HELIO – A step into the Future

R. D. Bentley
and the HELIO Team

(EPSC 2010, Rome 21 September 2010)

HELIO, the Heliophysics Integrated Observatory, is a Research Infrastructure funded under EC’s FP7 Capacities Specific Programme

- Started 1 June 2009, duration 36 months
- Consortium includes 13 partners from 7 countries

HELIO will provide the heliophysics research community with an integrated e-Infrastructure that has no equivalent anywhere else

- HELIO will provide the ability to identify interesting phenomena and access relevant solar and heliospheric data together with related magnetospheric and ionospheric data (for planets with magnetic fields and/or atmospheres)
- Need for capability driven by desire to study problems that span disciplines
- Search base on metadata increasingly important as data volumes increase

HELIO will address its challenges following the Integrated Infrastructure Initiative (I3) activities model of the EC’s Framework Programme:

- Networking Activities used to involve the community
- Service Activities used to implement structure of the virtual observatory
- Research Activities used to investigate/develop required capabilities
Science in Context

- Heliophysics, 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’s perspective of a Generic Problem

- Identify interesting things to study
  - Search undertaken in 4-Dimensions across several domains
    - Effects occur as phenomena propagate – whether, where and when to look
    - Follow phenomena through coordinate systems as they evolve
  - Search based solely on metadata and derived products
    - Event lists and feature lists from many domains used as primary selection criteria

- Review availability of suitable observations
  - Determine whether suitable instruments at the relevant locations
    - Science objectives dictate types of observations required
  - Determine whether instrument was making observations
    - Coverage and quality of observations are addition selection criteria

- 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)
  - Optionally process selected observations (extract and calibrate)
  - Optionally return data in different/desired format

- Analysis done with users own software tools (e.g. IDL)
Service Oriented Architecture

Design of HELIO intended to split task into components that can be developed and used independently or in a workflow

- Search process intended to refine selection of what is relevant to scientific objective
  - Search based on metadata and derived products
- Only locate and retrieve data when selection process complete
  - Interaction with data providers minimized and very targeted
- Tools to return only required parts of the data

Search Process

- Objective is to turn a study that needs certain observations into a set required observations at specific locations and times
- HELIO tries to provide as diverse a set of tools as possible including metadata from all relevant domains
  - Heliophysics Event Catalogue (HEC)
    - Catalogue of existing event data
  - Heliophysics Feature Catalogue (HFC)
  - Metadata Evaluation Service (MDES)
    - Allow user to evaluate time series data and generate own event list
  - 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 process

- Objective is to turn a set of required types of observations at specific locations and times into data from specific instruments at specific times

- 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
  - Observation Coverage Service (OCS)
    - Observation Coverage Table determines if observatory active
    - Unified Observing Catalogue used to handle special cases

- 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

Defined Services

In addition to the services used to identify interesting events and then find and retrieve the data, there are a number of Enabling Services

These provide capabilities such as processing, storage, security, etc.

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

HELIIO being implemented as a set of services:

- **Identify interesting things**
  - Heliophysics Event Catalogue (HEC)
  - Heliophysics Feature Catalogue (HFC)
  - Metadata Evaluation Service (MDES)
  - Context Service (CXS)
  - Auxiliary Information Service (AIS)
  - Propagation, etc. models

- **Match to observations**
  - Instrument Capabilities Service (ICS)
  - Instrument Location Service (ILS)
  - Observation Coverage Service (OCS)
    - Observation Coverage Table
    - Unified Observing Catalogue

- **Locate and retrieve data**
  - Data Provider Access Service (DPAS)
    - Provider Access Table

Workflows…

- Service can be used individually or as part of a workflow – Taverna base-lined for HELIO

- Within event selection, the user may iterate using several services and tools until they are satisfied

- Other services then help the user go from a list of interesting times and locations to a list of Instruments and times to a list of files

- The user can then retrieve the files they need to address their science problem
**Generic Use Case**

- Solar activity includes CME’s that propagate through the solar system – user wishes to look of the consequences
- HEC used to identify time of events and location on the Sun
  - Propagation model used to determine times & locations in other parts of the Solar System
- Search tools used to identify other effects elsewhere
  - HEC identifies major effects in planetary environments
  - MDES looks for more subtle effects
- ICS/ILS identifies instruments that were suitably located to make observations; OCS confirms
- DPAS locates and retrieves data
- Services used and their order depends on the scientist and the science problem...

