

# Developing Standards and an Environment for the Solar Orbiter data system

**Bob Bentley (UCL)**  
**David Berghmans (ROB)**  
**Andre Csillaghy (FHNW)**

## Background

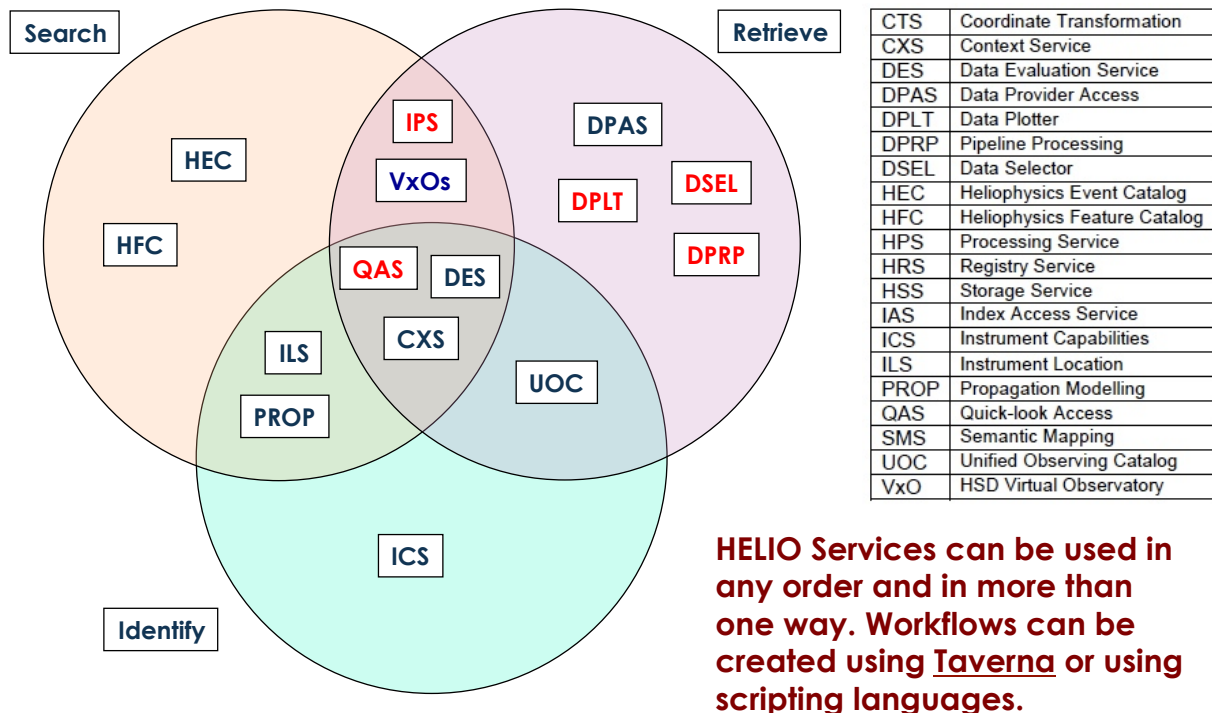
- **CASSIS is a Coordination Action that grew out of three projects funded under FP7**
  - Heliophysics Integrated Observatory (HELIO)
  - Europlanet Research Infrastructure
  - Solar Terrestrial Investigations and Archives (SOTERIA)
- **Objective is to improve interoperability in Solar System science**
- **Projects of CASSIS address similar problems in different areas**
  - All providing access to multiple archives although scale varies
  - Handling the foibles of the data created by different communities
  - HELIO addressing the widest range of disciplines and problems
    - Using a virtual observatory approach to many issues
- **Projects now looking at how to improve interoperability**
  - Many things are harder than they should be
  - Standards seen as the way to improve things

- **Solar Orbiter could benefit from the lessons learnt in order to maximize its scientific output**
- **Many problems caused by the lack of standards**
  - Each project has found its own ways of doing things
- **Working within an overall framework will improve things**
  - Standards needed for data and for service interfaces
- **Proposal is that CASSIS should help develop standards and guidelines in a number of areas**
  - Recommendations intended to improve quality and interoperability of data and capabilities
  - Desirable not to re-invent the wheel – better to improve it...
  - Very simple, low level decisions can make a bit difference

- **No matter how access to data is to be provided, following a few simple rules can help**
  - Hierarchical directory structure based on time
    - Granularity of tree depends on quantity of data
  - Filenames that help identify the file contents (include type, time?)
  - Filenames that have meaning outside of their "normal" environment
- **File format probably determined by the type of data**
  - FITS may not be the best format for everything
  - Other formats, including CDF, should be considered
  - General difficulties in supporting good annotation of parameters
- **Intelligent providers most useful to the community**
  - Requests by users are satisfied by the service (e.g. VOs)
- **Resource-poor providers could also be important**
  - Well structured archive makes access much simpler

