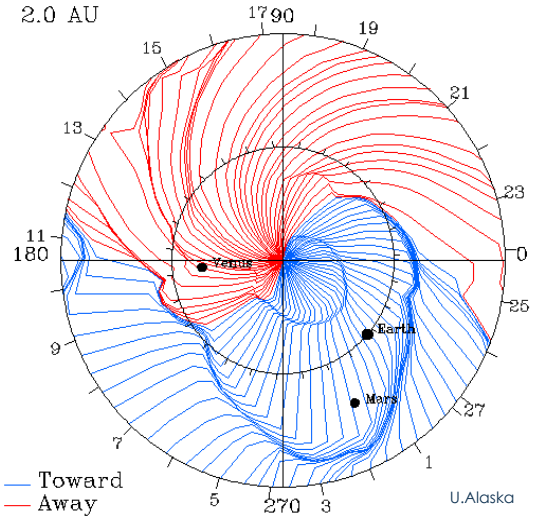
(EGU, Vienna – 26 April 2012)

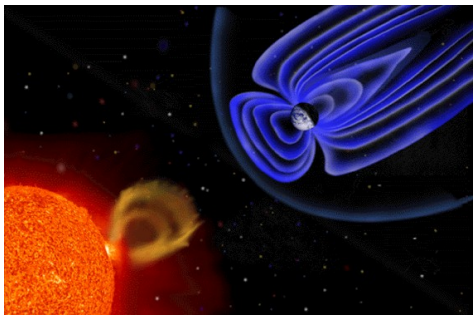
www.cassis-vo.eu

Background

- The Heliophysics Integrated Observatory, HELIO, is a Research Infrastructure funded under EC's FP7 Capacities Programme
- HELIO is creating a *collaborative environment* where scientists can discover, understand, and model the connection between solar phenomena, interplanetary disturbances, and their effects on planets, especially the Earth
  - Need driven by desire to study problems that span disciplines
  - Using metadata increasingly important as data volumes increase
- Many of the issues we are addressing are not unique
- Under CASSIS we are considering how the type of research environment that we are creating in heliophysics could be extended to other domains and communities
  - Also including overlapping collaborative environments

What is heliophysics?





- **Heliophysics is the effect of the Sun on the Solar System**
  - An event-driven science – something is observed and desire is to trace origins or subsequent effects
- **Nature of effect depends on causal phenomenon, type of emission, and the location of the observer**
  - Most effects have origins in emissions from solar activity; some related to propagating phenomena
  - Location of observer in relation to the source and with respect to a planet determines what is observed
  - Immediate and delayed effects result from the different types of emission
- **Presence of magnetic field and/or atmosphere influences effect on planetary environment**
- **Study requires an understanding of how phenomena evolve in space and time – how they propagate, interact…**

HELIO addressing a Generic Problem

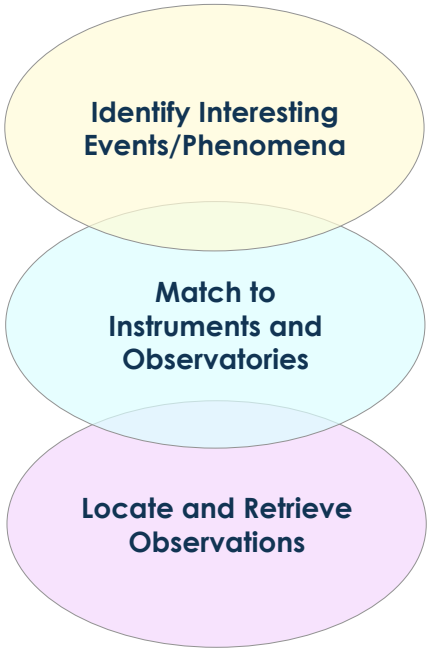
- **Identify interesting things to study**
  - Science Objectives define what user looking for
  - Search undertaken in 4-Dimensions across several domains
    - Effects occur as phenomena propagate and interact
    - Modelling required to understand whether, where and when to look
  - *Search based solely on metadata and derived products*
    - Event lists and feature lists from many domains used as a primary selection criteria
- **Review availability of suitable observations**
  - Determine whether suitable instruments are at the relevant locations
  - Determine whether the instruments were making observations
  - Decide if they could be showing something of interest
- **Locate, select and retrieve the required observations**
  - For all domains, system knows which types of data are held where and handles access no matter how data are stored (access protocols & formats)
- ***Analysis done with users own software tools (e.g. IDL/SSW)***

Service-Oriented Architecture

HELIO is an integrated system implemented with a service-oriented architecture

Design splits the tasks into a set of components or services

- Services to aid search process and turn science objectives into required instruments
- Services to locate and retrieve data when selection process complete
- The services can be used independently or as part of a workflow



