InterPARES 2

CS 26

MOST Satellite Mission
Preservation of Space Telescope Data

2005-11-17

Overview

- The Project
  - Introduction
  - Purpose
  - Facilities
  - Data
    - Collection
    - Analysis
    - Keeping
    - Access
- The Case Study

Introduction

- **MOST** (**M**icrovariability & **O**scillations of **ST**ars)
- Funded by CSA (Canadian Space Agency)
- Canada's first space telescope
- Unprecedented precision and long time coverage

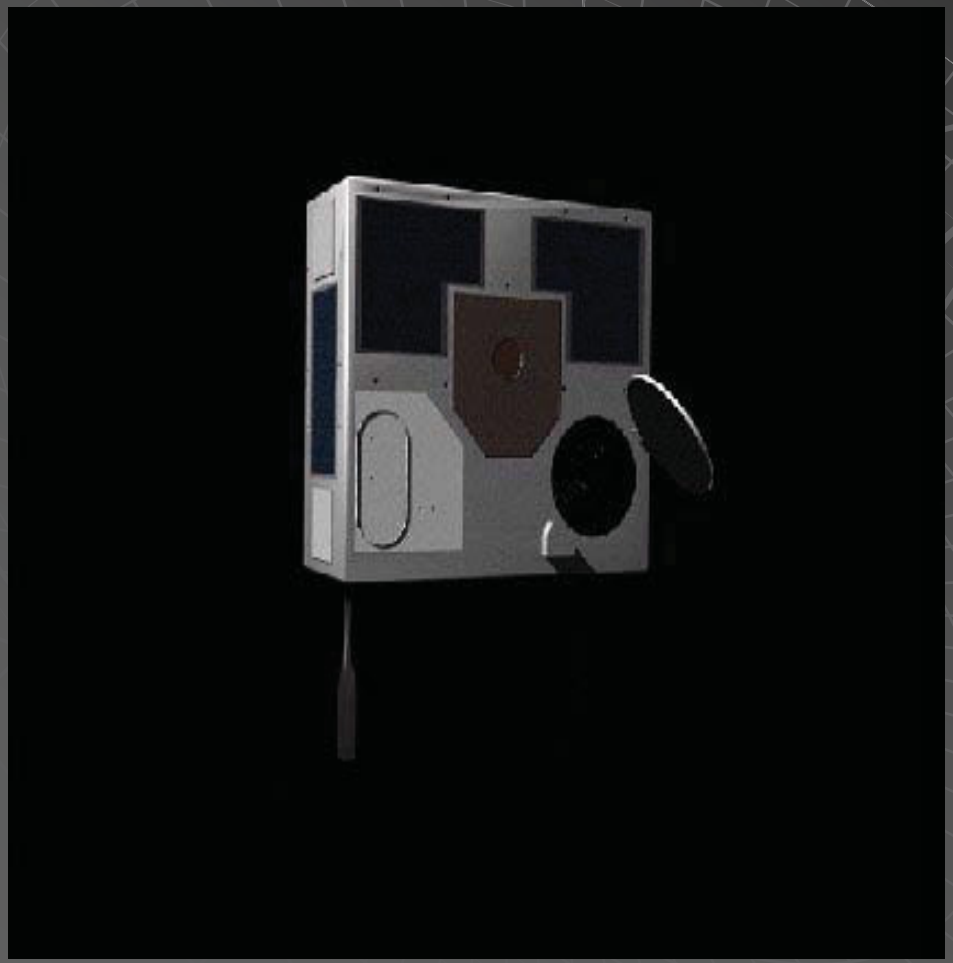
Introduction

- Main partners
  - UBC Physics & Astronomy
    - primary data analysis center
  - UTIAS Space Flight Lab (University of Toronto Institute for Aerospace Studies)
  - Dynacon Enterprises
- Other partners
  - SUMUS (Sumus Technology Limited)
  - Carnegie Observatories
  - Institute for Astronomy, University of Vienna
  - CADC (Canadian Astronomy Data Centre)