- **File metadata should be as comprehensive as possible**
  - information should be complete and accurate
    - Files could be of limited use if information is missing
  - Minimize the amount of assumed knowledge
  - Include key engineering information (incl. why not working...)
  - VOs should be able to accommodate all file formats
    - They cannot always manage if key information is missing...
- **Metadata can be thought of in two parts**
  - Common area with information that is always needed
  - Domain and/or instrument specific area
- **Try to maximize the common metadata**
  - Use restricted family to define temporal and spatial coordinates
  - Ensure other information provided in standard way
- **XML definition of file contents may be an option**
  - Allows full annotation of the parameters

- **HELIO has adopted a service oriented architecture (SAO)**
  - Proving to be a very flexible way of addressing issues
- **Required capabilities split into set of services**
  - Developed and maintained independently
  - Interfaces compliant with standards developed by the IVOA, etc.
  - Approach makes it easy to provide additional capabilities
  - Components can be used independently or within workflows
- **Capabilities needed by Solar Orbiter should be seen as components within larger collaborative environment**
  - Use or enhance existing capabilities where possible
  - Creating new complement capabilities to enrich the environment
  - Makes it possible to have an integrated set of operational and analysis tools (talk by H. Warren)
  - Different languages can be used within the services!



**Guidelines for Data Archives**

Following a simple set of rules makes it easier to integrate a data archive into any **virtual observatory (VO)**. Below are rules that are derived from those proposed by the IAU Div. II Working Group (*Sun and heliosphere*) on [International Data Access](#).

Although the Working Group concentrates on access to solar and heliospheric data, the rules have been expressed as generically as possible and they have relevance to any archive and any VO – we urge **data providers** to follow them as far as possible.

The rules are grouped into two halves:

- [Access Method and File Format](#)
- [File Names & Metadata and Directory Structure](#) within the Archive

Issues in the first group should primarily be decided by the providers and accommodated by the VO.

Although those in the second group are also in the province of the providers, following simple rules can make a lot of difference as to how easily the required observations can be found by the VO and supplied to the scientist.

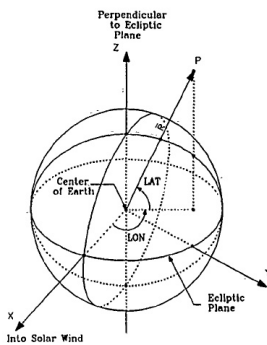
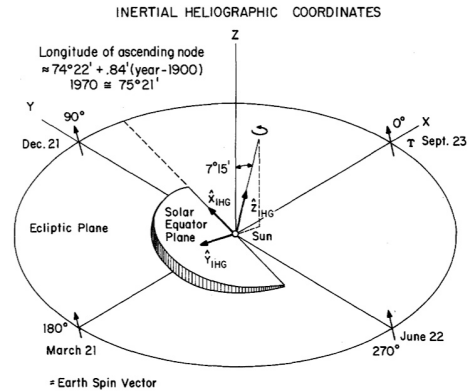
As much data as practical should be made available. From an analysis standpoint, a *“regular cadence”* with a minimum number of several observations per hour (6+) is desirable; this would make it possible to track the general evolution of phenomena although rapid changes would be missed.

*This document is open for discussion, [please contact us](#) if you have any comments.*

**Heliophysics  
Integrated  
Observatory**

**Project No.: 238969**  
Call: FP7-INFRA-2008-2

**Coordinate Systems  
Version 1.0**



Introduction.....

Remote-sensed observations related to the Sun.....

Features on (or near) the surface of the Sun.....

Features moving outward from the Sun.....

Coordinates for the location of objects in the heliosphere.....

In-situ observations of the local environment.....

Within the heliosphere.....

Far region of a planetary environment.....

Near region of a planetary environment.....

Naming convention for planetary coordinate systems.....

Questions:.....

References.....

- **Solar Orbiter must work with a wide range of data products from other parts of the Solar System to maximize the scientific return**
  - Improving its interoperability will help in this requirement
- **The project could save time and effort by building on the understanding develop by the three projects of CASSIS**
- **Working within a framework should facilitate progress**
  - Quality and content of data enhanced
  - Interoperability greatly improved
- **Beneficial to design system as a set of services that can be used in multiple ways**
- **Some guidelines are already available for discussion**
  - Data storage and coordinates
- **Recommendations for file contents available shortly**