HELIO The Heliophysics Integrated Observatory

R.D. Bentley, A. Csillaghy, J. Aboudarham, C. Jacquey, M.A Hapgood, K. Bocchialini, M. Messerotti, J. Brooke, P. Gallagher, P. Fox, N. Hurlburt, D.A. Roberts, L. Sanchez Duarte

PII: S0273-1177(10)00111-0
DOI: 10.1016/j.asr.2010.02.006
Reference: JASR 10083

To appear in: *Advances in Space Research*

Received Date: 31 March 2009
Revised Date: 5 February 2010
Accepted Date: 8 February 2010

Please cite this article as: Bentley, R.D., Csillaghy, A., Aboudarham, J., Jacquey, C., Hapgood, M.A, Bocchialini, K., Messerotti, M., Brooke, J., Gallagher, P., Fox, P., Hurlburt, N., Roberts, D.A., Duarte, L.S., HELIO The Heliophysics Integrated Observatory, *Advances in Space Research* (2010), doi: 10.1016/j.asr.2010.02.006

HELIO
The Heliophysics Integrated Observatory

R.D. Bentley [a,*], A. Csillaghy [b], J. Aboudarham [c], C. Jacquey [d], M.A Hapgood [e], K. Bocchialini [f], M. Messerotti [g], J. Brooke [h], P. Gallagher [i], P. Fox [j], N. Hurlburt [k], D.A. Roberts [l] & L. Sanchez Duarte [m]

[a] *University College London, MSSL, Holmbury St. Mary, Dorking RH5 6NT, UK*
[b] *Fachhochschule Nordwestschweiz, 5 Steinackerstrasse, 5210 Windisch, Switzerland*
[c] *Observatory of Paris, LESIA, 5 Place Janssen, 92190 Meudon, France*
[d] *Univ. Paul Sabatier, CESR, 118 Route de Narbonne, 31062 Toulouse, France*
[e] *STFC Rutherford Appleton Laboratory, Space Sci. Dept., Didcot OX11 0QX, UK*
[f] *Universite Paris-Sud, IAS, 15 rue Georges Clmenceau, 91405 Orsay, France*
[g] *INAF Observatory of Trieste, 302 Loc. Basovizza, 34149 Trieste, Italy*
[h] *Univ. Manchester, Computer Science, Oxford Road, Manchester M13 9PL, UK*
[i] *Trinity College Dublin, School of Physics, College Green, Dublin 2, Ireland*
[j] *Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA*
[k] *Lockheed Martin ATC, 3251 Hanover Street, Palo Alto, CA 94304, USA*
[l] *NASA-GSFC, HSD, 8800 Greenbelt Road, Greenbelt MD 20771, USA*
[m] *ESA-ESAC, PO box 78, Villanueva de la Caada, 28691 Madrid, Spain*

Abstract

Heliophysics is a new research field that explores the Sun-Solar System Connection; it requires the joint exploitation of solar, heliospheric, magnetospheric and ionospheric observations.

HELIO, the Heliophysics Integrated Observatory, will facilitate this study by creating an integrated e-Infrastructure that has no equivalent anywhere else. It will be a key component of a worldwide effort to integrate heliophysics data and will coordinate closely with international organizations to exploit synergies with complementary domains.

HELIO was proposed under a Research Infrastructure call in the Capacities Programme of the European Commission's 7th Framework Programme (FP7). The project was selected for negotiation in January 2009; following a successful conclusion to these, the project started on 1 June 2009 and will last for 36 months.

Key words: Heliophysics, HELIO, Virtual Observatory, Sun, Solar System, heliosphere, magnetosphere, ionosphere, services, semantics, ontology, workflow, metadata, data access, distributed data systems, interoperability, interdisciplinary

1 Introduction

Heliophysics is the study of the effects of the Sun on the Solar System and it addresses problems that span a number of existing disciplines – solar and heliospheric physics, and magnetospheric and ionospheric physics. The discipline is closely related to the study of space weather and the phenomena that affect it, but heliophysics is more generalized covering all parts of the Solar System rather that just the Sun-Earth connection.

The Heliophysics Integrated Observatory (HELIO) will provide the most comprehensive integrated information system in this domain. HELIO will deploy a distributed network of services that will address the needs of a broad community of researchers in heliophysics. It will coordinate access to the resources needed by the community, and will provide access to services to mine and analyze the data.

