InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

Title: **Case Study 26 Final Report: Microvariability & Oscillations of Stars (MOST) Satellite Mission – Preservation of Space Telescope Data**

Status: Final (public)

Version: 2.0

Submission Date: November 2005

Release Date: September 2007

Author: The InterPARES 2 Project

Writer(s): Bart Ballaux
School of Library, Archival and Information Studies,
The University of British Columbia

Project Unit: Focus 2

URL: http://www.interpares.org/display_file.cfm?doc=ip2_cs26_final_report.pdf

---

† A complementary report with the same title was subsequently published by this author in 2006 in *American Archivist* 69(1): 139–158.

Table of Contents

Version control ii
A. Overview of Case Study 1
B. Statement of Methodology 2
C. Description of Context 2
- Provenancial Context 2
  - Name 2
  - Location 2
  - Origins 2
- Juridical-Administrative Context 3
  - Legal status 3
  - Norms 3
  - Resources 4
  - Governance 4
  - Mandate 4
  - Philosophy 4
  - Functions 4
- Procedural Context 4
- Documentary Context 5
  - Document creation activities 5
  - Documents resulting from activities 6
  - File organization 7
- Technological Context 7

D. Narrative Answers to the 23 Core Research Questions 8
E. List of terms and abbreviations 14
F. IDEF0 Activity Model 15

Version control

Version 1_1	February 18, 2005	First draft	Bart Ballaux
Version 1_2	April 19, 2005	Second draft	Bart Ballaux
Version 1_3	November 25, 2005	Third draft	Bart Ballaux
Version 1_4	July 28, 2007	Editorial revisions	Jim Suderman
Version 2	September 2007	Copy editing, addition of Section F	Randy Preston

A. Overview of Case Study[2]

The MOST (Microvariability & Oscillations of STars) satellite mission is Canada's first space telescope. The telescope is 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 (CSA), 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 preselected 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 versus 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 Web site. 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 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 archive accessible to all members of the astronomical community and even to the general public. The layout and information content of MOST data are different than 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 times 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).

[2] This introduction is taken from the MOST Satellite Mission – Preservation of Space Telescope Data InterPARES 2 Case Study Project Proposal, 2004, pp. 1-2. Available at http://www.interpares.org/display_file.cfm?doc=ip2_satellite_mission.pdf.

B. Statement of Methodology

The results of this case study are based on the InterPARES 2 methodology for conducting case studies. Interviews and documents provided by the case study subjects form the core of the information. The interviews were conducted with the MOST researchers at UBC because they have the responsibility to make the data publicly available after one year.

The first interview was a group session in which five MOST researchers participated. Depending on the type and scope of the question, one of the five researchers provided information. This initial interview was taped and transcribed.

In addition to the initial interview, some follow-up interviews took place to get a more in-depth insight into the processes and techniques used by the MOST research group. Finally, some information was provided through e-mail.

The interviews were semi-structured and designed to provide the data necessary to answer the same 23 questions common to all case studies of the InterPARES 2 Project.[3]

C. Description of Context

Provenancial Context

Name

Microvariability & Oscillations of STars (MOST) project

Location

The MOST research project is headquartered at The University of British Columbia (UBC), Vancouver, B.C., Canada. Other sites of the MOST project include Toronto and Vienna. The researchers involved at these sites are not case study subjects, and are not directly involved in the issues that are under investigation in this case study. Since UBC is the 'official' repository of the data/records of the MOST project, only the UBC researchers were interviewed.

Origins

The MOST research project was initiated by researchers at UBC and the University of Toronto Institute for Aerospace Studies (UTIAS). It is funded by the Canadian Space Agency (CSA). The research began following the launch of the satellite in 2003 and will continue as long as the satellite and camera are functioning and the researchers of the various participating scientific institutions make use of the data that are transmitted from the space telescope.

The CSA normally provides annually renewed funding for projects that continue to operate successfully. Funding is limited to one year, so that the CSA's obligations cease if the satellite or camera fails to work, for example. Every year, MOST and CSA make a new scientific support agreement. This means that there is no foreseeable end in the creation of data/records.

The University of Vienna became a partner after the project was started. As a non-Canadian partner, it does not get funding from the CSA. It is considered to be a full research partner by the two Canadian research units. The University of Vienna built a ground station in Austria enabling increased data transmissions from the satellite.

[3] The 23 questions are available at http://www.interpares.org/ip2/display_file.cfm?doc=ip2_23_questions_for_case_studies.pdf.

Juridical-Administrative Context

Legal status

The MOST project is a collaborative research project involving the three participating research organizations (UBC, UTIAS, and the University of Vienna) and a commercial contractor. Dynacon Inc. is the engineering firm that created the satellite, telescope and camera that are currently used to generate the data. Dynacon Inc. has a contract with the CSA to build, operate, and maintain the satellite. All technical matters and issues related to the satellite are thus the sole responsibility of Dynacon Inc.

The project has a dual organization: technical and scientific. This case study focuses wholly on the scientific part of the project, specifically that relating to the activities and responsibilities of the UBC-based research group. All other aspects of the project are excluded.

As an approved CSA-project, MOST is subject to the guidelines and standards that apply to CSA and Canadian scientific missions. Although it is not a stipulation in the contract between the MOST project and CSA, the scientific data are expected to be publicly available after one year. The one year proprietary period starts at the moment that the data are transmitted.[4] MOST data becomes publicly available when it is transmitted to the Canadian Astronomy Data Centre (CADC). There is no written agreement with CADC for this process.

No intellectual property rights are asserted by MOST over public data, however, it is requested that the MOST project be acknowledged where its data is used.[5] During the proprietary period, the data are only distributed to researchers of the project and to a limited number of other researchers (e.g., referees of articles).

Intellectual property is an issue as far as the instrument (telescope) and the satellite are concerned, but this aspect is out of the scope of this case study.

Norms

As part of the university community, the MOST researchers are operating in an academic environment. The MOST research team reports to the Natural Sciences and Engineering Research Council of Canada (NSERC) and CSA. There are few formal rules, norms or standards that apply to the research.

As part of the astronomical research community, the MOST researchers use some technical guidelines and norms in regard to the data. For instance the choice of the Flexible Image Transport System (FITS) file format for storing the astronomical data is based on best practice in the field.[6] The metadata or descriptive fields that are attached to the FITS files were partly imposed by the file format, and partly chosen by the MOST researchers.

Because the MOST research is pioneering in the sense that it accumulates a specific type of data over time, new astronomical projects have contacted the MOST researchers with questions related to handling data.

[4] In other projects, such as the COROT mission, the proprietary period of one year (more or less the standard), starts at the moment that the reductions and analyses have been generated. Since one of the major scientific issues in the MOST project relates to the creation of a good, 'final' reduction scheme of the data, the scientists indicated that they would have preferred a proprietary period that is similar to the one of the COROT mission.

[5] Moreover, the FITS files contain a descriptive field that indicates that the files were created in the MOST project.

[6] Originally developed by NASA in 1979, FITS is the standard computer data format widely used by astronomers to transport, analyze and archive scientific data files. For more information about FITS, see http://fits.gsfc.nasa.gov/fits_documentation.html.

Resources

The satellite transmits data to three ground stations in Vienna, Toronto and Vancouver. The MOST research team at UBC has one large, secured office in the Astronomy Building. There are several computers on which data are stored, but one is specifically designated as the main data store. The office is only used by MOST researchers and participating students. The main computer and ground station devices are located in a separate compartment that can be locked. The satellite dish is located on the roof of the Astronomy Building.

Governance

There is no formal organizational structure within the research group, although the principal scientist, an Astronomy professor, is accountable to the funding agency. The other MOST researchers are hired for the MOST project and all have a background in astronomy, either at a Master or Ph.D. level. There have been no changes in the work force since the beginning of the project.

Tasks are assigned to individuals or groups. All researchers have a specific field of expertise, including an instrument scientist and a software developer.

Mandate

[Inferred] The MOST satellite mission is a space telescope, designed to collect data on 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.

Philosophy

The MOST researchers are astronomers and are thus guided by best practice, ethics, standards and guidelines relevant in that field of study.

Functions

The main functions of the MOST project are to gather and interpret data and publish findings. Data, captured from various sources and stored in various files, are brought together in one record: a FITS file. At this point 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.

Procedural Context

The MOST researchers have to fulfill some basic administrative duties related to the scientific process. Some of these duties are described in manuals (e.g., the MOST Archiving Manual). This manual indicates what should be done on a daily and weekly basis to successfully effect “processing, archiving and backing up [of the] MOST data.”[7]

The specific responsibilities of the MOST members are not formal (i.e., written assignments), but each team member has specific tasks. Most of the tasks that relate to the data/records are handled by the software developer and the instrument scientist.

[7] Jordon Johnson (2004), “MOST Archiving Manual,” draft version of 25 May 2004 (unpublished).

