The role of e-Infrastructures supporting Solar system science within general research environments

R.D. Bentley

University College London

Coordinator of HELIO and CASSIS FP7 projects

(EPSC 2012, Madrid, 26 September 2012)

Introduction

- The **Heliophysics Integrated Observatory**, HELIO, has created a research 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, considering how the type of *e-Infrastructure* created in heliophysics could be extended to other domains and communities
  - Also including overlapping collaborative environments

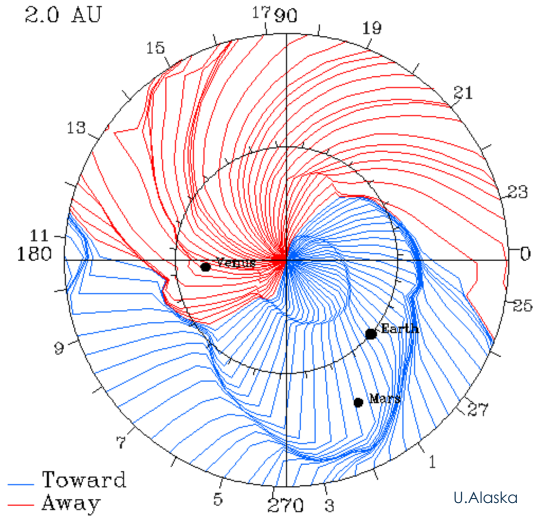
Defining a few terms

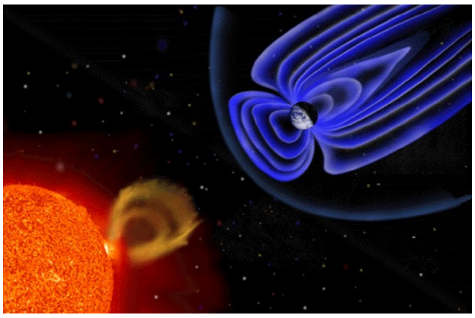
- Data Archives
  - There are many archives of data scattered around the world
  - Varying capabilities – resource poor to intelligent
- Research Tools/Facilities (?)
  - Provides integrated access to some capability with a specific objective – e.g. AMDA
- Virtual Observatories
  - What this means varies
    - NASA VxOs quite limited in scope – dedicated to sub-community
    - IVOA more general in scope – more than a VO?
- e-Infrastructures
  - Intended to be more general in scope, not just a VO
- Great scope for creating a larger collaborative environment in which all these can be used together

Archives and other capabilities

- Virtual Observatories provide *intelligent* access and expertise
  - In the US the VOs supporting different parts of the heliophysics have been developed as an alphabet soup
  - VxOs 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) data providers
  - National Space Science Data Center (NSSDC), NASA's Space Physics Data Facility (SPDF), NOAA's National Geophysical Data Center (NGDC), NASA's Planetary Data System (PDS), ESA's Planetary Science Archive (PSA)…
- Innumerable capabilities also scattered across Europe and Asia
- Similar mix of capabilities exist in other, related communities

Starting from HELIO…





- HELIO is a Research Infrastructure for the heliophysics community
  - Heliophysics is the effect of the Sun on the Solar System
- 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 addresses a Generic Problem

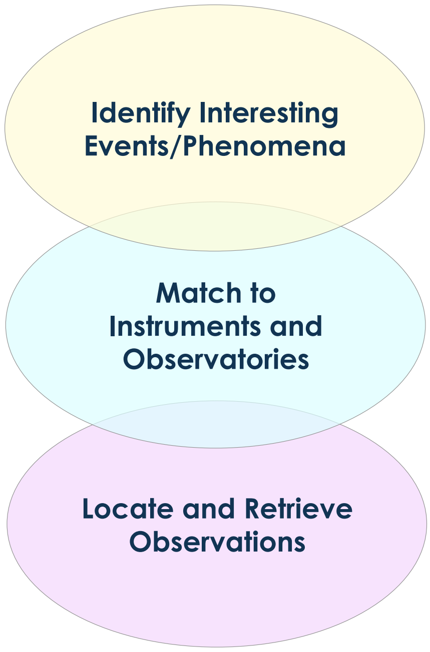
- Identify interesting things to study
  - Science Use Cases define these
  - Search undertaken in 4-Dimensions across several domains
    - Effects occur as phenomena propagate and interact
    - Modelling required to understand whether, where and when to look
  - *Search based solely on metadata and derived products*
- Review availability of suitable observations
  - Were suitable instruments at the relevant locations?
  - Were they observing?
  - Could they 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 how to access them (access protocols & formats)
- *Analysis done with users own software tools (e.g. IDL/SSW)*