Of necessity HELIO needs to provide access to data archives scattered around the world. This is required in order to ensure the greatest possible coverage of events and effects – individual ground-based observatories cannot provide continuous observations and space-based platforms have been funded by different combinations of nations.

HELIO was proposed under the *Research Infrastructures* part of the Capacities Specific Programme within the European Commission's 7th Framework Programme (FP7).

HELIO was proposed under a Research Infrastructure call in the Capacities Programme of the European Commission's 7th Framework Programme (FP7). The project was selected for negotiation in January 2009; following a successful conclusion to these, the project started on 1 June 2009 and will last for 36 months.

2 Overview

In order to facilitate the study of this new discipline, HELIO needs to address issues in a number of areas related to two basic requirements, to:

- Provide integrated access to data from all the domains of heliophysics that are held in archives around the world

* Corresponding author & HELIO Project Coordinator
Email address: rdb@mssl.ucl.ac.uk (R.D. Bentley).

- Provide the means to conduct searches across the domains to identify datasets of interest

There are a limited number of protocols that can be used to access data over the Internet and the principal issues related to access that HELIO needs to address are how the data are stored in files and how these are made available to users.

The communities involved in heliophysics have evolved independently over decades and even centuries. Even though the links between the effects observed in the domains are now evident, there have been virtually no attempts to coordinate the way the communities conduct their data analysis. As a consequence, there are considerable differences in the way the communities store, describe and think about data. HELIO must therefore be able to provide integrated access to a range of file formats with contents that differ in content and terminology, and completeness and quality.

The ways files are named and stored in an archive can affect how easily the data can be accessed. Since the capabilities of archives vary, HELIO must be able to accommodate these differences without placing any excessive requirements on the providers.

There are similar issues related to the metadata used to identify data sets of interest, but there are also additional considerations.

In order to conduct a search across the domains we need to be able to relate parameters in their data sets; the most basic of these are coordinate systems used to describe when and where a phenomena was observed. Time can be expressed in a number of ways but fundamentally these are all closely related. However, the spatial coordinates used can relate to different bodies within the Solar System as a phenomenon propagates – this may require translation between and rationalization of the systems that are used.

Most effects are caused by phenomena that originate on the solar surface and, when a phenomenon is in its onset phase, a heliographic coordinates system is required. Since many observations of the Sun are remote-sensed images, they are described in a heliocentric system – a translation to heliographic coordinates (and vice versa) will therefore be needed. As the phenomenon propagates through interplanetary space, an extended form of heliographic coordinates is required so that effects can be tied back to events on the solar surface. When the phenomenon interacts with a planetary environment the effects become related to planetary coordinate systems and depend on whether the planet has a magnetic field and/or atmosphere. It is therefore necessary to be able to translate seamlessly through a series of geomagnetic and geographic coordinate systems as the phenomenon approaches the planetary surface.