The "administrative" activities in the MOST project are minimal. In the data gathering phase of the MOST activities, various files with data are created. The data are then brought together in FITS files that form the basis of further processing. The processing of the data includes reductions and analyses. Finally, the research results are published or presented.

Documentary Context

The activities of the MOST researchers result in various types of files with data from the satellite and from other sources that are detailed below, reductions, and texts with graphs and images for presentations and publications.

There is no formal records management program or written policy related to records or data management in the MOST project. There is one person responsible for the preservation of the satellite transmitted data and gathered from other sources. One of his duties is for instance to back up parts of the data collection on DVDs, formerly CDs. Due to time constraints, the 'records management' work has not been a priority. Experience has resulted in some internal guidelines that are observed in the archiving of the files (e.g., a file naming convention).

There are some internal documents that describe aspects of the management of the records/data. One relates to the metadata that are included in the FITS files, another is an 'archiving manual' that explains the daily, weekly and target (i.e., relating to a specific targeted star) duties of the responsible person.

These internal documents were created by members of the MOST team, and are based on experience and best practice, and not on archival principles. In practice, the duties as described in the 'archiving manual' are not always executed as stated in the manual.

Document creation activities

There are several digital entities created during the scientific process of collecting, reducing and analyzing the data. The first digital entities are the three types of sds files that are transmitted from the satellite. The sds2 files are created immediately after the images are taken by the camera on the satellite and contain all critical data and some supplementary data. The sds1 files are made simultaneously but are compressed versions of the sds2 files. The only purpose of the sds1 files is to fulfill a CSA requirement that it has to be possible to store the data for seven days on the satellite in case there are transmission problems. The sds1 and sds2 files are then split into 100kB sds raw files. Sds raw files are sent to one of the ground stations. On earth, the sds raw files are reconstituted into complete sds1 and sds2 files. The sds2 files are used to create the FITS files.

In addition to these sds files, there are various metadata files, including contextual information about the camera on the satellite, information about how the image files are split,[8] and timing information associated with each of the exposures. However, metadata files do not accompany individual sds files. Orbital information is downloaded from the North American

[8] The camera takes one-mega pixel images. Since the bandwidth for transmission is not large enough, the pictures are cut, and only relevant sections are sent to earth. There is always a primary target (the star that is observed) and some secondary targets (stars that are in the same region as the primary target). Information about this process is captured in a RASTA (Really Annoying System Tool Acronym) file. RASTA is a framework for describing tasks on a computer system. For more information about RASTA, see http://oss.oracle.com/projects/rasta/documentation/.

Aerospace Defense Command (NORAD) Web site.[9] Orbital information is kept in Two Line Element (tle) files.

The sds2 files are converted into FITS files using a C++ program. The FITS files form the basis for all scientific reductions and analyses.

Documents resulting from activities

Scientific analysis is performed on the documents in the FITS file format because these contain the data and related metadata. Later in the research process, analyses and reductions of the data result in other types of records (e.g., light curves or images). These records are used in scientific publications and presentations.

File formats used to create and maintain project documents are listed below in the rough order of generation.

- **.lst** created on ground. Gives the list of rasters for a target and is usually referred to as the raster list. This is usually a single region of the Charge Coupled Device (CCD)[10] within which the sub-rasters (see next entry) are defined.
- **.rst** created on ground. Gives the list of sub-rasters with a raster and is usually referred to as the sub-raster list. Each sub-raster represents some of area of interest on the CCD, whether a secondary stellar target, dark or bias, region, primary fabry target, background fabry image,[11] etc.
- **.str** created from the .lst and .rst information. Usually referred to as the StreamID file. This is used to create a link between a given sds file and its raster and subraster settings. Each sds file contains a value indicating the StreamID file that it is associated with. Each StreamID file also contains the same value, which allows each sds file to be tied to a unique StreamID file. The StreamID file contains the same information as a raster list and subraster list combined.
- **.tle** (two-line-element). A standard format for the definition of a satellite's orbit. These are created by Norad and available online. These are typically updated once a week.
- **.xls** Microsoft Excel file. Pass Report Information files use the Excel format. These files contain any information that might be useful in the analysis of data, such as problems with the satellite or ground stations, status of the download buffers, etc. Pass Report files are created once per day except on weekends when no operator is on site. Operations Schedule: Used to store upcoming pass information for the next ten days and the operator(s) on duty.
- **.csv** created on ground. Timing tick logs use the Comma Separated Values file format to tie a timestamp on the satellite with a well-known time on the ground so that later calibration of the on-board clocks is possible. The csv format is also used for downloaded telemetry information. These files, of which there are thirteen, distinguished by their naming conventions, contain a wide range of telemetry information on all aspects of the satellite. These documents are not considered in this report's data analysis.

---

[9] Established in 1958, NORAD "is a bi-national United States and Canadian organization charged with the missions of aerospace warning and aerospace control for North America." See http://www.norad.mil/.

[10] "A CCD is a sensor for recording images, consisting of an integrated circuit containing an array of linked (coupled) capacitors. Under the control of an external circuit, each capacitor can transfer its electric charge to one or other of its neighbours" (MOST, "Glossary." Available at http://most.oracology.net/?q=glossary/term/10).

[11] "The projected image of a telescope's primary mirror through a Fabry lens" (MOST, "Glossary." Available at http://most.oracology.net/?q=glossary/term/9).

- **.bin** downloaded from the satellite. Usually referred to as “sds raw” files. Each file can contain one or more individual sds files encoded using KISS encoding, which is prevalent in the amateur satellite community. The maximum file size is 100KB and individual sds files may span across two consecutive sds raw files.
- **.dat** an individual sds file. The same file extension is used for both sds1 and sds2. The two file types are distinguished by their naming convention, all file names start with either SDS1 or SDS2, and a flag within the file itself. These files are also KISS encodedand compressed. These files are split into sds raw files for transmission, then reconstituted. Each file contains both image information and telemetry information (temperatures, pointing errors, voltages, time stamps).
- **.fits** (flexible image transport system) is a standard format for astronomical data. Each FITS file typically represents one image or spectrum, containing header information (date, exposure time, telescope, etc) and the image itself. FITS files are created from the sds files, with additional input from the streamID files, timing ticks, TLEs, etc.
- **.xls** format is used for the analysis of the data in a light curve
- **.jpg** image file format is used for creating presentation images.

File organization

The files are stored in a directory-structure organized by primary target (i.e., star and time). This is the way the MOST researchers use the data and records. The data sets that are generated by the satellite camera are preserved and never changed.

Files are copied on DVDs.

Technological Context

Since the MOST project’s software is mostly custom-made, preservation of the software is as important as preservation of the data. Back-ups are made of the various programs that are used in the project. Moreover, if anything is added to one of the programs, the old version of the software is always preserved. In this way, the researchers are always capable of recreating earlier results.

The files with scientific data are always preserved, even if they are corrupt or false. False or corrupt data are filtered out in the data reduction phase and do not affect the final outcomes of the scientific research.

At first, CDs were used as the back-up medium, but now DVDs are used. The earlier data stored on the CDs have not (yet) been migrated to DVDs.

File storage has also changed from a File Allocation Table (FAT) 32 format to the New Technology File System (NTSF) format. The FAT 32 format can only handle approximately 64,000 files in all the subfolders. Because the MOST research generates approximately 250,000 files per target, the FAT 32 format could not support the management, use and preservation of the data. The NTFS format eliminated the 64,000 file restriction.

Because errors were found when writing a large number of files to DVD, the files are first compressed (i.e., zipped) into a single file before being written to DVD.

The researcher who develops the software is responsible for the management and preservation of the data. Another researcher is responsible for controlling the integrity of the data as transmitted by the satellite. Due mostly to technical problems, data may be corrupted in transmission. The daily control of the integrity of the new data ensures that the raw data that are the basis for reductions and analyses are accurate. There are two integrity checks. The first is a computer executed technical analysis, and the second is an intellectual analysis performed by the instrument scientist.

The official set of data is stored on one computer in the MOST office and regularly backed up onto DVDs, with one set maintained in the MOST office, and the other off-site. Another researcher, responsible for the reductions of the raw data, has a second set of the data on his computer but this set is not considered an 'official' set.

D. Narrative Answers to the 23 Core Research Questions

1. What activities of the creator have you investigated?

The creator is the MOST project team. This case study focuses on the scientific data/records that are created during the project. The MOST researchers analyze stars on the basis of new data that are captured by a satellite camera. As such, the activities include capturing, transmission, maintenance, analysis (reduction), and preservation of the data/records.

2. Which of these activities generate the digital entities that are the objects of your case study?

Capture of the data by the satellite camera, processing of these data in the satellite, "packaging" of the data in 100kB files, capture of additional data, creation of FITS files in which relevant data are compiled, and the reductions and analyses of the data.