Using a 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



Search Process

- *Objective:*
  Aspects of phenomenon ➔ Set of Obs. Types @ Times, Locations
- **HELIO provides as diverse a set of tools as possible, based on metadata from all relevant domains**
  - Heliophysics Event Catalogue (HEC)
    - Catalogue of existing event data
  - Data Evaluation Service (DES)
    - Allow user to evaluate time series data and generate own event list
  - Heliophysics Feature Catalogue (HFC)
  - *Propagation, etc. models*
    - Helps determine *whether, where and when to look*
  - Context Service (CXS)
- ***How the tools will be use – which and in what sequence – depends on the science problem and the scientist***

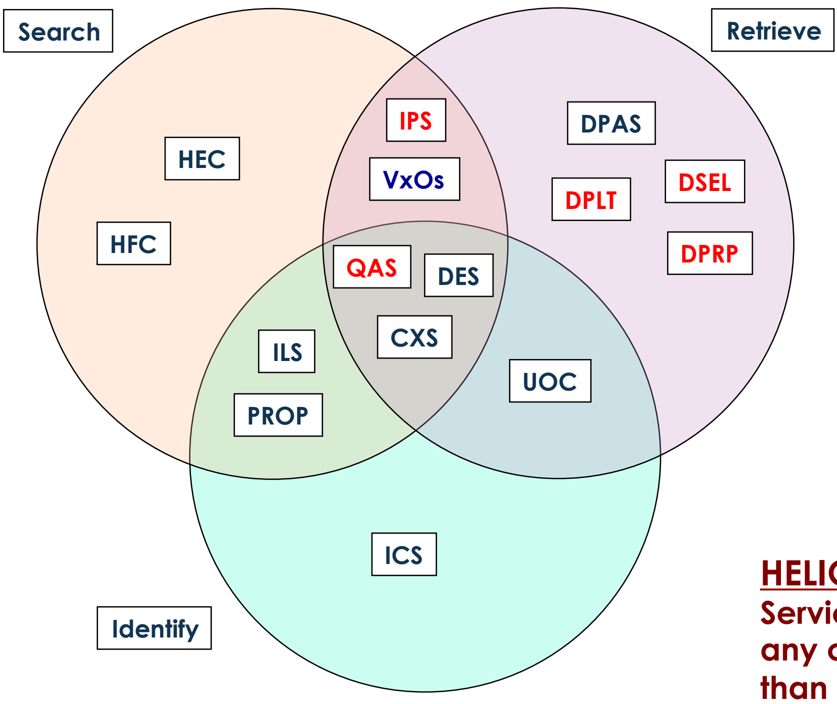
Data selection and retrieval

- ***Objective:***
  Obs. Type @ Time, Location ➔ Instrument/Observatory @ Time

- **Some services match type of observations to instruments**
  - Instrument Capabilities Service (ICS)
    - Type of observation each instrument is capable of
  - Instrument Location Service (ILS)
    - Locations of the observatories and hence the instruments
  - Observation Coverage Service (OCS)
    - Unified Observing Catalogue used to handle special cases
    - Observation Coverage Table determines if observatory active

- ***Once this is done the user should then be able to retrieve the observations they wish to use***
  - Data Provider Access Service (DPAS)
    - Location and method of retrieval described in Provider Access Table

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

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

Re-thinking the Environment

- 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
- Overlap with geo-sciences and *planetary science*
  - HELIO provides access to magnetospheric data but it is harder to relate effects caused by solar activity 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
    - Other driver of atmospheric effects become important
    - No longer how much does the Sun effect the atmosphere
- Considering how to create a more general collaborative research environment from this perspective

Services as Building Blocks

- The services should be thought of as building blocks 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
  - Services can be developed and maintained independently
  - New capability can easily be implemented as a new service
- Services interfaces need to be compliant with a set of standards in order to ensure interoperability
- If VOs and other capabilities conformed to compliant interfaces, these could become part of the tool kit
  - *Go beyond the interfaces defined in IVOA (and extensions)*
  - *May need some iteration to satisfy needs of all*

What are the advantages?

- The overlaps between environments are multi-faceted
  - Different communities have varied interests in each other
  - Technology and capabilities can be shared
- Limits of what could be shared between domains not defined
  - New science could emerge if were easier to share data, metadata and other resources
- Resources, if properly designed, could be employed elsewhere
- Components are part of a bigger picture should contribute to a general environment that can be tailored as required
- When building a capability think:
  - Can it be made into modules that are more flexible
  - Could be used by someone totally unfamiliar with the subject
- This should result in true interoperability

New File Formats?

- For the Collaborative Environment should NOT require 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

- Exploring relationships across disciplinary boundaries may lead to new types of science
  - Difficulties in providing integrated access is hindering science
- The e-Infrastructures are creating research environments for specific communities
  - More general capabilities than a virtual observatory
- A collaborative research environment for heliophysics could be established relatively easily by aligning the services and metadata
- Intersecting collaborative environments covering other sets of disciplines are also possible
- Standards that have already been developed need to be modified to facilitate the required interoperability
  - Need to involve a wide range of communities in formulating these