Purposes

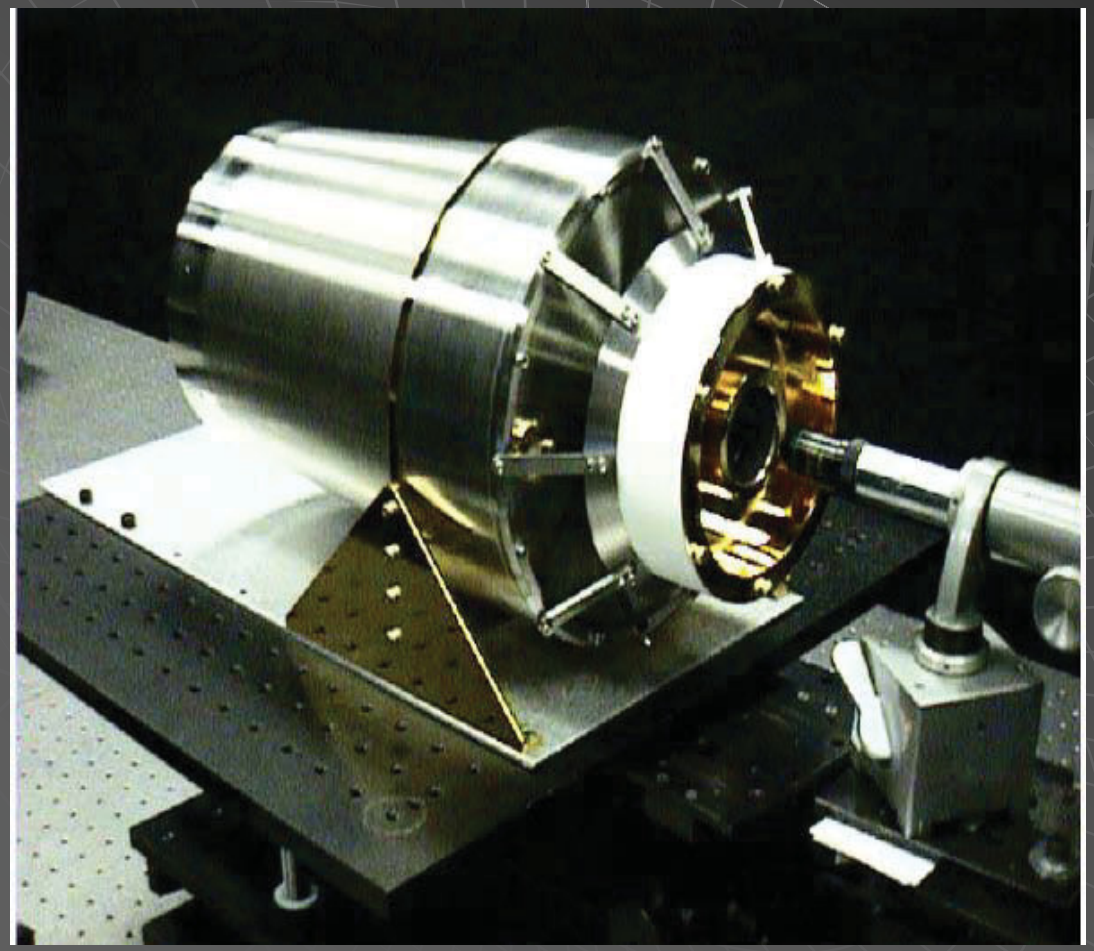
- Monitor & capture variations in the brightness of stars
- Understand the Sun in the context of other stars
- Study the structure and evolution of stars