The MOST researchers consider the original data and the FITS files as the most important entities in the project. The original data files are never deleted. The FITS files are considered as important because they are the first instantiations in which all relevant data are captured.

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

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

4. What form do these digital entities take? (e.g., e-mail, CAD, database)

The digital entities are, depending on the stage in the process, files that contain strings of data, graphs, or reductions of the original data (i.e., data formatted to be meaningful to humans). Digital entities take the form of text or graphic files.

4a. What are the key formal elements, attributes, and behavior (if any) of the digital entities?

The key elements are predominantly textual, but there are graphic elements as well.

4b. What are the digital components of which they consist and their specifications?

The researchers use various types of files detailed above.

4c. What is the relationship between the intellectual aspects and the technical components?

The intellectual aspects of the MOST research project refer to the data that are captured (i.e., document content). The technological components refer to the technology used to capture and process the data. The MOST researchers created custom made software to process the data.

To access the FITS files, various viewer applications are available. The viewer does not affect the document content, but the interfaces differ.

Since the beginning of the project multiple versions of programs have been used in the process of reducing the data. Results vary depending on the program used. In the end, the MOST researchers will make a decision about the official reduction and the official version of the program used for that reduction.

4d. How are the digital entities identified (e.g., is there a [persistent] unique identifier)?

Digital entities are uniquely identified using a file naming convention relating to: (1) the primary target (i.e., star) and (2) date. In addition to this, the metadata provide additional unique identifiers.

4e. In the organization of the digital entities, what kind of aggregation levels exist, if any?

Before transmission from the satellite to earth, captured data are split into 100kB files to accommodate the limited transmission bandwidth. After receipt, these files are used to reconstitute the original files.

In the management of the data/records, the entities are organized into files organized by primary target (i.e., star). Items within each file are in chronological order.

4f. What determines the way in which the digital entities are organized?

Technological constraints determine the organization of the digital entities for transmission. The way the MOST researchers use the records determines how the entities are organized and used because analyses are executed on the basis of a star through time.

5. How are those digital entities created?

The first digital entities are created at the moment that the satellite camera takes an image. Subsequent entities are created in the satellite prior to their transmission to earth. These files are created automatically by the various software programs, without human involvement.

The entities comprising the reductions of the data are created by the instrument scientist using custom-made software.

5a. What is the nature of the system(s) with which they are created? (e.g., functionality, software, hardware, peripherals etc.)

The operating system used in the MOST research project is Windows. The created files are managed in Windows Explorer. Other software that is used in the project is custom-made.

5b. Does the system manage the complete range of digital entities created in the identified activity or activities for the organization (or part of it) in which they operate?

There is no overall system that manages all data/records.

6. From what precise process(es) or procedure(s), or part thereof, do the digital entities result?

Most digital entities are the result of automated processes of data gathering, packaging, and reduction.

7. To what other digital or non-digital entities are they connected in either a conceptual or a technical way? Is such connection documented or captured?

Conceptually, the raw data files that the satellite transmits are ordered similarly to how they are maintained (i.e., by target and date). Other data files are related to the raw data files by date. These various data files are technically connected within the FITS files by the program that creates them. The program indicates where the different parts of information have to be extracted so that that information can be put in the FITS file.

Naming and directory conventions indicate the conceptual connections between the entities.

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

The digital entities are given unique names, 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.

9. Are those processes and procedures documented? How? In what form?

A MOST researcher has made an internal document ('MOST archiving manual,' dd. August 25, 2004) that outlines the procedure for creating, naming, and storing the digital entities. In the manual, a difference is made between daily, weekly and target[12] duties.

10. What measures does the creator take to ensure the quality, reliability and authenticity of the digital entities and their documentation?

Throughout the process of transmission of data, applications perform checksums to ensure that the data remains unchanged. This is a technical issue and has to be done for all raw data files to ensure that they are reliable. Besides checksums, the instrument scientist looks at individual files or sequences of files. From the moment these checks are completed, the MOST researchers consider the data to be reliable.

However, the researchers acknowledge that some raw data files contain corrupt data. These corrupt or false data are eliminated in the process of data reduction. The fact that these data are false or corrupt has no impact on the outcome of the reductions, which is one of the end results of the scientific research. The inclusion of false or corrupt data in the reduction calculations does not affect the reliability of the outcome.

Accuracy and reliability are important issues for the FITS files as well. In the process of creating FITS files, there are various checks to assure that the information that is input is

[12] A target star can be the same for a while (e.g., a month).

good. If errors occur, the researchers typically will examine the problem and recreate the FITS files. For instance, a magnetic field value is added in the FITS files. This value is based on the position of the satellite at a particular time and is calculated on the basis of a model. The model that was used went up on 1 January 2005 and initially showed all magnetic field values were 0. After this problem was identified, the FITS files were recreated on the basis of a corrected magnetic field model. A researcher noted that values may need to be added or corrected in the FITS file.

11. Does the creator think that the authenticity of his digital entities is assured, and if so, why?

There is no reason for the creator to assume that the digital entities are not authentic.

12. How does the creator use the digital entities under examination?

The digital entities are used as basis for data reductions and scientific analysis. Data reductions are used as basis for scientific presentations and publications.

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

In general, digital entities are not changed: sds files are never changed or deleted. FITS files are sometimes recreated. One of the descriptive fields of the FITS files indicates which version/program was used to create the file. FITS files might be recreated if they contain errors. Digital entities that are the result of reductions are not changed either. Every new reduction ends up as a new digital entity. If the results of that reduction are more satisfactory to the researchers than previous reductions, then the latter can be disposed of.

There is no version control on recreated FITS files. By recreating FITS files, the old one is superseded. Moreover, it is always possible to recreate the superseded FITS files using with the earlier, preserved versions of the application.

14. 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?

There are few external users are using digital entities of the MOST project.[13] These users are researchers of other scientific institutions who are receive data directly from the MOST researchers.

As data are made publicly available at the CADC Web site, a larger community of external users can access the FITS files.

15. Are there specific job competencies (or responsibilities) with respect to the creation, maintenance, and/or use of the digital entities? If yes, what are they?

The researcher responsible for the creation and maintenance of the raw data files and FITS files is both an astronomer and software developer. He is a specialist in designing software that creates and captures the data that is necessary for the scientific process.

The scientist who creates the reductions is also an astronomer. However, there are no formal or written job competencies.

[13] Researchers at the partner institutions, UTIAS and the University of Vienna, are not considered as external users. External users are scientists of the European COROT project, for example.

16. Are the access rights (to objects and/or systems) connected to the job competence of the responsible person? If yes, what are they?

There are physical access rights: there is only one computer with the official records in the MOST office; however, all MOST researchers have access to it.

17. Among its digital entities, which ones does the creator consider to be records and why?

The researchers use the word 'record' in another sense than archivists.[14] The sds and FITS files are considered the most important files that are created. The raw data (i.e., sds files) contain original data, while the FITS files are the first instantiation in which the data are represented in a comprehensible way. The raw data files are never deleted. Any other files may be deleted because it is possible to recreate them.

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

There is no recordkeeping system in place. All sds files are preserved without changes or additions. FITS files are preserved as well, but may be recreated. Other digital entities (e.g., reductions) can be removed without any audit trail. If results based on a specific reduction are published, that reduction is preserved.

18a. Do the recordkeeping system(s) (or processes) routinely capture all digital entities within the scope of the activity it covers?

Only the FITS and sds files are routinely captured and backed up. From the moment any other entity is superseded (e.g., a more satisfactory reduction is created), they may be removed.

18b. From what applications do the recordkeeping system(s) inherit or capture the digital entities and the related metadata (e.g., email, tracking systems, workflow systems, office systems, databases, etc.)?

Other than Microsoft Windows Explorer, there is no formal capture system in place.

18c. Are the digital entities organized in a way that reflects the creation processes? What is the schema, if any, for organizing the digital entities?

Yes, the digital entities are organized by target star and date and, thus, reflect the creation process.

18d. Does the recordkeeping system provide ready access to all relevant digital entities and related metadata?

It is possible to access all digital entities via Windows Explorer.

18e. Does the recordkeeping system document all actions/transactions that take place in the system re: the digital entities? If so, what are the metadata captured?

There is no audit trail.

[14] Record means a field value.

19. How does the creator maintain its digital entities through technological change?

For the creation of the FITS files and reductions, the old versions of the software applications are preserved.

19a. What preservation strategies and/or methods are implemented and how?

All records/data are backed up initially onto CDs and currently onto DVDs. There is a backup procedure set out in the MOST Archiving Manual. Because the FITS files and reductions are created using specific, custom-built software, the MOST researchers also backup all versions of the software so that each reduction can be redone in the same software environment.

19b. Are these strategies or methods determined by the type of digital entities (in a technical sense) or by other criteria? If the latter, what criteria?

The MOST researchers backup the data for reasons of security.

