



# InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

## Diplomatic Analysis

### Case Study 26:

## Microvariability & Oscillations of Stars (MOST) Satellite Mission - Preservation of Space Telescope Data

Sherry Xie, UBC

May 2006

### INTRODUCTION

The InterPARES 2 case study 26, *Microvariability & Oscillations of Stars (MOST) Satellite Mission - Preservation of Space Telescope Data*, proposed in December 2004 at the Sicily workshop, examines the process of gathering and maintaining space telescope data.

The MOST 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.

The following text presents the results of the diplomatic analysis on the digital entity identified in the case study Final Report. The Report indicates that "[t]he MOST researchers consider the original data and the FITS files as the most important entities in the project."<sup>1</sup> Since the FITS files are created from the original raw data and can be recreated at any time needed, the original raw data is hence considered by this diplomatic analysis the most important digital entity in the project, and will be examined in the following process with respect to the identification of its status as record.

The purpose of the diplomatic analysis is to assess the status of the identified digital entity as a record. Once the status of the digital entity has been determined, preservation strategies may be proposed by Domain 3.

### IDENTIFICATION OF RECORD(S)

A record, as defined by the InterPARES glossary, is a document made or received and set aside in the course of a practical activity. A record must also possess all of the following five

---

<sup>1</sup> In answer to core research question 2: *Which of these activities generate the digital entities that are the objects of your case study?*

components, as established by InterPARES 1 research conclusions: fixed content and form, embedded action, archival bond, persons and contexts. The application of the definition of a record to the creator's digital entities is therefore analyzed according to the following parameters:

**1. To be identified as a record, the digital entity must possess fixed content and form,<sup>2</sup> and be affixed to a stable medium (or physical carrier).**

- **The content of the raw data is fixed.**

The raw data in this case study refers to the packaged sds1 and sds2 data. The sds2 files are created immediately after the images of target stars are taken by the camera on the satellite, containing all critical data and some supplementary data. The sds1 files are a compacted version of the sds2 files created for storing more data on the satellite in case there are communication problems and data cannot be transferred to earth. The sds1 and sds2 files are processed in the satellite to form sds raw files and then sent to one of the ground stations participating in the MOST project. The sds raw data are decomposed into sds1 and sds2 files on earth, and sds2 files are used to create FITS files. Since the sds1 files are compact versions of sds1 files, they are not used for further data processing unless sds2 files are not available due to certain circumstances. As revealed by this process in which the activities of taking images of stars, packing, sending and unpacking data are all done automatically, there are no changes made to the images after they are taken. Also, being the original captures of the movement and brightness of the target stars, these data serve the function of being the basis of future research activities. The content of the images is stable after creation and remains stable in subsequent research activities.

According to the answers to the core research question in the Final Report:

Q.13. *How are changes to the digital entities made and recorded?*

In general, digital entities are not changed: sds files are never changed, nor deleted.

Q.18. *Does the creator keep the digital entities that are currently being examined? That is, are these digital entities part of a record keeping system? If so, what are its features?*

All sds files are preserved without changes or additions.

---

<sup>2</sup> The InterPARES1 Authenticity Task Force has defined fixed form as the following: 1) binary content of the record, including indicators of documentary form, must be stored in a manner that ensures it remains complete and unaltered, and 2) technology must be maintained and procedures defined and enforced to ensure that the content is presented or rendered with the same documentary form it had when set aside. (See ATF Research Methodology Statement, available at: [http://www.interpares.org/documents/interpares\\_ResearchMethodologyStatement.pdf](http://www.interpares.org/documents/interpares_ResearchMethodologyStatement.pdf)).

- **The documentary form<sup>3</sup> of the raw data is fixed.**

The individual sds file generated on the satellite is stored in .dat file format and is encoded using KISS code that is prevalent in the amateur satellite community. Each file contains both image information and telemetry information, such as temperatures, pointing errors, voltages, time stamps. The sds raw file is stored in .bin file format which contain one or more individual sds files. Its maximum file size is 100KB and individual sds files may span across two consecutive sds raw files. This is handled by the software. The use of such software to store and encode the data files gives the raw data its documentary form, which remains stable as long as the same software is used to interpret/display the data.

- **The raw data is fixed to a stable medium.<sup>4</sup>**

After the transfer from the satellite to the ground station, the sds raw data is stored in the computers of the research unit. The hard drives of the computer serve as a stable physical carrier of the data.

According to the answers to the core research question in the Final Report:

Q.8. *What are the documentary and technological processes or procedures that the creator follows to identify, retrieve, and access the digital entities?*

The files are stored on the central computer and a second computer, and are regularly backed up.

**2. A record must also participate in an action, defined as the conscious exercise of will by an officer of the creator or by an external person, aimed to create, maintain, modify or extinguish situations. A record results as an unintended by-product or product of the action.**

The raw data are generated in order to for scientists to conduct astronomic research, which is mandated by the MOST project's mission. These data are instruments in the process of carrying out this action.