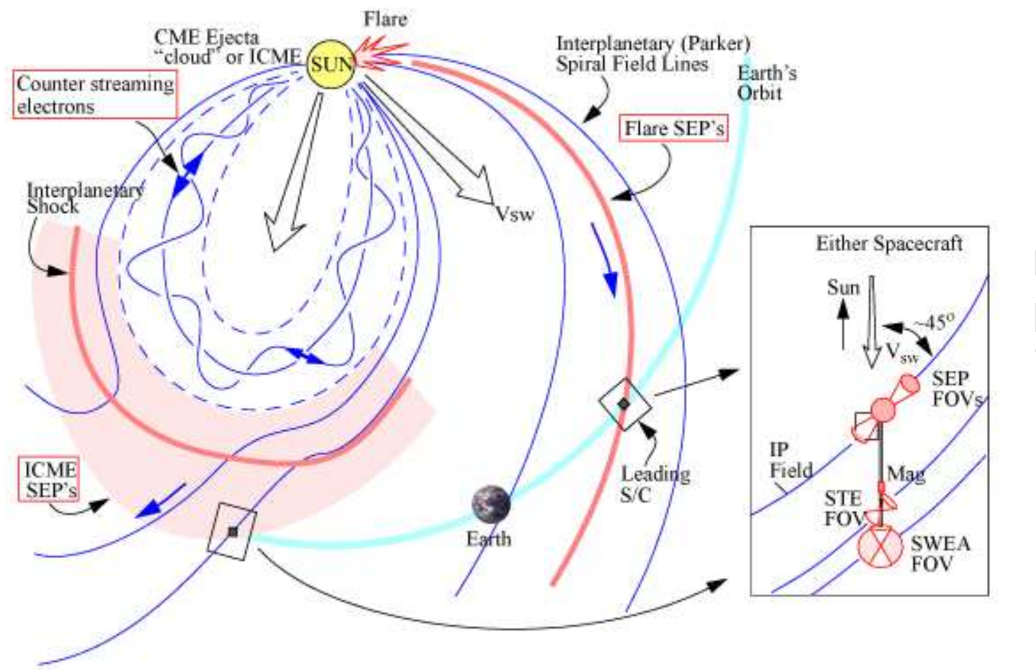

Fig. 1. Material moves out from the Sun along the lines of the interplanetary magnetic field (IMF). As the Sun rotates, the field lines lags behind and are wound into a spiral – known as the "Parker Spiral" – as a consequence, active regions producing geo-effective emissions must be located ahead of the sub-Earth point (towards the solar west limb). Variations in the solar wind speed distort the spiral a small amount since the high speed portions of the wind associates with areas like coronal holes follow a more open spiral path. Coronal mass ejections drive through the spiral at high velocity causing severe distortions – Type II radio waves and energetic particles are emitted from the shock front of the CME. The figure shows how important the location of the observer is: the STEREO-B (behind) spacecraft is affected by the CME while STEREO-A (ahead) is affected by the flare rather than the CME. *Figure courtesy of the STEREO Team, NASA*

Effects caused by photon emissions require line-of-sight view of the source and any delays are related to light travel times; those that are caused by particles occur with much longer delays and in most cases the effects are only experience if the propagating phenomena passes the observer. Knowing which effect will be observed when and where requires the solution of a 4-Dimensional problem (see Figure 1).

In order to undertake a search, not only do we need to know how to track a phenomenon as it moves between coordinate systems, we also need to understand how it propagates through interplanetary space, i.e. the path it follows and the time scales involved. HELIO will therefore provide access to a propagation model to help determine which observatories making the required type of observations are located where they could satisfy the user's search criteria (at the required time).

3 Project Implementation

HELIO will address the challenges it faces following the Integrated Infrastructure Initiative (I3) activities model of the Framework Programmes; this is split into three components:

- Networking – We will cooperate closely with the community to define the required capabilities of the HELIO infrastructure and develop new standards for heliophysical data
- Joint Research – We will develop search tools that span disciplinary boundaries and explore new types of user-friendly interfaces
- Service – We will integrate the services to produce an infrastructure that can easily be extended to satisfy the growing needs of the community

Many of the partners of HELIO were also involved in *European Grid of Solar Observations*, EGSO (Bentley et al., 2002; Csillaghy et al., 2003; Aboudarham et al., 2006), and we will build on the experience gained in that and other projects.

3.1 Networking Activities

The first tasks undertaken by the project will be within the Networking Activities. The community will be involved in the project through a series of User Groups and Focus Groups; these will be used to provide information needed to define the project requirements and resolve a number of issues. The User Groups will also help test the services established by HELIO and provide the feedback needed to improve their quality.

Principally we need to involve the community in discussions to:

- Identify the data resources and other capabilities (from all the domains of heliophysics) that should be included in the HELIO infrastructure. Items to be discussed are: *i)* which event lists, indices and other types of search metadata should be integrated into the Metadata Servers; *ii)* which data sets should be integrated into the Provider Access Servers; *iii)* which processing capabilities should be integrated into the Processing Servers. Given that there are limits to the resources that can be deployed, part of the discussion will be to prioritize these lists.
- Gather information on the data models used within the communities so that HELIO can use ontological tools to develop an over-spanning data model

in order to tie the heterogeneous sources of data together.
- Identify ways of improving the contents and interoperability of the data and metadata and promote their use within the community.
- Define new capabilities that need to be developed by HELIO in order to realize its objectives.