20. To what extent do policies, procedures, and standards currently control records creation, maintenance, preservation and use in the context of the creator's activity? Do these policies, procedures, and standards need to be modified or augmented?

There are hardly any procedures beyond the MOST Archiving Manual due to the organizational culture, the size of the research team and resources available.

21. What legal, moral (e.g., control over artistic expression) or ethical obligations, concerns or issues exist regarding the creation, maintenance, preservation and use of the records in the context of the creator's activity?

There are no legal, moral or ethical issues.

22. What descriptive or other metadata schema or standards are currently being used in the creation, maintenance, use and preservation of the recordkeeping system or environment being studied?

The metadata schema used was created by the MOST researchers and is specific for the data/files that are created in the MOST project. The metadata refer to information such as orbital parameters, observational parameters, telemetry information and target image information. Also, the FITS file format requires some metadata/descriptive fields.

In general, no metadata standards are used; the MOST researchers created their own scheme of important descriptive fields.

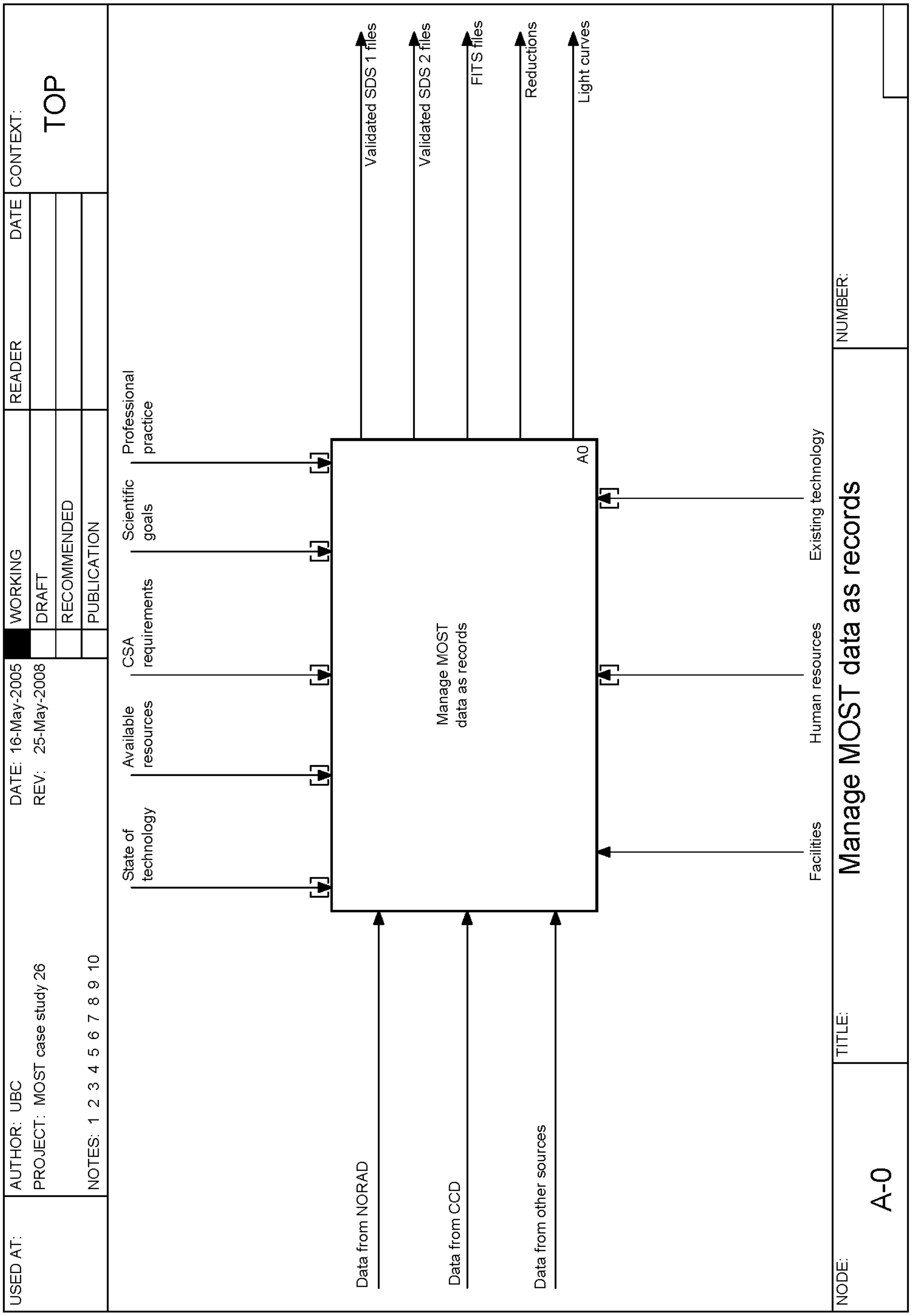
23. What is the source of these descriptive or other metadata schema or standards (institutional convention, professional body, international standard, individual practice, etc.?)

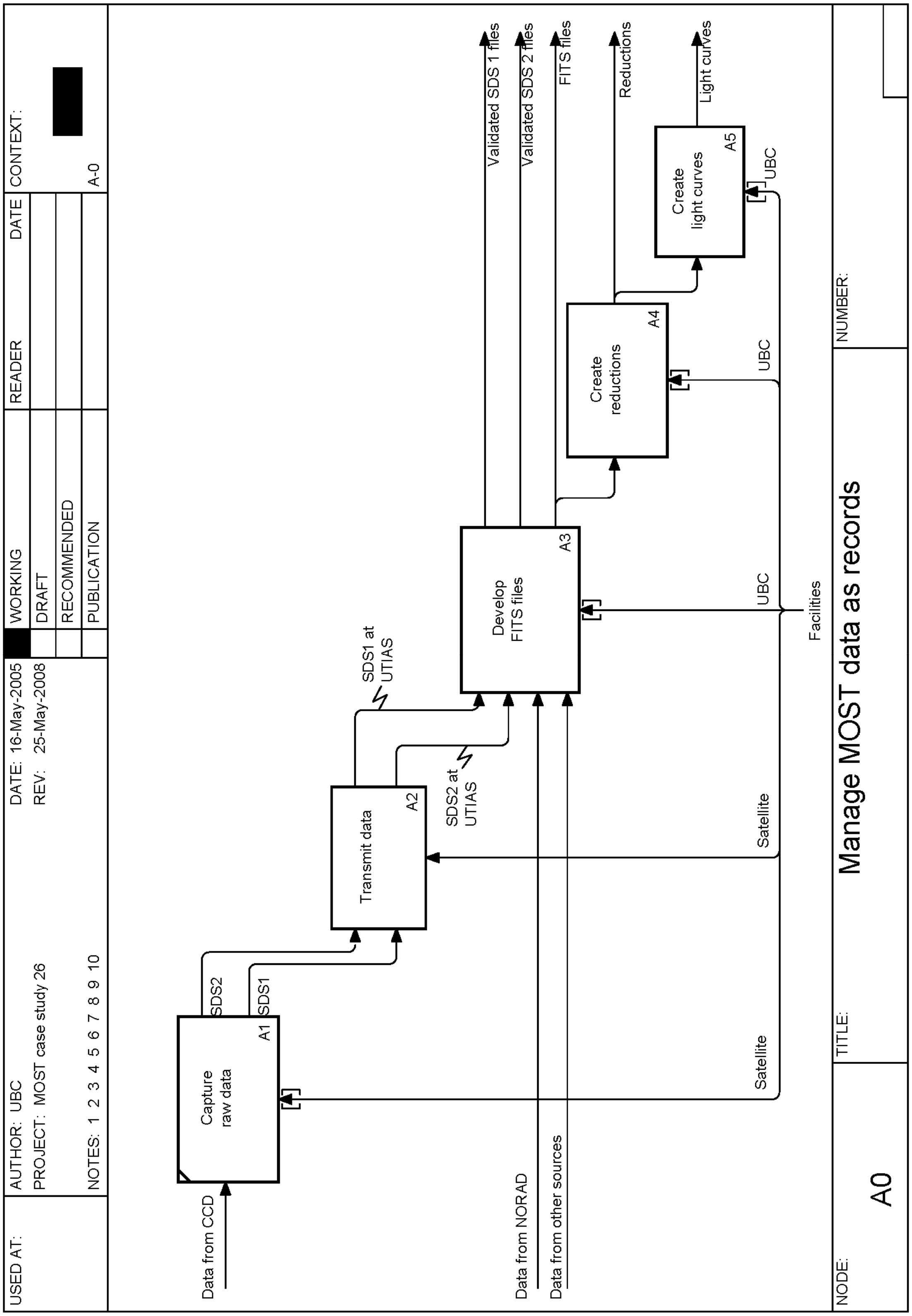
The metadata that are used for the various files are based on experience and best practice in the astronomical research community, and on the foreseeable use of the records. There is an internal MOST document that describes the descriptive fields of the FITS files.