According to the answers to the core research question in the Final Report:

Q.3. *For what purpose(s) are the digital entities you have examined created?*

All digital entities are created to analyze the behaviour of stars. As such, the data in the digital entities will be used and analyzed in scientific publications and presentations to enhance our understanding of stars.

---

<sup>3</sup> The definition of "documentary form" in the InterPARES 2 Terminology Database is: "The rules of representation according to which the content of a record, its administrative and documentary context, and its authority are communicated. Documentary form possesses both extrinsic and intrinsic elements."

<sup>4</sup> The definition of "medium" in the InterPARES 2 Terminology Database is: "The physical material or substance upon which information can be or is recorded or stored."

**3. A record must possess an archival bond, which is the relationship that links each record to the previous and subsequent record of the same action and, incrementally, to all the records which participate in the same activity. The archival bond is originary (i.e., it comes into existence when a record is made or received and set aside), necessary (i.e., it exists for every record), and determined (i.e., it is characterized by the purpose of the record).**

The sds raw data possesses an archival bond with other digital entities generated in the process of carrying out the action, i.e., the study of stars.

As introduced in the above sections, the sds raw data is the original capture of the images and related information about the star under examination. They are saved and accumulated on the hard drives of the computers in the research unit. They are the basis on which FITS files are created and then used for all subsequent scientific data reduction and analysis. The set(s) of sds raw file used for the creation of the FITS files is indicated in the FITS files' metadata field, which are required by the file format. The relationships between the accumulated sds raw data and between the sds raw data and FITS files are considered archival bond, since the relationships:

- a. *are originary*: the saving of sds raw data on the hard drives of the central computer and the saving of created FITS files upon completion to the researcher's computers for next-step processing, satisfy the requirements of setting records aside;<sup>5</sup>
- b. *are necessary*: every data set, as identified primarily by the name of the target star and secondly by the capture time, has linkages with the previous and subsequent data sets. The study of the star involves the analysis of all data sets resulting from the same target over time; and
- c. *are determined*: the capture and organization of the data sets serve the purpose of conducting astronomic research. As indicated in the Final Report, the organization of the digital entities reflects the way how researchers use them.<sup>6</sup>

**4. Record creation must involve at least three persons, whether or not they explicitly appear in the record itself. These persons are the author, addressee and writer; in the electronic environment, one must also take into account two additional necessary persons: the creator and the originator.**

- The record's **author** is the physical or juridical person having the authority and capacity to issue the record or in whose name or by whose command the record has been issued.

The author of the sds raw files is the MOST project, which is established to conduct the research and responsible for its management and operation.

- The **addressee** the physical or juridical person(s) to whom the record is directed or for whom the record is intended.

Within the first year after the data creation, the addressee of the data is the Project (including researchers in all ground stations).<sup>7</sup> After this time period, they are open to all members of

---

<sup>5</sup> The definition of "set aside" in the InterPARES 2 Terminology Database is: "Of a record that is filed or archived, by assigning to it a classification code, including it in a folder or consigning it to a recordkeeping system, or that is associated with other records by any other means."

<sup>6</sup> See the answer to core research question 4f in the Final Report.

the astronomical community and even to the general public. Since this is set in the contract between the Project and the Canadian Space Agency (CSA), meaning the public access is part of the Project's objective, the astronomical community and general public are also addressees of the data.

- The **writer** is the physical or juridical person having the authority and capacity to articulate the content of the record.

Although the content of the sds raw data is captured by the satellite camera automatically, the writer of the data is also the Project since it is its capacity that brings the technologies and contractors together to create the data.

- The **creator** is the person in whose fonds the record exists.

The creator of the raw data is the MOST project. It qualifies as the fonds creating agency because it has distinct mandate and structure and operates on its own standing.

- The **originator** is the person to whom the Internet account issuing or the server holding the record belongs.

Each ground station participating in the Project has its own Web site for public communication purpose, which include some of the records generated from the Project. The complete sets of data, after the one year proprietary time, are hosted by the Canadian Space Agency's Canadian Astronomy Data Center (CADC).

**5. Finally, a record must possess an identifiable context, defined as the framework in which the action in which the record participates takes place. The types of context include juridical-administrative, provenancial, procedural, documentary, and technological.**

- The **juridical-administrative context** is the legal and organizational system in which the creating body belongs.

The MOST research project is a project that was initiated and founded by researchers at UBC and UTIAS (University of Toronto Institute for Aerospace Studies), and was approved and funded by the Canadian Space Agency (CSA). As part of the university community, the MOST researchers are operating in an academic environment. There are few formal rules, norms or standards that apply to the research.

The researchers acknowledge that they are not aware of any rules or norms that would specifically apply to the MOST research.

The MOST research team reports to NSERC and CSA.

---

<sup>7</sup> In its answer to core research question 4, *Do external users have access to the digital entities in question? If so, how, and what kind of uses do they make of the entities?*, the Final Report mentions that a very limited number of external users, who are researchers of other scientific institutions, also have access to the data. However, since the data are sent to these external users by the MOST researchers, they are not the addressees of the data intended in the first place.