**Status and Networking**

- **Status**
  - HELIO being developed as a series of Releases – in middle of R3
  - Working prototypes of several service exist, some can be accessed
    - HEC, ICS & ILS, MDES, UOC, CXS, DPAS and HRS
    - Taverna Server just becoming available – workflows “invisible to the user”
      - Hand-crafted Workflows executed on Workbench on user’s system until now
    - Simple user interface
  - Lists of metadata and data that could be added has been defined

- **Networking**
  - Trying to implement a series of User Groups
    - Users that will help define Use Cases and refine system requirements
    - Users that will test and validate the system (over the duration of the project)
  - First CDAW planned for November 2010
    - Will allow us to test capabilities and get feedback
    - Driving/prioritizing the selection of metadata, data and services
  - Email user_groups@helio-vo.eu if you are interested
Instrument Location and Capabilities Services

- Prototype Location Service (ILS) based on a database
  - Ability to calculate locations on demand may be added later
- Most of the observatories and instrument that need to be added to the ICS have been defined
  - Interested in instruments that address heliophysics problems
  - Note: Inclusion does not imply that all will be available through HELIO
    - ICS key table used to assist in search for interesting events

<table>
<thead>
<tr>
<th>Obs. Name</th>
<th>Observatory Location</th>
<th>ILS-L</th>
<th>ILS-9</th>
<th>Launch Date</th>
<th>End Clues</th>
<th>End Clues</th>
<th>Status</th>
<th>Observatory Name -elongform</th>
<th>Sat. ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voyager 2</td>
<td>HP2</td>
<td>T-Sat1D</td>
<td>Heliosphere</td>
<td>1977-08-20</td>
<td>0000-00-00</td>
<td>0000-00-00</td>
<td>Voyager 2</td>
<td>V 2</td>
<td>1977-026A</td>
</tr>
<tr>
<td>Voyager 1</td>
<td>HP2</td>
<td>T-Sat1D</td>
<td>Heliosphere</td>
<td>1977-09-08</td>
<td>0000-00-00</td>
<td>0000-00-00</td>
<td>Voyager 1</td>
<td>V 1</td>
<td>1977-026A</td>
</tr>
<tr>
<td>Galileo</td>
<td>HP4</td>
<td>T-Sat1D</td>
<td>Jupiter</td>
<td>1986-02-02</td>
<td>0000-00-00</td>
<td>0000-00-00</td>
<td>Galileo</td>
<td>Galileo</td>
<td>1986-026A</td>
</tr>
<tr>
<td>Galileo</td>
<td>HP4</td>
<td>C-Sat1D</td>
<td>Jupiter</td>
<td>1989-11-18</td>
<td>1990-12-07</td>
<td>2003-09-21</td>
<td>Galileo</td>
<td>Galileo</td>
<td>1989-026A</td>
</tr>
<tr>
<td>Galileo</td>
<td>HP4</td>
<td>Jupiter</td>
<td>Jupiter</td>
<td>1990-12-07</td>
<td>0000-00-00</td>
<td>0000-00-00</td>
<td>Galileo</td>
<td>Galileo</td>
<td>1990-026A</td>
</tr>
<tr>
<td>Ulysses</td>
<td>HP3</td>
<td>C-Sat1D</td>
<td>Saturn</td>
<td>1990-10-20</td>
<td>0000-00-00</td>
<td>0000-00-00</td>
<td>Ulysses</td>
<td>Ulysses</td>
<td>1990-026A</td>
</tr>
<tr>
<td>Galileo</td>
<td>HP4</td>
<td>C-Sat1D</td>
<td>Mars</td>
<td>2001-04-07</td>
<td>2001-10-24</td>
<td>2001-10-24</td>
<td>Galileo</td>
<td>Galileo</td>
<td>2001-026A</td>
</tr>
<tr>
<td>Huygens</td>
<td>HP4</td>
<td>C-Sat1D</td>
<td>Mars</td>
<td>2003-09-27</td>
<td>2003-12-28</td>
<td>2003-12-28</td>
<td>Huygens</td>
<td>Huygens</td>
<td>2003-025B</td>
</tr>
<tr>
<td>Rosetta</td>
<td>HP2</td>
<td>T-Sat1D</td>
<td>Comet</td>
<td>2004-03-02</td>
<td>0000-00-00</td>
<td>0000-00-00</td>
<td>Rosetta</td>
<td>Rosetta</td>
<td>2004-025B</td>
</tr>
<tr>
<td>VEX</td>
<td>HP4</td>
<td>C-Sat1D</td>
<td>Mars</td>
<td>2006-11-09</td>
<td>2006-09-11</td>
<td>2006-09-11</td>
<td>VEX</td>
<td>VEX</td>
<td>2006-025B</td>
</tr>
<tr>
<td>New Horizons</td>
<td>HP2</td>
<td>T-Sat1D</td>
<td>Pluto</td>
<td>2006-04-29</td>
<td>0000-00-00</td>
<td>0000-00-00</td>
<td>New Horizons</td>
<td>New Horizons</td>
<td>2006-025B</td>
</tr>
<tr>
<td>STEREO-A</td>
<td>HP1</td>
<td>T-Sat1D</td>
<td>Sun</td>
<td>2006-11-09</td>
<td>0000-00-00</td>
<td>0000-00-00</td>
<td>STEREO</td>
<td>STEREO-A</td>
<td>2006-025B</td>
</tr>
<tr>
<td>STEREO-B</td>
<td>HP1</td>
<td>T-Sat1D</td>
<td>Sun</td>
<td>2006-10-25</td>
<td>0000-00-00</td>
<td>0000-00-00</td>
<td>STEREO</td>
<td>STEREO-B</td>
<td>2006-025B</td>
</tr>
</tbody>
</table>