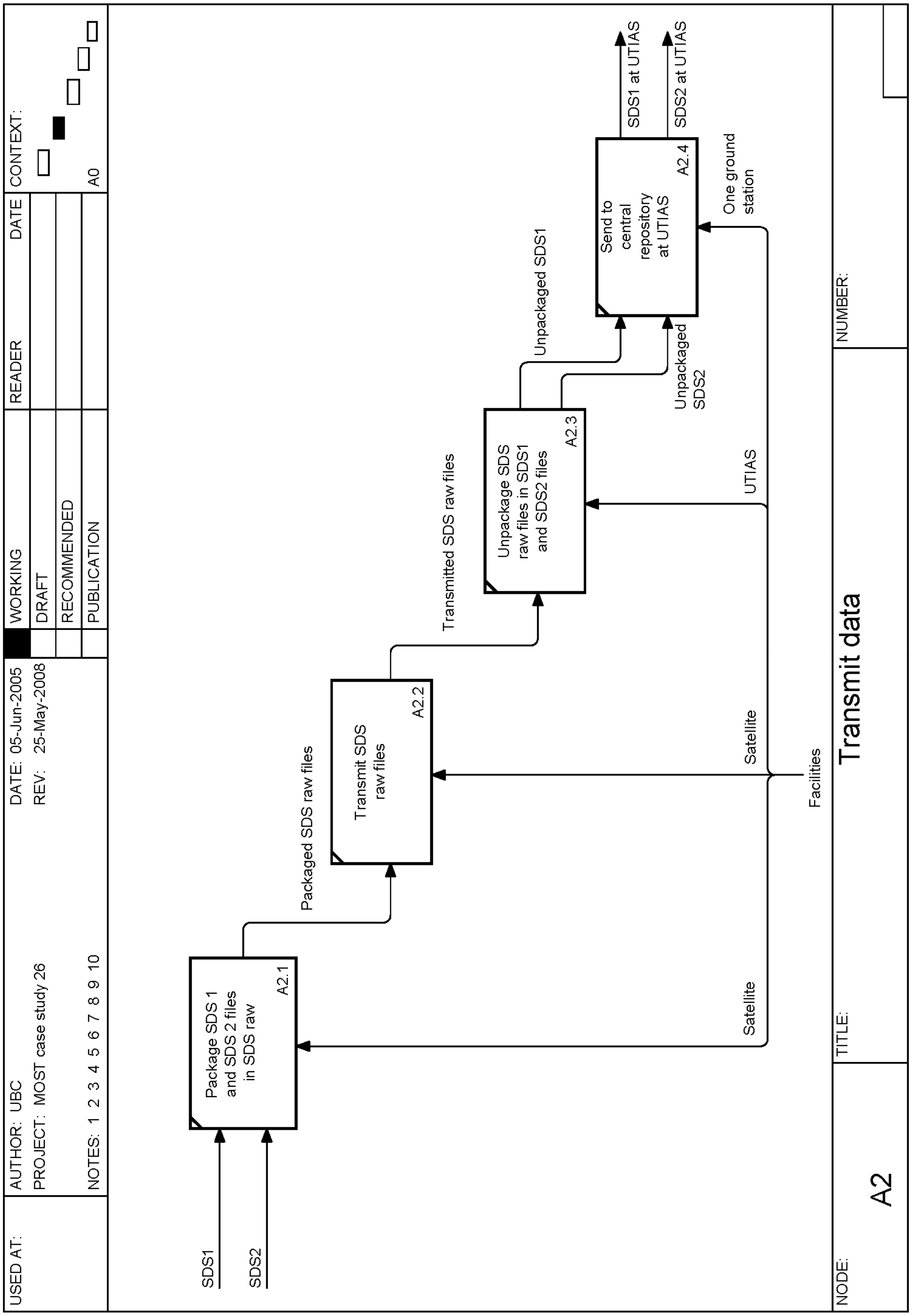
E. List of terms and abbreviations

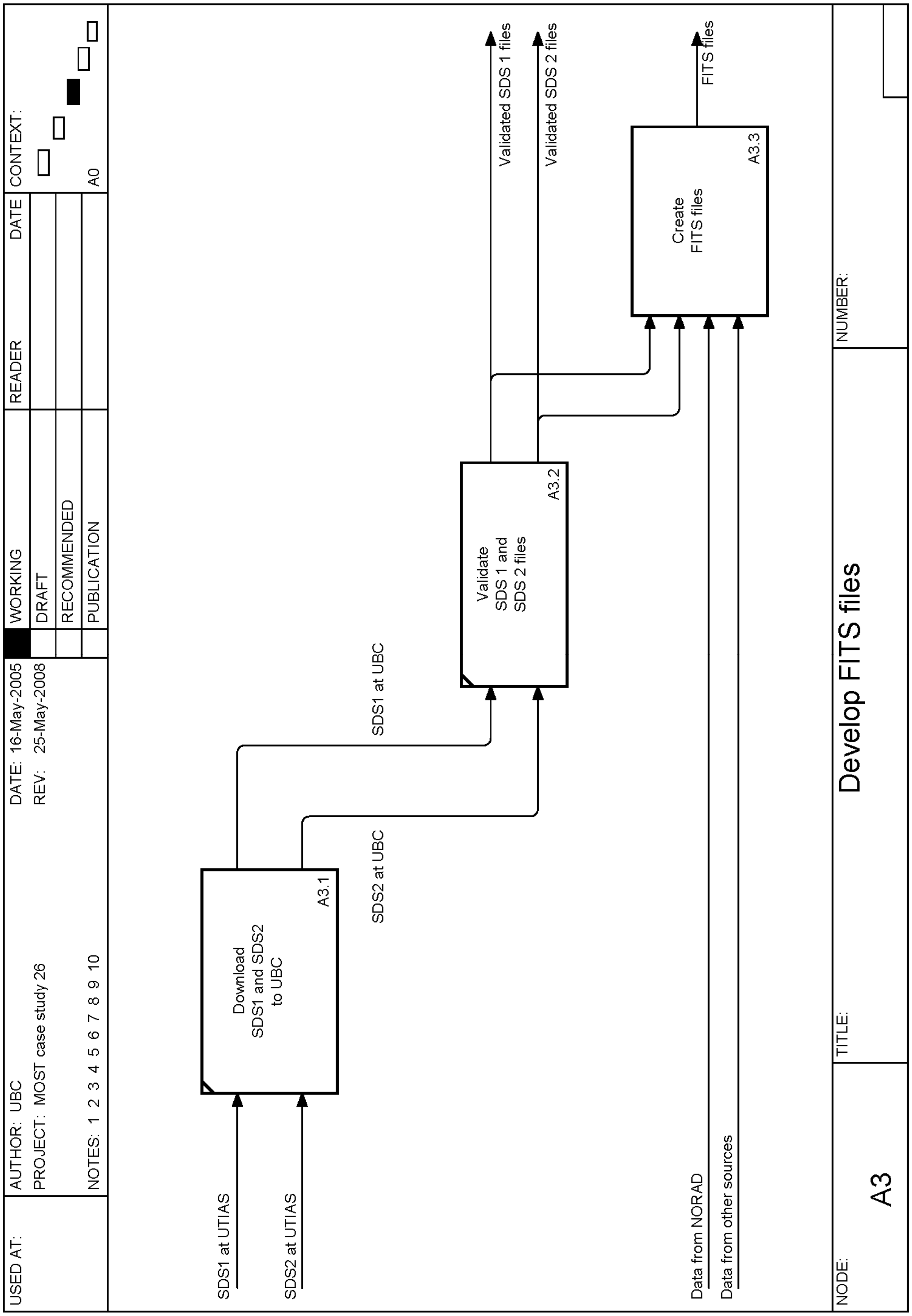
CADC	Canadian Astronomy Data Centre
CCD	Charge Coupled Device. A sensor for recording images, consisting of an integrated circuit containing an array of linked (coupled) capacitors
CSA	Canadian Space Agency
csv	Comma Separated Values
Fabry image	The projected image of a telescope's primary mirror through a Fabry lens.
Fabry lens	A lens that forms an image of the telescope’s primary mirror onto the science CCD. A mask placed in front of it, isolates the object whose light intensity is to be measured. The use of such a lens was first proposed by the French physicist Charles Fabry.
FAT	File Allocation Table
file	Used as computer file in this report.
FITS	Flexible Image Transport System
MOST	Microvariability & Oscillations of STars
NSERC	Natural Sciences and Engineering Research Council of Canada
NORAD	North American Aerospace Defense Command
NTFS	New Technology File System
RASTA	Really Annoying System Tool Acronym. A framework for describing tasks on a computer.
reduction	The transformation of data from a “raw” form to some useable form.
sds	science data stream
Tle	Two Line Element
UTIAS	University of Toronto Institute for Aerospace Studies

