Using Standards to Make Data and Services in Solar system Science more Interoperable

R. D. Bentley, D. Berghmans, J. Aboudarham
and the CASSIS-Team

(SCIOPS 2013, Madrid)

Background

- Solar System Science has traditionally been undertaken within a number of separate disciplines
  - Many aspects of the system are inter-related
  - Difficult to address them because of the lack of the integrating tools and techniques
- Advances in technology have removed some barriers
  - Data are much more accessible that 20 years ago
  - Processing power not a an issue for many applications
- Very different attitudes towards managing and handling data
  - Solar data are usually quite open “missions”
  - Planetary instruments often PI oriented
- Intrinsic differences between disciplines must be addressed
  - Manifest by differing data formats and other dependencies
- Lack of interoperability is inhibiting interdisciplinary science

- CASSIS is a Coordination Action funded under Research Infrastructures within the Capacities programme of FP7
- Objective is to facilitate science within the Solar System by improving the interoperability between data and services in all domains
  - Taking things to the next level by cooperating in a number of areas
  - Enabling new combinations of interdisciplinary studies
- Builds on work of three relevant FP7-funded projects
  - HELIO, EuroPlanet RI and SOTERIA; IMPEx now of interest
- Desire is to engage as many other groups as possible in the discussions, from Europe and the rest of the world

HELIO

- The Heliophysics Integrated Observatory, HELIO, is a Research Infrastructure established under EC's FP7 Capacities Programme
- HELIO has created 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
  - Increasing data volumes require means of focusing the search
  - Search based on metadata can accommodate greater complexity
- HELIO based around a Service-oriented Architecture
  - Tasks are implemented as a set of stand-alone services
  - Easy to include services developed externally
    - If compliant with the standards for interface, etc.

URL – helio-vo.eu

Making intelligent choices

- HELIO provides integrated access to a wide range of types of observations in all the domains of heliophysics
  - Over 200 instruments from >60 observatories through >30 sources
  - Similar access no matter how the data are stored
- Many of the tools provided by HELIO are intended to help the user make an intelligent selection of the data available
  - Help the user answer a series of questions

Number of instruments and observatories difficult to count. It depends on how you class constellations (Cluster and Themis) and networks (GONG)

Observatories:

ACE, BBSO, BLEN, CALLISTO, CASSINI, CLUSTER-1, CLUSTER-2, CLUSTER-3, CLUSTER-4, CTIO, CUCS, GEOTAIL, GOES, GOES-12, GONG, HINODE, IMAGE, KANZ, KPNO, KSAC, KSFO, LEAR, MESSENGER, MEUD, MEX, MGS, MLSO, MWSO, NANC, NEAR, NOBE, OACT, OAUC, OVRO, PDMO, POLAR, PROBA2, RHESSI, SDO, SODA, SOHO, STEREO-A, STEREO-B, TEID, THEMIS-A, THEMIS-B, THEMIS-C, THEMIS-D, THEMIS-E, TIMED, TRACE, UDPR, ULYSSES, VEX, VOYAGER, WIND, YNAO, YOHKOH

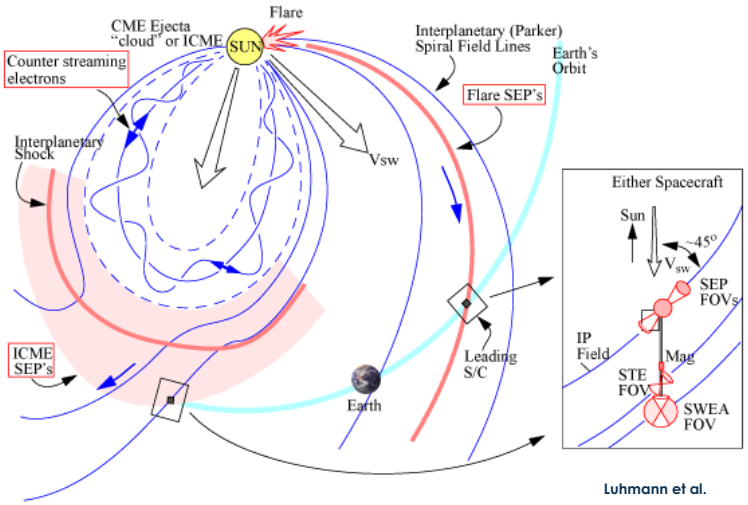
(58 observatories can be accessed)

Data Providers:

BASS2000, CDAWEB, DARTS, FHNW, HANET, HAO, HASTA, HSDCE, INAF-OACT, JSOC, KANZ, KSO, LSSP, MEDOC, MSU, MWSPADP, NAOJ-NRO, NGDC, NSO, NSO-GONG, OBSPM, OVRO, P2SC, PDS-GEO, PDS-PPI, SAO, SDAC, SFO, SHA, SODA, SSC, UKSSDC

(32 providers are being sourced)

Location and timing



Luhmann et al.