The third item is of particular importance for the future. We know that a virtual observatory such as HELIO can solve many problems using technology, but some require the communities to take a more coordinated and consistent approach to data and metadata. In implementing HELIO we will undoubtedly find areas where changes could improve the accessibility and quality of data. From this enhanced understanding we will formulate a set of guidelines and suggestions for data providers that we hope will provide a template for future data sets. While it could be helpful if existing access capabilities could be made compliant with such guidelines, it is essential the HELIO can handle access even if such changes are not possible so as not to place an excessive burden on the providers.

This work will be undertaken in collaboration with the data providers, in discussion with standards bodies such as the *International Virtual Observatory Alliance*[1] and the *International Planetary Data Alliance*[2], and in coordination with the *IAU Working Group on International Data Access*[3] *(solar and heliospheric).*

3.2 Joint Research Activities

If the data and metadata from the different domains are to be used collectively, it is necessary to establish the relationship between parameters in the information that they contain. Each domain has well established data models that describe its own data and we have concluded that developing a single data model to integrate the data from all the domains is not practical, nor is it realistic to expect the communities to change their practices in order to conform to such a model.

So that we can relate the data to each other, we will create an over-spanning data model to link the individual domain data models. This will be based on an ontology that we will develop from the existing models and other useful sources of information – e.g. the EGSO data model[4] (Reardon et al., 2004).

[1] IVOA: http://www.ivoa.net/
[2] IPDA: http://planetarydata.org/
[3] IAU Working Group: http://www.mssl.ucl.ac.uk/grid/iau/
[4] EGSO Report: http://www.egso.org/documents/DataModelv1.4.pdf

Space Physics Archive Search and Extract[5] data system (Harvey et al., 2004). The ontology will be used to implement a semantic driven approach needed to realize some of the HELIO services. While this will handle many of the issues, it is important that the communities work together to improve interoperability so that everything becomes simpler in the longer term – this is discussed under *Networking*.

Another area of research is related to the recognition and tracking of features. While establishing *European Grid of Solar Observations* we developed the image processing tools to detect sunspots, active regions and filaments in solar data and used these to build the EGSO Solar Feature Catalogue (Fuller et al., 2005; Zharkova et al., 2005). For HELIO, these techniques will need to be adapted to recognize features in data from other domains, including time-frequency and other types of two-dimensional plots. New techniques will be developed to allow features to be tracked when moving through and evolving in the field of view. The new catalogues that this work will facilitate will be used within the search capability provided by the Service Activities.

If our resources permit, we will also investigate new ways of presenting data through user interfaces. We have determined that many users find the interfaces provided by virtual observatories difficult to use because they do not work in the way the scientist wishes to think about the data. HELIO will try to address this problem through the innovative use of existing technology.

3.3 Services Activities

HELIO is based on a Service-Oriented Architecture and will establish a number of services handling metadata, data provider access, processing and storage.

The metadata servers will provide access to metadata products used to search for interesting observations. Given the diverse nature of the metadata, it is anticipated that in some cases it will be necessary for HELIO to ingest the metadata in order to condition it to match its data model; improving the interoperability between coordinate systems will be one core requirement. We will also look closely at the contents and quality of some of the metadata (Bentley et al., 2009) and if necessary try to improve these by regenerating certain lists in collaboration with the providers.

The search for interesting events and phenomena will produce a list of times and locations within the Solar System where observations of a particular type are required to address the science problem being considered. These will be

[5] SPASE: http://www.spase-group.org/

translated by another service into requests for observations from individual instruments (and observatories) during specific time intervals.

The data provider access service will be used to determine where observations that match the time interval and location of interest can be found; it will then allow the user to retrieve the selected observations. HELIO will not ingest any data itself, instead the provider access service will present a unified style of access into the wide variety of existing data archive interfaces.

The archives have differing levels of resources available and hence different capabilities. For archive with intelligent interfaces, the service will let the provider do most of the work; for those with limited resources, the HELIO service will establish the mapping into the archive in order to satisfy a data request.

So that we can provide access to data from as many different observatories and instruments as possible early in the project, we will initially concentrate on the more capable providers, including existing virtual observatories. The virtual observatories provide value-added access to data from the domains that they cover and could save the project a lot of effort. As mentioned earlier (see *Networking*), we are planning to work with the wider community to establish standards that will facilitate the exchange of information and data between the different domains within heliophysics; it is hoped that the virtual observatories will play a major role in this activity.