Access to Planetary Data

- Access to NASA’s Planetary Data System yields best results
- Temporary solution will be used until PDAP comes into service
- Data Provider Access Service used to access archives, etc.
  - Provider Access Table describes access method, directory structure, etc
  - For planetary data, will refer to the UOC for access information
- Unified Observing Catalogue (UOC) handles special cases
  - Helps where access is indirect - e.g. for planetary data
  - Provides URL that can used to access the data
Standards and Strategy

- Developing data models to help describe the systems that we are trying integrate
  - Based on models from EGSO, SPASE, etc.
- Developing recommendation for providers about how to naming and storage strategies that would make their data archives more interoperable
- Developing standards for files that we are producing that could be extended to other groups
  - Related to IAU working group
- Addressing issues on how to ensure event metadata from different sources are interoperable
- Developed a Query Interface that is used for all services that are driven by time and other quantities
- Working on Registry services to provide easy access to services
  - Some heritage from AstroGrid and the IVOA for both

Annotation of data that are used/created

- Output of HEC and other services in VOTable format
- HELIO has decided that the output of the HEC, and other lists that could be used by external communities, should be fully annotated with UCDs, utypes etc.
- Need to develop a comprehensive data model
- Looking for help to define utypes and UCDs for the Solar System

UCDs and utypes are concepts defined by the International Virtual Observatory Alliance (IVOA)
- Existing UCDs, etc not really applicable to heliophysics: HELIO needs to develop a set!
Importance of coordinate systems

- The same feature can be seen from different vantage points
  - Need to be able to relate features seen in pairs of images
  - Need to be able to interpret coronagraph (plane-of-sky) observations from different viewpoints and relate them to motion of material in the heliosphere
  - Need to relate these to in-situ observations made at various locations
  - Need to be able to track features and follow effects into planetary environments
- Working set of coordinates has been selected; being used for development

Coordination Action for the integration of Solar System Infrastructures and Science

