



InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

Case Study Proposal MOST Satellite Mission: Preservation of Space Telescope Data Focus 2 - Science

Luciana Duranti
The University of British Columbia

December 2004

Description of the project¹

The MOST (Microvariability & Oscillations of STars) satellite mission is Canada's first space telescope, designed to monitor variations in the brightness of stars with unprecedented precision and time coverage. These measurements can be used to study the structure and evolution of stars, as well as the properties of planets beyond the Solar System. Launched in late June 2003, MOST is funded by the Canadian Space Agency, and operated jointly by Dynacon Inc., the University of Toronto Institute for Aerospace Studies (UTIAS) and the University of British Columbia (UBC).

The scientific data consist of series of nearly uninterrupted measurements of star fields lasting up to two months, sampled at rates of 1 - 8 times per minute, which can result in up to about 500,000 individual files for a single target. Each file contains a set of counted charges from pre-selected subsets of light-sensitive pixels on the satellite's Science CCD (Charge Coupled Device) electronic detector, along with detailed information about the exposure (e.g., time and duration, pointing position of the telescope, which pixels were recorded, temperatures of the detector and electronics, etc.)

The scientific data and engineering telemetry are transmitted from the satellite to a network of three radio ground stations, at UTIAS, UBC and the University of Vienna. The repository for the scientific data is currently UBC, where the raw data from the satellite are transformed into FITS (Flexible Image Transfer System) format – the most widespread standard in the global astronomical community. The data are reduced into series of intensities vs. times by the 7 members of the MOST Science Team and their immediate collaborators. The raw and reduced data are made available to the Team members via an internal password-protected website. The measurement values in the raw data should never be updated, although the format and supporting information could be augmented. There are various methods to reduce the data, so several different data products can be associated with a single set of raw data. A set

¹ For more information about the project, see: <http://www.astro.ubc.ca/MOST/>.

of MOST observations is subject to a proprietary period of one year, during which it is available only to the project's Science Team members for their research. After this time, they are to be placed in a public archives accessible to all members of the astronomical community and the general public

The layout and information content of MOST data are different from other existing astronomical archives, which tend to consist of 2-dimensional CCD images or 1-dimensional spectra. The accuracy of some of the information (e.g., exposure time and duration, detector temperature, etc.) is much more critical to extract scientific information from these data than in many other astronomical catalogues. Therefore, the MOST Team hopes that the collaboration with InterPARES will lead to improved strategies and protocols for storing and distributing the data efficiently (at the MOST end) and productively (at the user end).

Rationale for case study

MOST falls within the research scope of Focus 2 (scientific records) of InterPARES 2. The goal of the MOST project is to collect and analyse astronomical data. This case study will provide the opportunity to address the issue of the long-term authenticity and preservation of digital entities that do not belong in the traditional sphere of responsibility of archives and libraries and partake in the characteristics of both. The InterPARES 2 research questions on which it will be necessary to focus are those related to:

1. The nature, characteristics and attributes of the digital entities in question;
2. The processing of those entities and what it does to their accuracy, reliability and authenticity;
3. The determination of the authoritative version(s) of each data set, that is, the identification of the data sets whose authenticity must be verifiable through time;
4. The existing identifying metadata, and the metadata sets necessary for long term authentic preservation;
5. The legal, ethical and moral obligations regarding the use, accessibility and preservation of the data sets;
6. The technological and intellectual tools that would assist creators to generate records that can be authentically preserved over time;
7. The applicability of the preservation concepts, methods and models developed by InterPARES 1 for the administrative and legal records created in databases and document management systems; and
8. The existence of policies, procedures, and standards that control these data sets creation, use, maintenance, preservation, and their adequacy.

Methodology

The MOST case study will be a collaboration between InterPARES researchers, and MOST researchers at the University of British Columbia

Leader: Luciana Duranti, the University of British Columbia, InterPARES Project director

InterPARES Researchers:

- Reagan Moore, Data Intensive Computing, San Diego Supercomputer Center, and InterPARES Focus 2 researcher
- Bart Ballaux, UBC Graduate Research Assistant;

MOST Researchers:

- Jaymie Matthews, Astronomy & Astrophysics, the University of British Columbia, and MOST project
- Rainer Kuschnig, Department of Physics and Astronomy, the University of British Columbia, and MOST project
- Andrew Walker, MOST project
- Greg Burley, MOST project
- Heather King, MOST project
- Jordon Johnson, MOST project

Timeline

An information meeting involving the first four MOST researchers, the case study leader and the GRA took place on November 22nd in the InterPARES office to discuss the goals, timeline and specific focus of the case study. Following approval of the case study proposal by the InterPARES International Team in December 2004, and approval by the Ethics Committee at the University of British Columbia, the project will commence at the beginning of January and progress according to the timeline below:

Jan. 2005	Analysis of all accessible documentation related to the processes and procedures followed by the MOST project in the collection, processing, maintenance and distribution of the data sets
24 Jan. – Feb. 15, 2005	Interviews (the first formal on the 24th, and then follow-ups as needed)
Feb. 20-24, 2005	Submission to the Focus 2 researchers of the answers to the “23 case study questions”
Feb. – Mar. 2005	Development of activity models and diplomatic analysis
Apr. 2005	Draft of final report
May-June 2005	Refinement of final report after feedback
June 2005	Submission of final report to InterPARES International Team