The processing services will support the production of quick-look products that are required to assist the user when searching for interesting observations; they will also execute the propagation model used to provide the 4D link between observations. Once observations have been selected, it will be possible to use the capabilities of the processing services to extract and calibrate a subset of the data and return the results in any desired format.

The storage service will be provide the option to store the results of searches and processing activities; space will be available for limited interval for those users that have logged into the system. Other services support the semantic mapping needed to relate information form the different communities, and where necessary translate between the coordinate systems used in different parts of the Solar System.

The services can be used individually or combined through a workflow capability based on the Taverna[6] workflow engine. The advanced search capability that HELIO is providing will use this workflow capability to combine the various services, including the metadata catalogues, propagation model and quick-look products.

[6] Taverna: http://www.taverna.sourceforge.net

4 Organization

The partners of HELIO provide expertise in a number of different areas. Many have an extensive knowledge of the different domains of heliophysics and experience in managing and archiving their data. Others have expertise in different areas of computer science but also have some familiarity with handling data. The partners from US will provide access to useful technologies and will help ensure that discussions and decisions include input from as large a part of the community as possible.

Involving the community is important to the HELIO project. We will do this through a number of activities: User Groups will be established to provide input and feedback; Community Coordination Meetings held in conjunction with interdisciplinary meetings like the EGU; Coordinated Data Analysis Workshops (CDAWs) are planned to give the community the chance use tools developed by the project. We hope that collectively this will lead to the acceptance of the HELIO project in the wider community.

The HELIO project Web site can be found at *http://www.helio-vo.eu*

References

Aboudarham, J., Scholl, I., Csillaghy, A., Bentley, R.D., Antonucci, E., Ciminiera, L., Finkelstein, A., Ipson, S., Messerotti, M., Pike, D., Vial, J.C. & Zharkova, V., "European Grid of Solar Observations (EGSO)", in Proc. International Heliophysical Year – First European General Assembly, 10-13 January 2006, Paris, France, C. Briand (Ed.), p. 18, 2006.

Bentley, Robert D. & the EGSO Consortium, "EGSO – the European Grid of Solar Observations", in: Solar variability: from core to outer frontiers. The 10th European Solar Physics Meeting, 9 - 14 September 2002, Prague, Czech Republic, A. Wilson (Ed.), ESA SP-506, Vol. 2. Noordwijk: ESA Publications Division, ISBN 92-9092-816-6, 2002, p. 923 - 926, 2002.

Bentley, R.D., "Building a Virtual Observatory for Heliophysics", Earth, Moon, and Planets, Volume 104, Issue 1-4, pp. 87-91, 2009.

Csillaghy, A., Aboudarham, J., Antonucci, E., Bentley, R.D., Ciminiera, L., Finkelstein, A., Gurman, J.B., Hill, F., Pike, D., Scholl, I., Zharkova, V., & EGSO development team, Integrated Access to Solar Observations with EGSO, Abstract #U21B-02, American Geophysical Union Fall Meeting, 2003.

Fuller, N.; Aboudarham, J.; Bentley, R.D., Filament Recognition and Image Cleaning on Meudon Hα Spectroheliograms, Solar Physics 227, Issue 1, pp.61-73, 2005.

Harvey C.C., Thieman J.R., King, T. and Aaron-Roberts, D., SPASE – Space

Physics Archive Search and Extract, in "Ensuring the Long Term Preservation and Adding Value to Scientific and Technical Data", ESA/ESRIN Publication WPP-232, 2004.

Reardon, K.P., Bentley, R.D., Messerotti, M. and Giordano, S., "A Solar Data Model for Use in Virtual Observatories", American Astronomical Society Meeting 204, #70.03; Bulletin of the American Astronomical Society, Vol. 36, p.796, 2004.

Zharkova, V.V.; Aboudarham, J.; Zharkov, S.; Ipson, S.S.; Benkhalil, A.K.; Fuller, N., Solar Feature Catalogues In EGSO, Solar Physics 228, Issue 1-2, pp. 361-375, 2005.