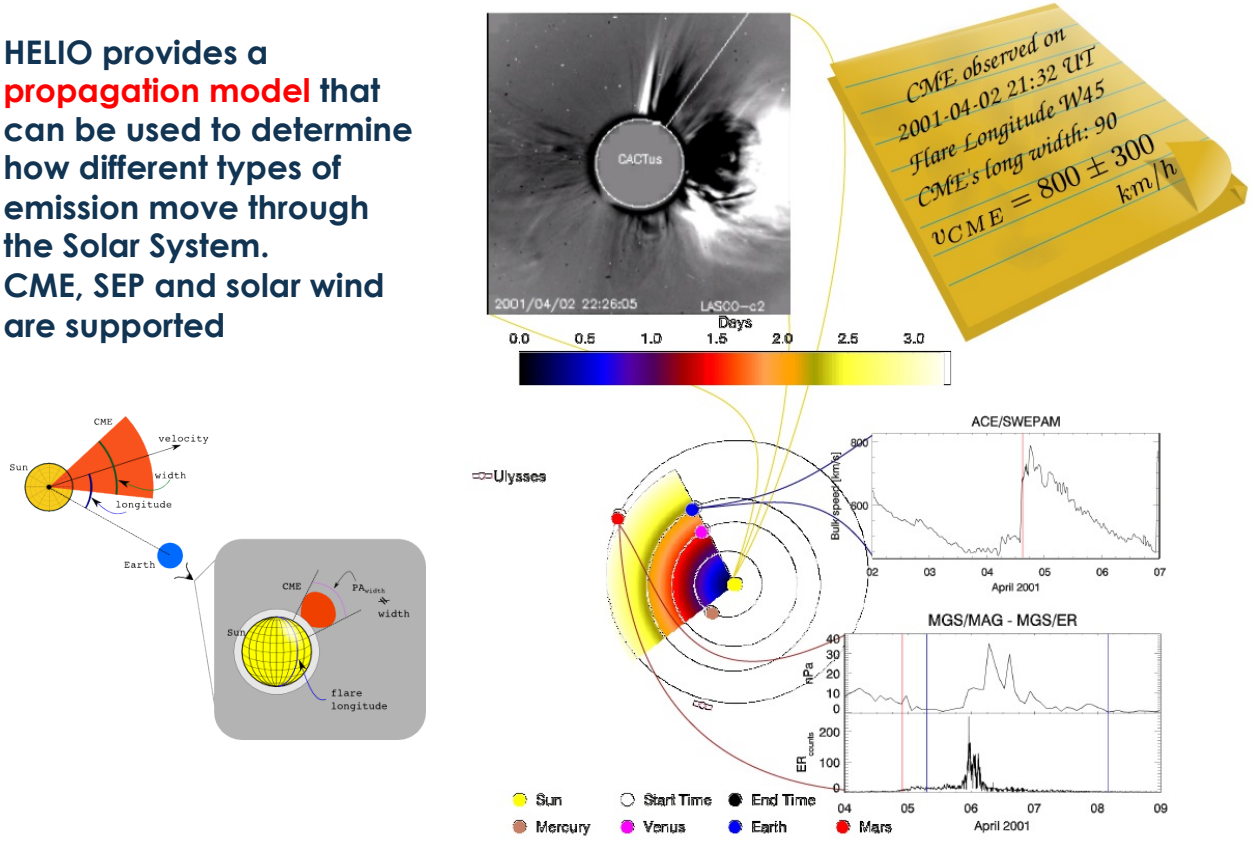
- Nature of the effect depends on the type of emission
  - EM radiation and particulate
  - Immediate and delayed
- Photons almost immediate and line of sight
- Propagation of particles influenced by interplanetary magnetic field
  - Energetic and plasma…
- CMEs crash through everything
- Modelling needed to help understand when, where and whether to look

What is happening, when and where?