The Satellite



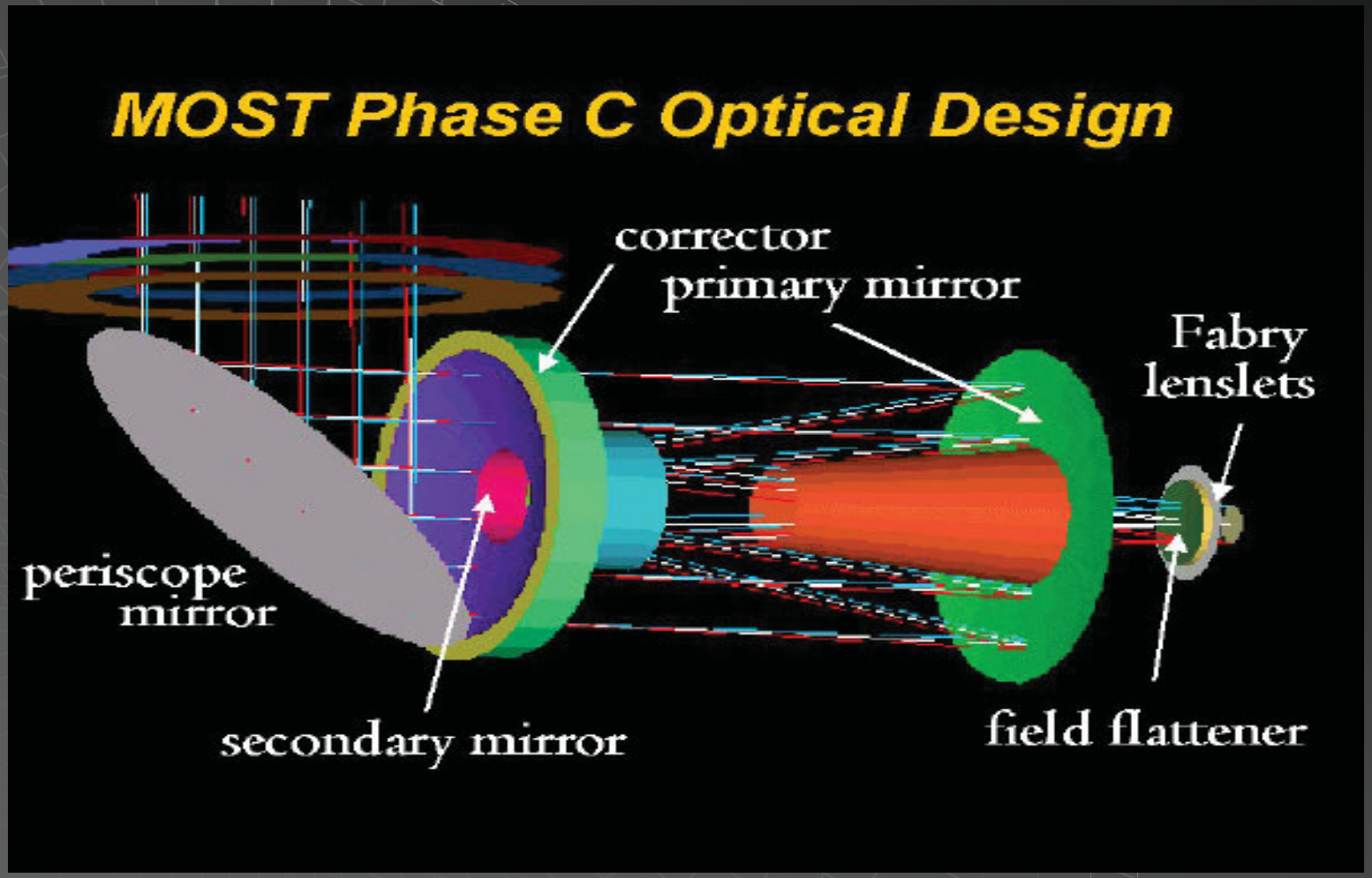
A suitcase-sized
(65cm x 65cm x 30cm, 60 kg)
micro-satellite housing various devices needed for the project

The Telescope



A small telescope (15 cm aperture) produced by the new Canadian attitude control technology, measuring tiny light variations that are undetectable from the Earth

The Optical Design



The Launch

- Russian Rockets
- June 2003

Data Collection

- Series of uninterrupted measurements of star fields lasting up to two months
- Result in up to about 500,000 individual files for a single target
- Each file contains a set of counted charges recorded by light-sensitive pixels on the CCD (Charge Coupled Device)
- Metadata of the file

The Data

- Sds (science data stream) files
- 3 types
  - sds raw data, created immediately after the images are taken by the camera on the satellite
  - sds2 file contains all critical data and some supplementary data
  - sds1 file is a compacted version of the sds2 file