Defined Services

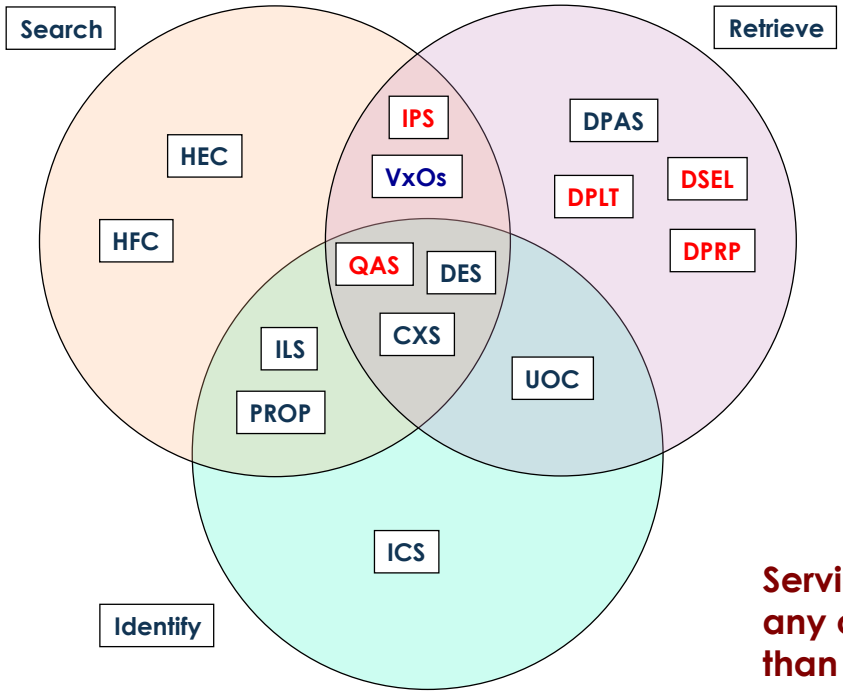
HELIO's capabilities are defines as a set of principal services that address these tasks

There are also a number of Enabling Services that provide capabilities such as processing, storage, coordinate transformation, security, etc.

The HELIO Web pages will provide more information as services are developed:
http://www.helio-vo.eu/

Service Name	Purpose
Search Metadata	
Heliophysics Event Catalogue (HEC)	Maintain and provide access to *existing* event data from all domains
Heliophysics Feature Catalogue (HFC)	Maintain and provide access to existing feature data from all domains
MetaData Evaluation Service (MDES)	Allows the user to create an auxiliary event list based on a *newly-derived* parameter, etc.
Context Service (CXS)	Provide context information to help the user make a selection
Review suitable observations	
Instrument Capabilities Service (ICS)	Match required observation type to one or more instruments (each part of an observatory)
Instrument Location Service (ILS)	Determine the location of an instrument (part of an observatory) at a specified time
Observation Coverage Service (OCS)	Provide information on whether an instrument was making suitable observations at a specified time
Locate and Retrieve the Data	
Data Provider Access Service (DPAS)	Provide integrated access to data archives in all domains no matter how the data are stored or accessed
Enabling Services	
HELIO Registry Service (HRS)	Maintain and provide access to a registry that describes all the services available to HELIO
Community Interaction Service (CIS)	Manages interactions with the community, including authentication and usage statistics
Processing Service	Support processing on demand
Storage Service	Provide storage for user information
Ancillary Information Service (AIS)	Provide integrated access to external resources that do not conform to HELIO interface standards
Coordinate Transformation Service (CTS)	Translated between the different coordinate systems used by the communities
Semantic Mapping Service (SMS)	Maps terms used in the metadata from the different communities
HELIO Monitoring Service (HMS)	Keeps track of the status and performance of the services that the HRS knows about
Resource Usage Service (RUS)	Keeps track of usage of HELIO so that the project can provide statistics to users, providers, etc.

Services as Building Blocks



CTS	Coordinate Transformation
CXS	Context Service
DES	Data Evaluation Service
DPAS	Data Provider Access
DPLT	Data Plotter
DPRP	Pipeline Processing
DSEL	Data Selector
HEC	Heliophysics Event Catalog
HFC	Heliophysics Feature Catalog
HPS	Processing Service
HRS	Registry Service
HSS	Storage Service
IAS	Index Access Service
ICS	Instrument Capabilities
ILS	Instrument Location
PROP	Propagation Modelling
QAS	Quick-look Access
SMS	Semantic Mapping
UOC	Unified Observing Catalog
VxO	HSD Virtual Observatory

Services can be used in any order and in more than one way

Services as Building Blocks

- **The capabilities that HELIO is developing should be thought of a building block in a larger capability – parts of a tool kit…**
- **Service-oriented architecture has advantages**
  - Services can be used individually or as part of a workflow
  - Method of implementation is hidden from the user
  - New capability can be implemented as a new service
  - Services can be developed and maintained independently
- **Services interfaces need to be compliant with a set of standards in order to ensure interoperability**
- **HELIO has defined Web Service interfaces for its services based on IVOA specifications (with extensions)**
  - If adopted by other the VxOs and other capabilities, these could become part of the tool kit
  - *May need some iteration to satisfy needs of all*