- The **provenancial context** refers to the creating body, its mandate, structure and functions.

#### Mission

(Inferred). The MOST satellite mission is a space telescope, designed to monitor variations in the brightness of stars to use for study of the structure and evolution of stars, as well as the properties of planets beyond the Solar System.

#### Mandate

The MOST researchers are given the responsibility (by the CSA) to gather and use the data that are transmitted from the satellite. Whether the project is prolonged or not, is based on the successful continuous completion of this responsibility as evaluated on a yearly basis by the funding agency.

#### Functions

The main function of the MOST project is scientific research. From the moment the data are captured, a process of data collection starts. Data, captured from various sources and stored in various files are brought together in one record: a FITS file. After creating the FITS file, the interpretation phase begins. The interpretation process is done by reducing the data. Because various techniques and theories can be used to come to satisfactory results, this part of the research takes a long time. Reductions and interpretations of the reductions are finally presented and published.

#### Structure

The University of Vienna became a partner after the project was started. It has a 'special' institutional status as partner. As a non-Canadian partner, it does not get funding of the CSA. However, in a scientific perspective, it is considered as a full partner by the two Canadian research units. The University of Vienna built a third ground station in Austria which resulted in the opportunity to transmit more data from the satellite.

There is no official written organizational structure within the research group. Specific tasks are assigned to one or various researchers. All researchers have a specific field of expertise, for instance the instrument scientist and the software developer. The principal scientist is accountable to the funding agency.

The principal researcher is a professor in Astronomy. The other MOST researchers are hired for the MOST project. They have a background in astronomy, either on a Master or Ph.D. level. There have been no changes in the work force since the beginning of the project.

- The **procedural context** comprises the business procedure in the course of which the record is created.

A MOST researcher has made an internal document ('MOST archiving manual,' date August 25, 2004) in which the procedure for creating (naming and storing) the digital entities, is explained.

Diplomatic analysis of **procedural phases** related to the creation of the FITS files can be broken down as follows:<sup>8</sup>

- a. **Initiative:** the introductory phase of any procedure is “constituted by those acts, written and/or oral, which start the mechanism of the procedure.”<sup>9</sup>

The initiative phase of the procedure of creating FITS files is to receive the raw data from the satellite.

- b. **Inquiry:** this preliminary phase “is constituted by the collection of the elements necessary to evaluate the situation.”<sup>10</sup>

The inquiry phase of the procedure of creating FITS files is to gather all metadata necessary for understanding the raw data.

- c. **Consultation:** this phase is “constituted by the collection of opinions and advice after all the relevant data has been assembled.”<sup>11</sup>

The consultation phase of the procedure of creating FITS files is to study the raw data and their metadata with regard to data analysis and reduction.

- d. **Deliberation:** this phase is “constituted by the final decision-making.”<sup>12</sup>

The deliberation phase of the procedure of creating FITS files is to decide how to analyze and reduce the raw data.

- e. **Deliberation control:** this phase is “constituted by the control exercised by a physical or juridical person different from the author of the document embodying the transaction, on the substance of the deliberations and/or on its forms.”<sup>13</sup>

The deliberation control phase of the procedure of creating FITS files is to test the decided means of analyzing and reducing raw data.

- f. **Execution:** “the documents created in this phase are the originals of those embodying the transactions.”<sup>14</sup> In other words, the execution phase results in the issuing of the first record capable of producing the consequences intended by its author.

The execution phase of the procedure of creating FITS files is to generate data products, such as light curves, from raw data through data analysis and reduction.

---

<sup>8</sup> The phases of procedure as dictated by Diplomatic Analysis; see Luciana Duranti, *Diplomatics: New Uses for an Old Science* (Lanham, Maryland and London: The Scarecrow Press in association with the Society of American Archivists and the Association of Canadian Archivists, 1998), 115.

<sup>9</sup> Ibid.

<sup>10</sup> Ibid.

<sup>11</sup> Ibid.

<sup>12</sup> Ibid.

<sup>13</sup> Ibid.

<sup>14</sup> Ibid., 116.

- The **documentary context** is defined as the archival fonds to which a record belongs and its internal structure.

All data and data products generated in the process of carrying out research activities belong to the Project, and they are saved on the Project-owned server. The digital entities are given a unique name, based on the target (star) and the time. The files are stored on the central computer and a second computer, and are regularly backed up. The files are retrieved as they are used in the reductions. The digital entities are organized by target (star) and date and, thus, reflect the creation process.

- **The technological context** is defined as the characteristics of the technological components of an electronic computing system in which records are created.

The technological requirements for the Project are equipment (satellite camera, and satellite processors, transmission equipment, computers), special custom made software, and astronomical file formats.

## CONCLUSIONS

Based on the above analysis, the sds raw data collected by the satellite and the various data products generated using the raw data are all records of the MOST project.