The Metadata

- information about observational parameters
- information about telemetry parameters
- information about the target image
- information about the exposure
  - **time and duration**
  - **pointing position of the telescope**
  - **which pixels were recorded**
  - **temperatures of the detector and electronics. etc.**
- Orbital information is downloaded from the website of Norad (N.A. Aerospace Defense Command)

The Metadata

- Existing metadata standards are not used
- The metadata schema was created by the MOST researchers and is specific for the files generated by the project

Data Transfer

- sds1 and sds2 are gathered to form a parsed file through a processor on the satellite
- transmitted from the satellite to one of the three ground station sites
- UBC is the repository for the data
- the parsed file is chopped again in its constituting sds files
- sds2 files are used to create FITS files

FITS Files

- Data captured from various sources and stored in various file formats are used to create FITS files
  - sds2 files
  - various metadata
- Considered important because they are the first instantiations in which all relevant data are captured
- Form the basis for all later on scientific analyses

FITS Format

- **F**lexible **I**mage **T**ransfer **S**ystem
- the most commonly used file format in astronomy for information exchange
- consist of a series of “extensions” and a set of headers
- can contain information about the imaging device used to capture the image, and more importantly, the time, date, and location of the telescope used
- MOST researchers create their own metadata/descriptive fields in FITS files

Accuracy

- Data check
  - technical analysis (done by software)
  - intellectual analysis (done by the instrument scientist)
- Recreation of FITS files
- After the creation/reduction, files are not changed

Data Analysis

- Reductions and analyses
- The data are reduced into series of intensities versus times
  - various data products can be generated from a single set of raw data by using different reduction methods
- Data analyses results in other types of files, e.g. light curves or images
- Reductions and analyses are used for presentations and publications

Access

- The raw and reduced data are made available to the Team members via an internal password-protected website
- Placed on CADC (Canadian Astronomy Data Center) accessible to all members of the astronomical community and the general public

Data Keeping

- MOST Archiving Manual
  - Processing
  - Archiving
  - backing up MOST data
- One computer as the main storage space of the data
- Only the FITS and sds files are routinely captured
- Data sets that are generated by the satellite camera are preserved and never changed or updated
- Files with scientific data are always preserved, even if they are corrupt or false

Data Keeping

- Name conventions are based on experience
- Directory-structure for storing files is based on
  - primary target (i.e. star)
  - time
- Regularly backed up on two DVD's (originally CD)

Overview

- The Project
- The Case Study
  - Background
  - Purpose
  - Model
  - Records identification
    - Diplomatic analysis
  - Records preservation
    - Issues
    - Recommendations

Background

- Proposed to the International Team of InterPARES2 in December 2004 at the Sicily workshop
- InterPARES 2 case study # 26
- Science focus
- The site of UBC is the subject
- MOST UBC Team
  - Astronomy scientist
  - Instrument scientist
  - Software consultant
  - Students

Purpose of the Study

- Preservation of scientific data
- Preservation of associated metadata that is critical for interpreting the data
- Collaboration b/w MOST and InterPARES to improve strategies and protocols for storing and distributing the data efficiently