- CASSIS is a Coordination Action funded under Research Infrastructures within the Capacities programme of FP7
  - Started 1 June 2010, duration 36 months
- CASSIS intended to facilitate science within the Solar System by improving the interoperability between data and services in all domains
- CASSIS brings together three projects that are directly relevant to this issue
  - HELIO, EuroPlanet RI and SOTERIA – all funded under FP7
  - Relevant partners from each project, plus other key groups
- Desire is to engage as many other groups as possible in the discussions, from Europe and the rest of the world
- Web site: http://www.cassis-vo.eu
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 means that the intrinsic differences between disciplines are being addressed
  - Manifest by differing data formats and other dependencies
- Three FP7 projects – HELIO, EuroPlanet RI & SOTERIA.
  - Each is making significant improvements to the infrastructure that supports their communities, increasing ability to do science
  - Necessary to coordinate the efforts of these and other projects in order to help break down the inter-disciplinary boundaries barriers
- CASSIS intended to take things to the next level by cooperating in a number of areas
  - Enabling new combinations of interdisciplinary studies

Cooperation & Discussion

- Areas of cooperation include:
  - Investigating ways to improve the interoperability between data and metadata from the domains, and the possibility of sharing services, including metadata resources.
  - Coordinating the use of standard within the projects and reflect any changes that are required to organizations like the IVOA and IPDA.
  - Coordination of the dissemination activities of the projects in order to create a more coherent and comprehensive approach
- Two principle means of discussion:
  - Community Consultation Meeting will be used to gather input from the wider community
  - Vision for Solar System Science Workshops are planned to bring key players together in order to lobby the case for solar system science with the decision makers and funding agencies
<table>
<thead>
<tr>
<th>Short Name</th>
<th>Organization</th>
<th>Organization Details</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCL</td>
<td>University College London (MSSL and Physics &amp; Astronomy)</td>
<td>UK</td>
<td></td>
</tr>
<tr>
<td>KULeuven</td>
<td>Katholieke Universiteit Leuven</td>
<td>BE</td>
<td></td>
</tr>
<tr>
<td>FHNW</td>
<td>Fachhochschule Nordwestschweiz</td>
<td>CH</td>
<td></td>
</tr>
<tr>
<td>ROB</td>
<td>Koninklijke Sterrenwacht van Belgie</td>
<td>BE</td>
<td></td>
</tr>
<tr>
<td>INAF</td>
<td>Istituto Nazionale di Astrofisica (IFSI and Obs. Trieste)</td>
<td>IT</td>
<td></td>
</tr>
<tr>
<td>UPST</td>
<td>Universite Paul Sabatier Toulouse III (CESR)</td>
<td>FR</td>
<td></td>
</tr>
<tr>
<td>OBSPARIS</td>
<td>Observatoire de Paris (LESIA)</td>
<td>FR</td>
<td></td>
</tr>
<tr>
<td>UCLA-IGPP</td>
<td>University of California, Los Angeles (IGPP)</td>
<td>US</td>
<td></td>
</tr>
<tr>
<td>RPI</td>
<td>Rensselaer Polytechnic Institute</td>
<td>US</td>
<td></td>
</tr>
<tr>
<td>ESA</td>
<td>European Space Astronomy Centre</td>
<td>ES</td>
<td></td>
</tr>
<tr>
<td>NOAA</td>
<td>Space Weather Prediction Center</td>
<td>US</td>
<td></td>
</tr>
<tr>
<td>NASA</td>
<td>GSFC Heliophysics Science Division</td>
<td>US</td>
<td></td>
</tr>
</tbody>
</table>

Participants include groups from the three project – HELIO, Europlanet RI and SOTERIA. Groups from the US were added during the negotiation phase. CASSIS interested in broadening the international aspects of the project by extending invitations to key organisations elsewhere. Whether other disciplines should be included is being considered – e.g. terrestrial climate studies.

Summary

- HELIO is developing infrastructure in a way that could be useful to other virtual observatories and related capabilities.
- Projects have common problems even if they do not know it
  - Quality of data and metadata is a concern to all projects
  - Improved interoperability would help us all
- User involvement in important
- You local VO needs you!!