- **HELIO provides tools that help the user identify interesting event and phenomena**
  - **Heliophysics Event Catalogue (HEC) –** over 65 catalogues of different types of events from various sources
  - **Heliophysics Feature Catalogue (HFC)** – 10 features solar/heliospheric
- **HEC and HFC contained events and features that have been derived a priori in a variety of ways**
  - Several types of event list from various vantage points
  - Several different *techniques* used to detect solar features
  - Have tried to standardise the parameters used to define events
- **Some events are subtle and need to know when/where to look**
  - **Data Evaluation Service (DES)** – provides the ability to examine time series data for "unrecognized" events
  - **Context Service (CXS)** – allows user to plot context information (light curves, flare locations, spacecraft location)

How are events related?

HELIO provides a propagation model that can be used to determine how different types of emission move through the Solar System.
CME, SEP and solar wind are supported



Are there suitable observations?

- HELIO provides to tools to identify suitable instruments at locations and times determined by the propagation model
  - Instrument Capabilities Service (ICS) knows about the capabilities of relevant instruments – the type of observations they make, their rough location, etc.
  - Instrument Location Service (ILS) contains detailed information about the location of observatories
  - Unified Observing Catalogue (UOC) helps filter searches involving small FOV, pointed solar instruments

- Collectively these help refine the request that will be made to the Data Provider Access Service (DPAS)
  - Provides access to large amounts of data – selection required
  - Steps needed to ensure different instrument do not collide
  - Greater standardization of file naming and metadata content would be beneficial

Europlanet RI

- The Europlanet Research Infrastructure (RI) was funded under the Capacities specific programme of FP7

- Europlanet RI followed the Integrated Infrastructure Initiative (I3) activities model, but with a different emphasis to HELIO:
  - Networking Activities – aimed at further fostering a culture of cooperation in the field of Planetary Sciences.
  - Trans-national Access Activities was to provide:
    - Trans-national access to a range of laboratory and field site facilities tailored to the needs of planetary research
    - On-line access to the available planetary science data, information and software tools, through the IDIS e-Service
  - Joint Research Activities – aimed at improving the services provided by the ensemble of Trans-national Access Activities.

- Europlanet RI complements HELIO providing access to the planetary atmospheric and ground observations

URL – europlanet-ri.eu

IMPEx

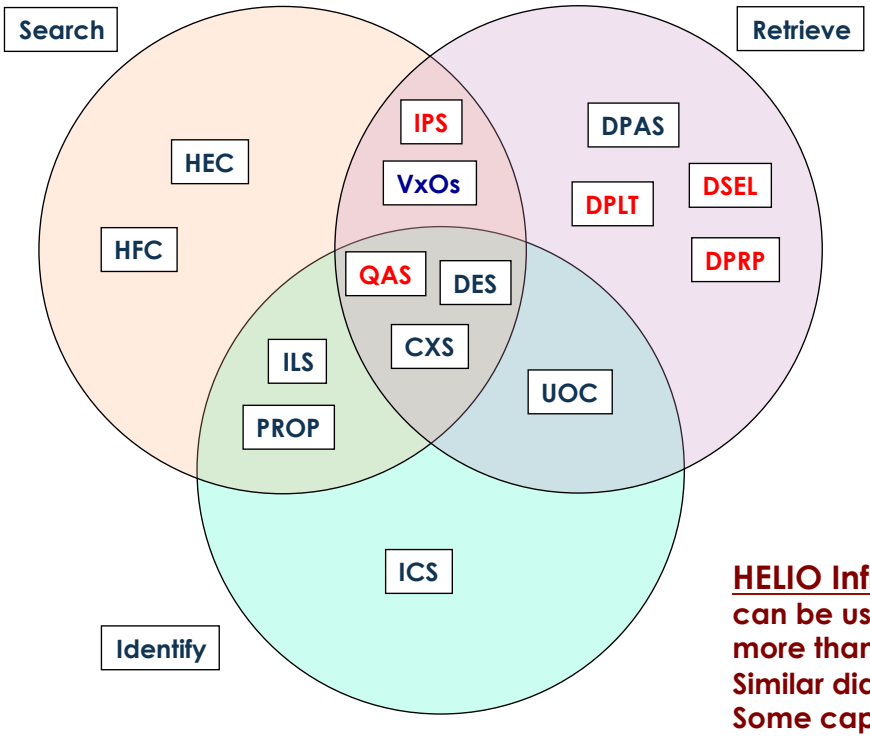
- **Integrated Medium for Planetary Exploration (IMPEx) aimed at the creation of an integrated interactive computational framework where data from planetary missions will be interconnected with numerical models**
- **Relatively new project that provides a possibility to:**
  - Simulate planetary phenomena and interpret space missions measurements
  - Test models versus experimental data and perform further improvements of models
  - Fill gaps in the measurements by appropriate modelling runs;
  - Perform preparation of specific mission operations and solve various technological tasks, including preparation of new missions
- **Complements HELIO and provides greater modelling capabilities**