Records Identification

Diplomatic Analysis

Record Components

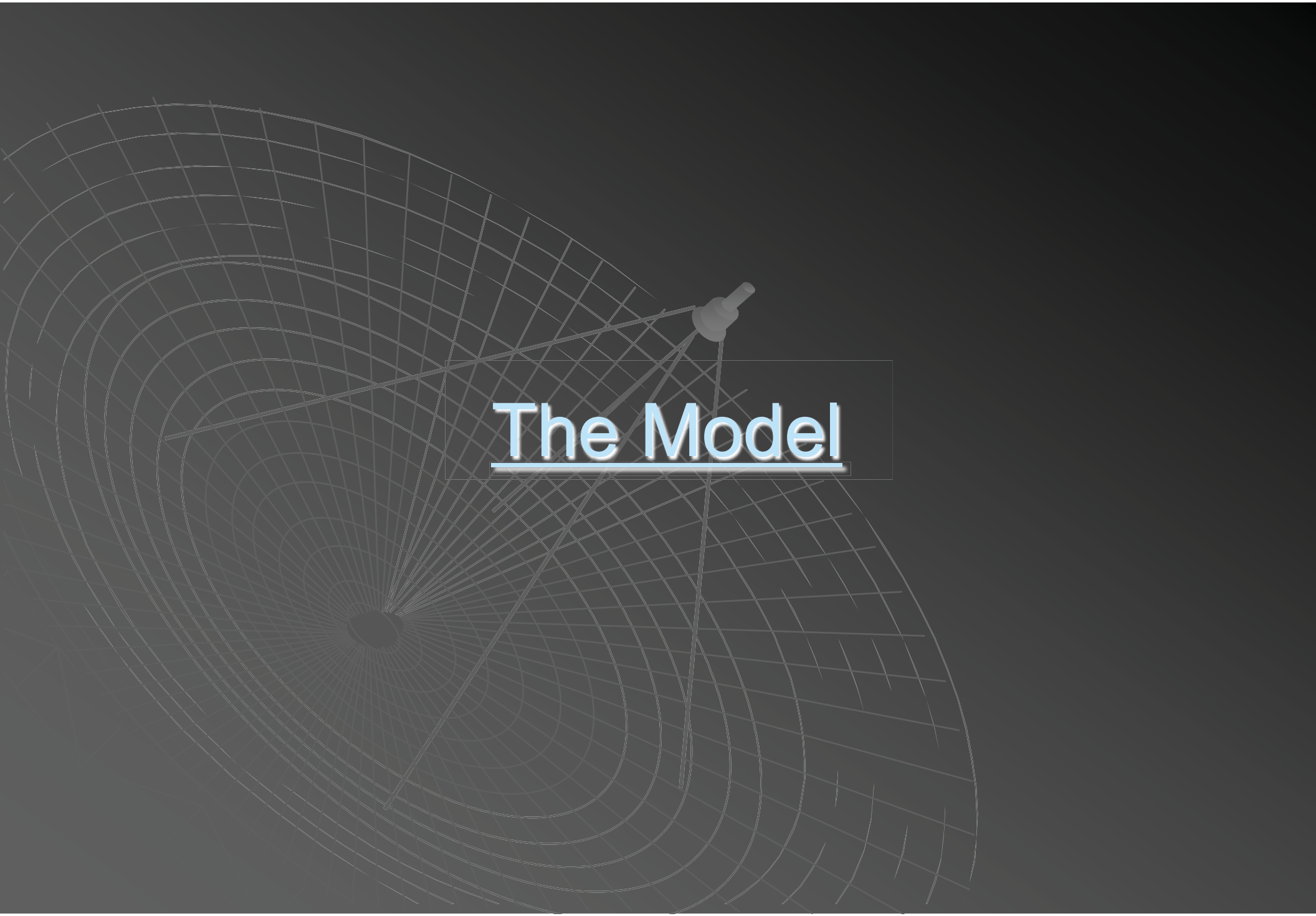
- Fixed content and format after the creation of data
- Data are captured on stable medium
- Data are instruments of fulfilling the projects purposes (action)
- Persons
  - Author
  - Writer
  - Addressee

Records Identification

Cont

- Relationships exist among data/files generated in the same action (archival bond)
- Identifiable contexts
  - Juridical-administrative context
  - Provenancial context
  - Procedural context
  - Documentary context
  - Technological context

Data/files are electronic records!


The Model

Preservation Issues

- No official written organizational structure within the research team
- No formal records management program
- No formal policies or procedures (e.g. selection of data for preservation)
- No preservation of commercial software
- Previous back-up data on CD have not yet been migrated to DVD

Recommendations

- Policies & procedures (written, formal)
  - access policy
- Establish RM with assigned responsibility
  - records inventory
  - records classification
  - records retention
  - technology monitoring
    - medium selection
    - file format selection
    - software preservation
    - migration

Recommendations

- Focus on recordkeeping & preservation
  - Raw data
- Appraisal criteria
- Descriptive fields in metadata for keeping & preservation (authenticity)

InterPARES 2: *Manage Chain of Preservation*