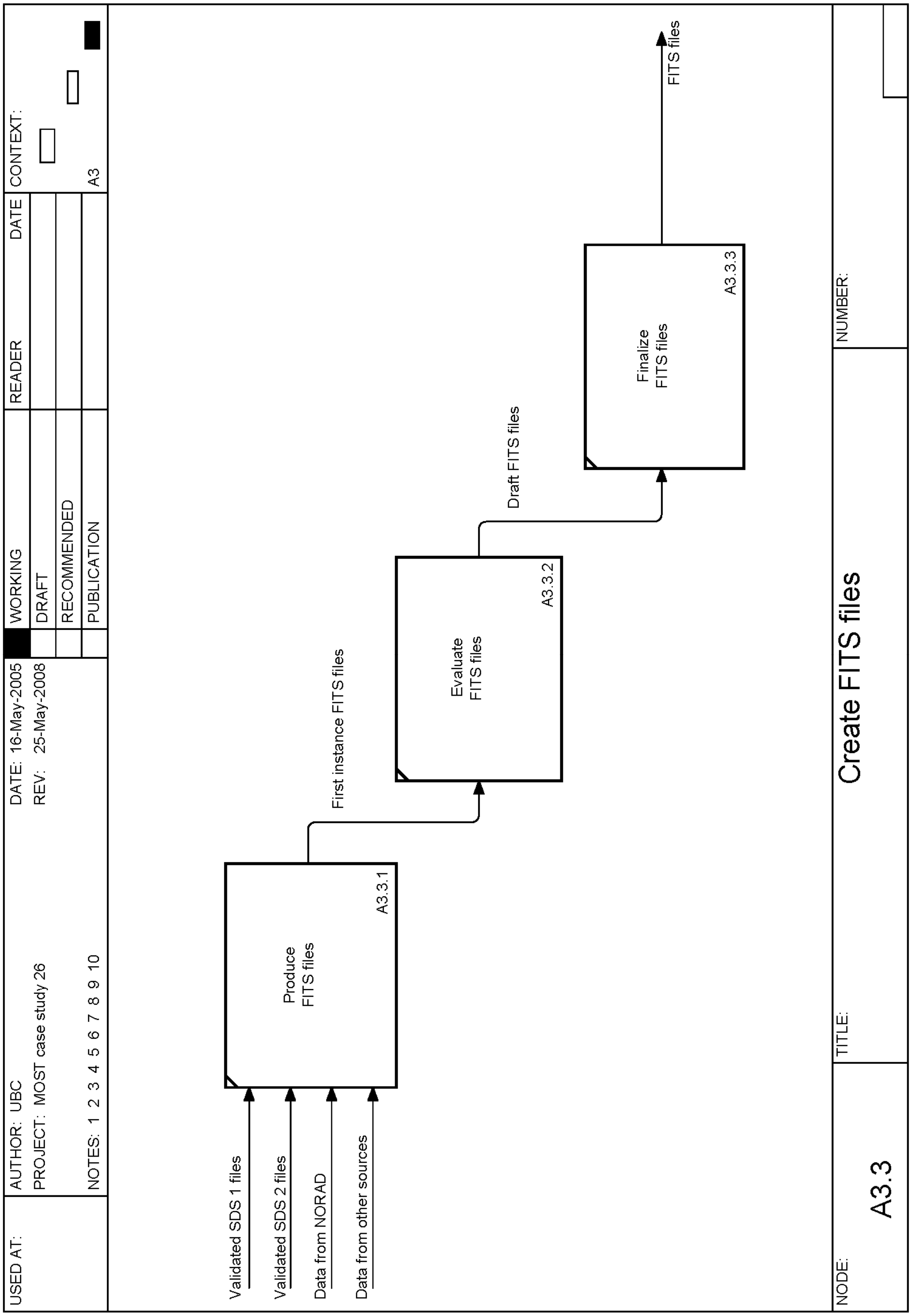
F. IDEF0 Activity Model

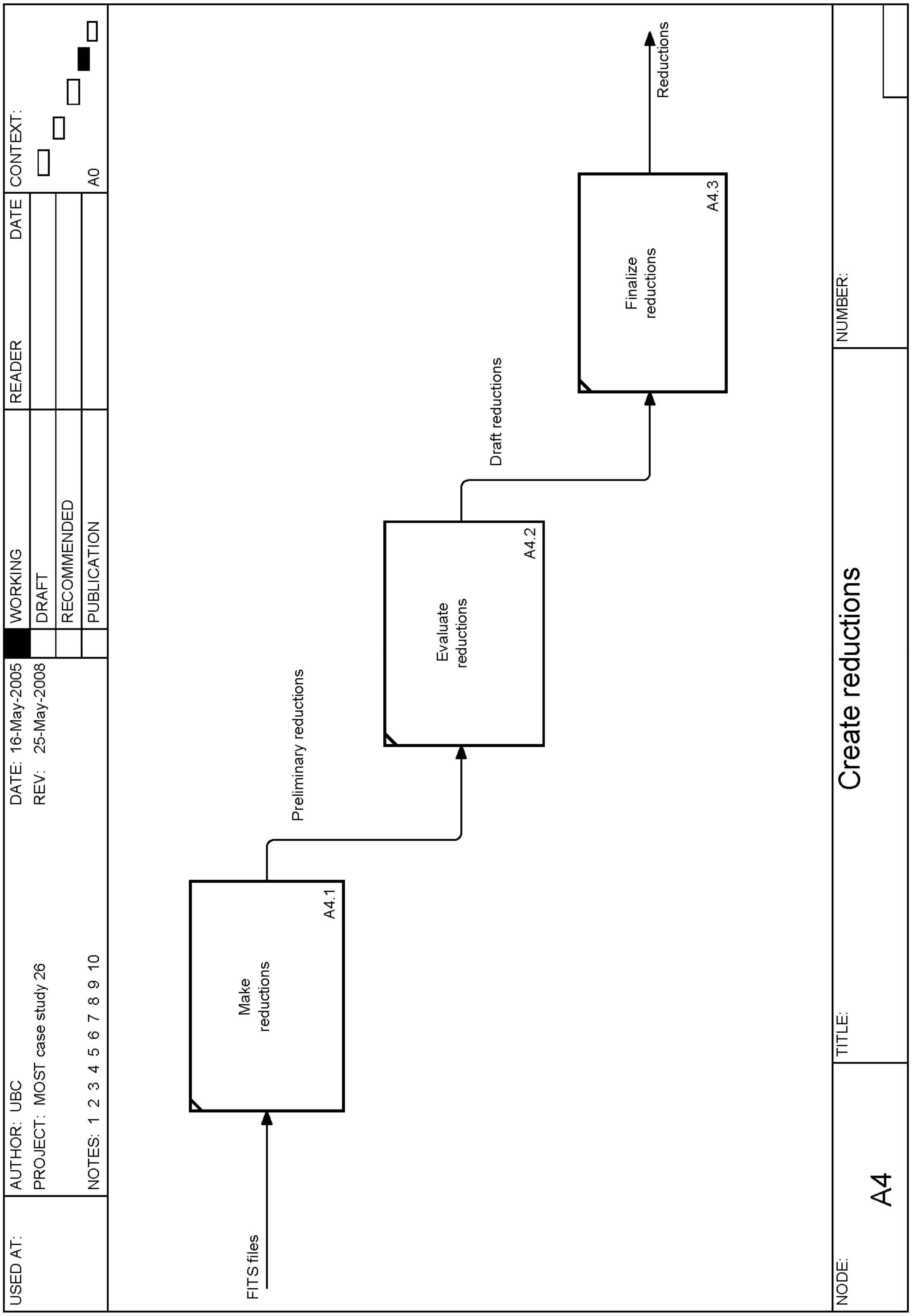


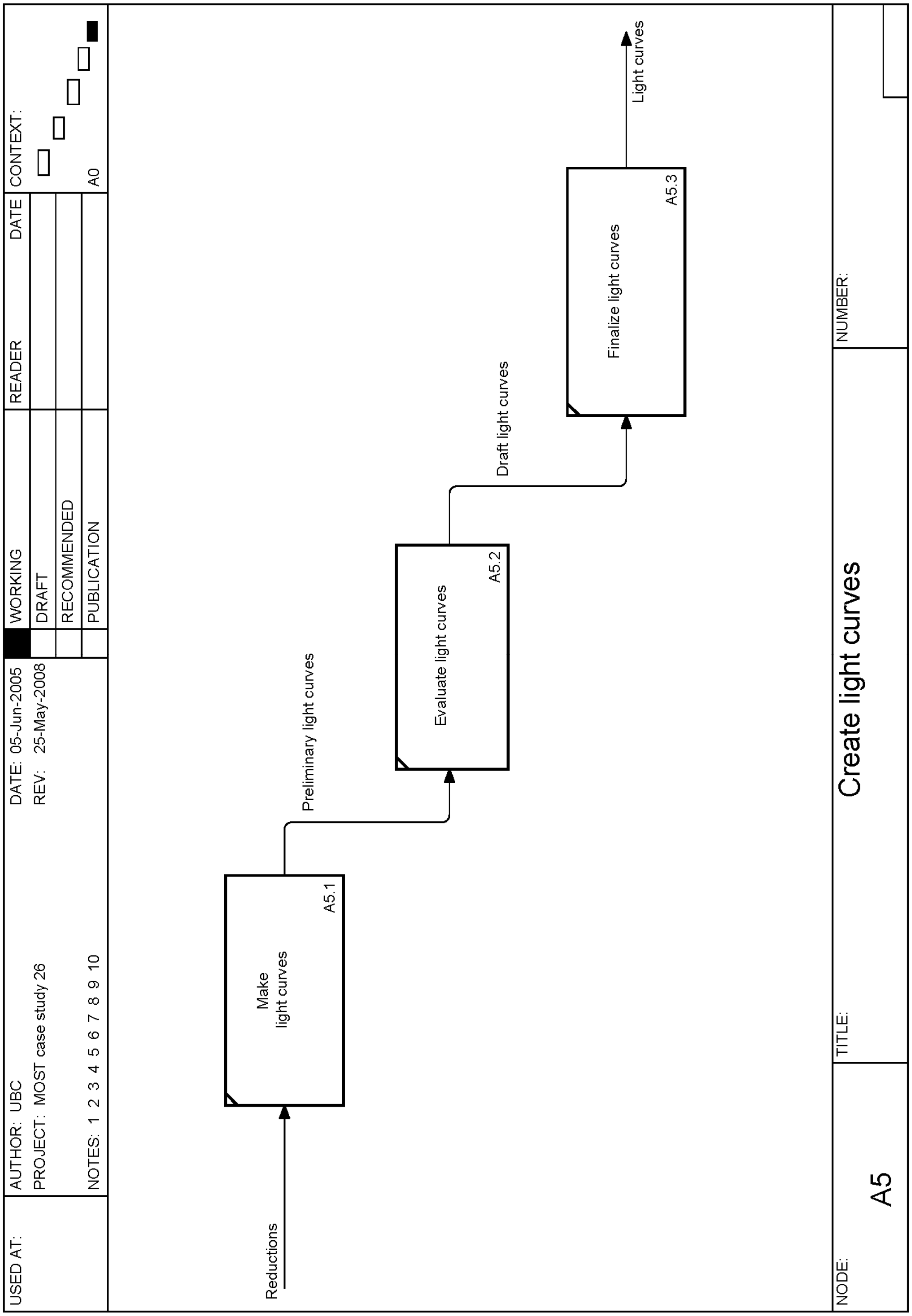

USED AT:
AUTHOR: UBC
PROJECT: MOST case study 26
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 16-May-2005
REV: 25-May-2008
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A-0
Data from CCD
Capture raw data
A1
SDS2
SDS1
Transmit data
A2
SDS1 at UTIAS
SDS2 at UTIAS
Develop FITS files
A3
Create reductions
A4
Create light curves
A5
Data from NORAD
Data from other sources
Satellite
Satellite
UBC
UBC
UBC
Facilities
Validated SDS 1 files
Validated SDS 2 files
FITS files
Reductions
Light curves
NODE:
A0
TITLE:
Manage MOST data as records
NUMBER:

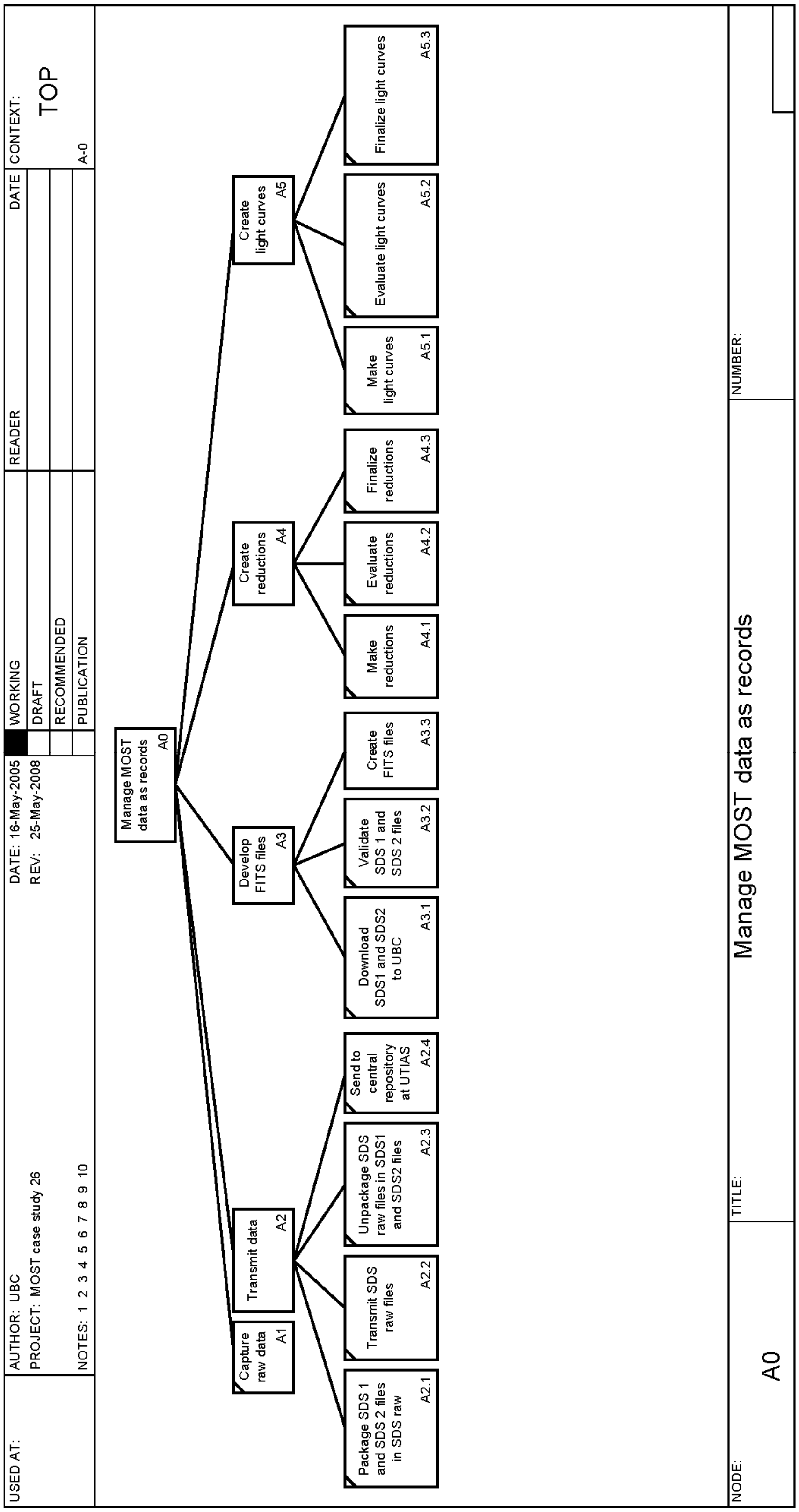

USED AT:
AUTHOR: UBC
PROJECT: MOST case study 26
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 05-Jun-2005
REV: 25-May-2008
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A0
SDS1
SDS2
Package SDS 1 and SDS 2 files in SDS raw
A2.1
Packaged SDS raw files
Transmit SDS raw files
A2.2
Transmitted SDS raw files
Unpackage SDS raw files in SDS1 and SDS2 files
A2.3