Many things can be services

- The Virtual Observatories in the US have been developed as an alphabet soup supporting different parts of the heliophysics...
  - VxOs etc. under NASA-GSFC Heliophysics Science Division:
    - Virtual Solar Observatory (VSO), Virtual Heliospheric Observatory (VHO), Virtual Space Physics Observatory (VSPO), Virtual Magnetospheric Observatory (VMO), Virtual Ionosphere Thermosphere Mesosphere Observatory (VITMO), Virtual Radiation Belt Observatory (ViRBO), Virtual Wave Observatory (VWO), …
  - VxO project funded by NSF:
    - Virtual Solar Terrestrial Observatory (VSTO)
- Other relevant (resource rich) US data providers
  - National Space Science Data Center (NSSDC)
  - NASA's Space Physics Data Facility (SPDF)
  - NOAA's National Geophysical Data Center (NGDC)
- Innumerable capabilities scattered across Europe and Asia
- Some can be integrated as services
- The same mix of capabilities exist in other, related communities

Building outwards and sideways

- In trying to integrate other capabilities into HELIO we are rethinking how the environment should be established
- Heliophysics is the effect of the Sun on the Solar System but the boundary is fuzzy
  - HELIO provides access to magnetospheric data but it is harder to relate effects to ionospheric data
  - We know that changes in solar activity and output can affect the Earth's atmosphere
    - Possible effects on weather and climate
  - But, as we move into Earth's environment, the emphasis of the search shifts
    - How much of observed effects are caused by the Sun in comparison to other drivers
    - No longer how much does the Sun effect the atmosphere
- Consider how to create a more general collaborative research environment from this perspective

I – Align the Services

- Many existing *Services* already have have query interfaces, a number as Web Services
- Interfaces to the HELIO services are mostly of two types, based on standards developed by the IVOA
  - HQI based on TAP and PQL; UWS used for services that process
  - Additions to the IVOA standards have been required
    - Added parameters to support a 3D spatial query
    - Added support full SQL query – PQL had deficiencies
- Query interfaces ought to support queries that are agnostic to science of communities
  - IVOA standards were born out of needs of the astrophysics community and still contain some remnants of that bias
  - HELIO query interface probably more closely aligned to the needs of the Earth sciences than the raw IVOA standards
- Behind the interface, services can be built using any technology with whatever structure is desired

II – Align the Metadata

- Several of the HELIO services are dedicated to managing and providing access to metadata
  - Use events list as an example
- HELIO tries to maintain structure of lists while standardizing terms
  - A significant amount of conditioning has been necessary because the information is not provided in a standard way
    - Groups have developed their own ways of describing things including time, position and pointing
  - On the fly translation and semantic mapping are possible
  - More efficient to translate once and use many times
- Problem is that there are no standards to follow
- Input from other communities needed to develop standards and extend what has been learnt by HELIO
  - Standards developed by IVOA for event lists not optimized for science within the Solar System
  - Need to discuss a standard that encompasses Earth Sciences

III – Standardize data file *Common Area*

- Metadata of usually derived from the data and need to make changes indicates issues related to the underlying data
- Currently users from one community can have difficulty in understanding contents of data from another community
  - This is hindering cross-disciplinary science
- The metadata describing observational data usually have to describe a number of common items
  - Maximize the common area by standardizing the parameter names used and their contents
  - Use an agreed, related sets of coordinates to describe time and position and pointing; standard names/contents for other terms
- There will always be parts of files that relate to specific instruments or domains but maximizing the common area will greatly simplify interoperability

New File Formats?

- For the Collaborative Environment we are NOT requiring all providers to switch to new file formats, but:
  - Existing formats are now decades old and were not created with interoperability in mind
  - Frequently difficult to just open up a file and know what it means and what it contains
  - Difficult to require all formats to properly annotate (unambiguously describe) the parameters
- Tidal wave of data is heading in our direction and we must maximize interoperability before we are swamped
- The existing formats, FITS, CDF, netCDF, Text, etc. have their merits and are suited to certain types of data
- Should we consider a move to a family of new formats that are designed to facilitate interoperability
- VOs would manage to use of existing data sets...

Conclusions

- **Difficulties in providing integrated access across disciplinary boundaries is hindering science**
- **A collaborative research environment for heliophysics could be established relatively easily by aligning the services and metadata**
  - **HELIO and CASSIS have identified much of what needs to be done**
  - **Significant benefits if also try to maximize interoperability of the data**
- **Collaborative environments covering other sets of disciplines are also possible – these could be intersecting**
- **Beneficial to establish standards to facilitate interoperability**
  - **Need to involve a wide range of communities in formulating these**
- **Exploring relationships across disciplinary boundaries may lead to new types of science**