URL – impex-fp7.oeaw.ac.at

Re-thinking the Environment

- **Heliophysics is the effect of the Sun on the Solar System but the boundary with other domains is fuzzy**
- **Overlap with geo-sciences and *planetary science***
  - Changes in solar activity & output can affect the Earth's atmosphere
    - Probable effects on climate; possible effects on weather
  - Emphasis of the search shifts as move into planetary environment
- **Overlap with astrophysics**
  - Sun and Solar System example of what observed at a distance
- **In trying to integrate other capabilities with HELIO we are rethinking how the environment should be established**
  - Considering how to create a more general collaborative research environment from this perspective
  - HELIO provides access to outer layers of planetary environment
  - Europlanet RI and IMPEx complement HELIO in different ways

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
Similar diagrams for other projects
Some capabilities used in all

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 new capabilities have 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***

Standards needed for Metadata

- Quality of available metadata is extremely variable
  - Some projects are better that others
  - Dependent on the community involved
- As metadata has been ingested into the HELIO services
  - Names used for parameters were standardized
  - The way that time is described is standardized
  - Try to make spatial coordinates more interrelated
  - Try to standardize the way observations described
- Problem partly because of lack of standards or guidelines
  - There are standards defined by a number of organizations
  - No overall agreement on what to use!!!!!
- Urgently need to address the problem

Services need Standard Interfaces

- HELIO Query Interface (HQI) used on most services
  - Input based on the IVOA's Parameter Query Language (PQL) and Table Access Protocol (TAP)
  - Output generally as annotated VOTable
  - Both synchronous and asynchronous requests supported
- Eurplanet RI and IMPEx have interfaces based on TAP
- HELIO Services that involve processing follow the IVOA's Universal Worker Service (UWS) pattern
  - Input usually via an XML file (more flexible)
  - Output could be data file (VOTable or otherwise), or an image
  - Wait for completion flag from service
- Services usually have both SOAP and REST interfaces
  - REST interface used to implement service GUIs and some APIs

New File Formats?

- Tidal wave of data is heading in our direction and we must maximize interoperability before we are swamped
- Should we consider a move to a family of new formats that are designed to facilitate interoperability
  - Existing formats are now decades old and were not created with interoperability in mind
    - Each has merits and is suited to a certain type of data
  - 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
- Collaborative Environment should NOT require all providers to switch to new file formats, but:
  - Lack of interoperability if hurting science
- External XML file could provide required additional metadata

Greater Interoperability

- Increased interoperability needed between the capabilities developed by the heliophysics community
  - Existing capabilities start to become part of a larger picture
  - Framework in which to contribute new capabilities (*cf.* SolarSoft)
  - *Everyone would benefit !!!!!!*
- Making the services and other capabilities interoperable is only the start
  - Need to improve the quality and contents of the metadata
  - Needs to improve the quality of the data
- The heliophysics community need to act on this sooner rather than later
  - New capabilities should be part of the larger picture

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 it be used by someone totally unfamiliar with the subject
- This should result in true interoperability

Conclusions

- Increased interoperability needed between the data and capabilities developed by the communities
- This discussion does not just relate to science
  - Philosophy of how to define systems valid elsewhere
- Scientific analysis coupled with planning and operations
  - Some overlapping of information that is needed/used
- Operational information must be stored in electronic form and preserved
  - Why something happened or did not happen could be relevant to the analysis of the data
  - Information should become part of the analysis metadata
- Including a consideration of how to enhance interoperability should be a requirement of funding