Unpackaged SDS1
Unpackaged SDS2
Send to central repository at UTIAS
A2.4
SDS1 at UTIAS
SDS2 at UTIAS
Satellite
Satellite
UTIAS
One ground station
Facilities
NODE:
A2
TITLE:
Transmit data
NUMBER:


USED AT:
AUTHOR: UBC
PROJECT: MOST case study 26
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 16-May-2005
REV: 25-May-2008
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A0
SDS1 at UTIAS
SDS2 at UTIAS
Download SDS1 and SDS2 to UBC
A3.1
SDS1 at UBC
SDS2 at UBC
Validate SDS 1 and SDS 2 files
A3.2
Validated SDS 1 files
Validated SDS 2 files
Create FITS files
A3.3
FITS files
Data from NORAD
Data from other sources
NODE:
A3
TITLE:
Develop FITS files
NUMBER:


USED AT:
AUTHOR: UBC
PROJECT: MOST case study 26
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 16-May-2005
REV: 25-May-2008
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A3
Validated SDS 1 files
Validated SDS 2 files
Data from NORAD
Data from other sources
Produce
FITS files
A3.3.1
First instance FITS files
Evaluate
FITS files
A3.3.2
Draft FITS files
Finalize
FITS files
A3.3.3
FITS files
NODE:
A3.3
TITLE:
Create FITS files
NUMBER:


USED AT:
AUTHOR: UBC
PROJECT: MOST case study 26
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 16-May-2005
REV: 25-May-2008
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A0
FITS files
Make reductions
A4.1
Preliminary reductions
Evaluate reductions
A4.2
Draft reductions
Finalize reductions
A4.3
Reductions
NODE:
A4
TITLE:
Create reductions
NUMBER:


USED AT:
AUTHOR: UBC
PROJECT: MOST case study 26
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 05-Jun-2005
REV: 25-May-2008
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A0
Reductions
Make
light curves
A5.1
Preliminary light curves
Evaluate light curves
A5.2
Draft light curves
Finalize light curves
A5.3
Light curves
NODE:
A5
TITLE:
Create light curves
NUMBER:


USED AT:
AUTHOR: UBC
PROJECT: MOST case study 26
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 16-May-2005
REV: 25-May-2008
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
TOP
A-0
Manage MOST data as records
A0
Capture raw data
A1
Transmit data
A2
Develop FITS files
A3
Create reductions
A4
Create light curves
A5
Package SDS 1 and SDS 2 files in SDS raw
A2.1
Transmit SDS raw files
A2.2
Unpackage SDS raw files in SDS1 and SDS2 files
A2.3
Send to central repository at UTIAS
A2.4
Download SDS1 and SDS2 to UBC
A3.1
Validate SDS 1 and SDS 2 files
A3.2
Create FITS files
A3.3
Make reductions
A4.1
Evaluate reductions
A4.2
Finalize reductions
A4.3
Make light curves
A5.1
Evaluate light curves
A5.2
Finalize light curves
A5.3
NODE:
A0
TITLE:
Manage MOST data as records
NUMBER:

CS26 – MOST Satellite Mission - Preservation of Space Telescope Data, IDEF0 Model Activity Definitions			
Activity Name	**Activity No.**	**Activity Definition**	**Activity Note**
Manage MOST data as records	A0	To capture raw data, transmit data, create FITS files, reductions and light curves.	
Capture raw data	A1	To transform raw data from CCD into SDS1 and SDS2 files.	
Transmit data	A2	To package SDS1 and SDS2 files in SDS raw, transmit SDS raw files, unpackage SDS raw files in SDS1 and SDS2 files, and send to central repository at UTIAS.	
Package SDS 1 and SDS 2 files in SDS raw	A2.1	To combine some SDS1 and SDS2 files in a SDS raw file.	
Transmit SDS raw files	A2.2	To send the packaged SDS raw files to one of the three ground stations according to the vicinity of the satellite.	Three ground stations: Vienna, Vancouver and Toronto.
Unpackage SDS raw files in SDS1 and SDS2 files	A2.3	To undo the packaging of the SDS raw files so that they can be used for processing.	
Send to central repository at UTIAS	A2.4	To transmit the unpackaged SDS1 and SDS2 files from one of the three ground stations to the central data repository in Toronto.	UTIAS: University of Toronto Institute for Aerospace Studies.
Develop FITS files	A3	To download SDS1 and SDS2 to UBC, validate SDS1 and SDS2 files and create FITS files.	
Download SDS1 and SDS2 to UBC	A3.1	To transmit the SDS files from UTIAS in Toronto to the MOST project team in Vancouver.	
Validate SDS 1 and SDS 2 files	A3.2	To check the files and to ensure that the data are valid.	
Create FITS files	A3.3	To develop FITS files.	
Produce FITS files	A3.3.1	To bring together various data that are put in the FITS file format so that the astronomers can use this new format as the basis of their research activities.	FITS: Flexible Image Transfer System.

CS26 – MOST Satellite Mission - Preservation of Space Telescope Data, IDEF0 Model Activity Definitions			
Activity Name	**Activity No.**	**Activity Definition**	**Activity Note**
Evaluate FITS files	A3.3.2	To control whether the FITS files are logical, and whether they contain errors.	
Finalize FITS files	A3.3.3	If errors occurred in the process of composing the FITS files, or if the MOST researchers decide to add values to the FITS files, to change and add data to the FITS files, or to confirm the draft version.	
Create reductions	A4	To make, evaluate and finalize reduction.	
Make reductions	A4.1	To produce, on the basis of the FITS files and of an astronomical model, a product that can be used to analyze stars in a period of time.	
Evaluate reductions	A4.2	To control whether the reduction is logical, whether it contains errors, and whether it is the most appropriate method to analyze the brightness of the star.	
Finalize reductions	A4.3	To confirm the draft version or to edit it as needed.	
Create light curves	A5	To make, evaluate and finalize light curves.	
Make light curves	A5.1	To produce a graph on the basis of the reductions that shows the variations of the brightness of the star in a period of time.	
Evaluate light curves	A5.2	To control whether the light curve is logical, and whether it contains errors.	
Finalize light curves	A5.3	To confirm the draft version or to edit it as needed.	

CS26 – MOST Satellite Mission - Preservation of Space Telescope Data, IDEF0 Model Arrow Definitions		
Arrow Name	**Arrow Definition**	**Arrow Note**
Available resources	Human, financial and time resources available to the MOST research project.	
CSA requirements	Legal and other (e.g., scientific and administrative) requirements from the funding agency.	CSA: Canadian Space Agency
Data from CCD	Raw data that is captured by an exposure of the camera on the MOST satellite.	CCD: Charge Coupled Device
Data from NORAD	TLE data that is captured from the Web site of the NORAD and put in the FITS files.	TLE: Two Line Element NORAD: North American Aerospace Defense Command FITS: Flexible Image Transport System
Data from other sources	Data that are put in the FITS files that are neither originating from the CCD or from NORAD.	I.e. timing ticks, RASTA files and streamID files (generated internally in project).
Draft FITS files	Version of the FITS files after evaluation.	
Draft light curves	Evaluated light curves before approval by the research committee.	
Draft reductions	Evaluated reductions before approval by the research committee.	
Existing technology	Technology available to the MOST research project at any given time.	
Facilities	Resources available to the MOST research project in terms of building space and location.	
First instance FITS files	Version of the FITS files before evaluation.	
FITS files	Definitive version of the file format that is used in the astronomical community as the basis for further research and analysis.	FITS: Flexible Image Transfer System
Human resources	Amount of persons and person-time devoted to managing the MOST data/records.	
Light curves	Graphs in which brightness of a star is shown over a period of time, which are approved by the research committee.	

CS26 – MOST Satellite Mission - Preservation of Space Telescope Data, IDEF0 Model Arrow Definitions		
Arrow Name	**Arrow Definition**	**Arrow Note**
One ground station	One of the three stations that are connected to the satellite and are equipped to receive data.	i.e., Vienna, Vancouver and Toronto.
Packaged SDS raw files	SDS raw files to be sent to one of the ground stations.	
Preliminary light curves	Light curves that are created, but not yet evaluated.	
Preliminary reductions	Reductions that are created, but not yet evaluated.	
Professional practice	Best practice as used by the astronomy community.	Includes manuals that the MOST researchers created on the basis of their experience and the field's best practice.
Reductions	Files that result from the application of the "reduction", the scientific method used to analyze the brightness of a star over time, and approved by the research committee.	There are different ways of reducing.
Satellite	MOST instrument with camera and computer in space.	
Scientific goals	The final outcomes that the MOST researchers want to achieve through the MOST project.	
SDS1	Security file format in which the original raw data of the CCD are captured.	
SDS1 at UBC	SDS1 files received at UBC.	
SDS1 at UTIAS	SDS1 files that are received by UTIAS.	
SDS2	File format in which the original raw data of the CCD are captured.	SDS2 contains more information than SDS1. SDS2 is the official set of data
SDS2 at UBC	SDS2 files received at UBC.	
SDS2 at UTIAS	SDS2 files that are received by UTIAS.	
State of technology	Technology currently available to the MOST research project.	
Transmitted SDS raw files	SDS raw files that are transmitted from the satellite to one of the ground stations.	
UBC	MOST office in the Astronomy Department on the University of British Columbia Campus in Vancouver, B.C.	

CS26 – MOST Satellite Mission - Preservation of Space Telescope Data, IDEF0 Model Arrow Definitions		
Arrow Name	**Arrow Definition**	**Arrow Note**
Unpackaged SDS1	SDS1 file that is taken apart from the SDS raw in one of the ground stations.	
Unpackaged SDS2	SDS2 file that is taken apart from the SDS raw in one of the ground stations.	
UTIAS	University of Toronto Institute for Aerospace Studies.	
Validated SDS 1 files	SDS1 files after they have gone through the process of validation.	
Validated SDS 2 files	SDS2 files after they have gone through the process of validation.	