EPSC
European Planetary Science Congress

A new data mining tool for the Cluster Science Archive

A. Masson (1), E. Lopez (1), B. Martinez (1), H. Laakso (2), C. Perry (3), C.P. Escoubet (2), C. Arviset (1)
(1) ESAC/ESA, Madrid, Spain, (2) ESTEC/ESA, Noordwijk, The Netherlands, (3) RAL, UK (Arnaud.Masson@esa.int)

Abstract

Heliophysics archives mainly offer to download data for a specific time period. On the other hand, scientists usually look for data measured in a specific region of space, which fulfills a number of conditions for key physical parameters (e.g. DC magnetic/electric field, electron/ion density, temperature, velocity...), no matter when. Moreover, multi-spacecraft space missions such as Cluster and Double Star have been orbiting Earth for the last 15 years with various inter-spacecraft distances for specific scientific targets; in other words data campaigns. Such information is often difficult to find out by scientists not directly related to PI teams. For these reasons, a new data mining tool for the Cluster Science Archive is being developed by the European Space Agency. Its main features will be presented. This tool is tailored to the specificities of the Cluster and the Double Star missions and high-level datasets generated in the framework of two EU FP7 projects (ECLAT and MAARBLE). While focused on these two space missions, the approach used in this tool might be of interest for other online tools and the development of future tools. Needless to say that the goal of this presentation is also to trigger interaction with the community to improve the scientific relevance of this tool.

1. Introduction

Solar-terrestrial physics, magnetospheric dynamics and space weather combined is a major field of research. It is supported by a significant number of space missions supported by all space agencies worldwide and coordinated by the International Living with a Star (ILWS) program.

The SOHO and Cluster space missions represent a major contribution of the European Space Agency (ESA) to ILWS. SOHO is a solar observatory located at Lagrangian point L1, and operated since 1995. Cluster is the first constellation of four spacecraft flying in formation around Earth [1], enabling the first in-situ 3D mapping of the terrestrial space

Since 2000, the Cluster spacecraft relay the most detailed information on how the solar wind affects our planet in three dimensions. Science output from Cluster is a leap forward in our knowledge of space plasma physics: the science behind space weather. It has been key in improving the modeling of the magnetosphere and understanding its various physical processes. Cluster data have enabled the publication of more than 2000 refereed papers and counting.

This substantial scientific return is often attributed to the online availability of the Cluster data archive, now called the Cluster Science Archive (CSA). It is being developed by the ESAC Science Data Centre (ESDC) team and maintained alongside other ESA science missions archives at ESAC (ESA's Space Astronomy Centre, Madrid, Spain).

2. The Cluster Science Archive

The Cluster Science Archive (CSA) is the successor of the Cluster Active Archive [2], which provides access to the Cluster data since 2006 (see acknowledgements section).

CSA is a unique online archiving effort which contains the entire set of Cluster high-resolution data, Double Star high resolution data, and other related products in a standard format and with a complete set of metadata. The total amount of data format now exceeds 100 TB. The data archive is publicly accessible and suitable for science use and publication by the worldwide scientific community.

The CSA aims to provide user-friendly services for accessing data (Graphical User Interface, command line, data streaming, SAMP protocol); quick data browsing (with more than 15 TB of pre-generated plots), on-demand visualization (including particle distribution) and finally detailed documentation and various ancillary products (e.g. instrument caveat files, position of the spacecraft...). Accessing CSA requires to be registered to enable user profiles. CSA accounts more than 1,800 users and is accessible at the following address

http://cosmos.esa.int/csa

3. A new data mining tool

The CSA data mining tool is a web tool composed of three main parts:

- Key physical parameters conditions (e.g. |B|<10 nT)

- Constellation configuration&locator

- Cluster data campaigns

Once the conditions are specified, the tool will look for all time periods fulfilling these conditions. The output are eventually sent to the user by email and logged on her/his CSA user profile as a list of time periods. This list can then be used by the user to download all the relevant Cluster and Double Star datasets in one go.

4. Summary and Conclusions

The main workload for any heliophysics space mission archive is to provide the best calibrated data, with the highest time resolution measured to the scientific community. Once the data quality reaches a satisfactory level (i.e. including detailed instrument documentation, caveats, etc...), higher level services can be provided such as fast data browsing, on-demand data visualization and user profiles. The CSA data mining tool is the next step to provide further high-level data service to the community. The goal is simple: increase the scientific return of the Cluster and the Double Star missions

References

[1] Escoubet, C.P., M. Fehringer, M. Goldstein, The Cluster mission, Ann. Geophys., 19, 1197, 2001.

[2] Laakso, H., C. Perry, S. McCaffrey, D. Herment, A.J. Allen, C.C. Harvey, C.P. Escoubet, C. Gruenberger, M.G.G.T. Taylor, and R. Turner, Cluster Active Archive: Overview, 3-37, The Cluster Active Archive, Astrophysics and Space Science Proceedings, H. Laakso et al. (eds.), Springer, 2010