The PERICLES Long Term Data Preservation Project: Application to a Solar Database for Space Weather Applications.

C.Muller and the PERICLES consortium
B.USOC (Belgian Users Support and Operation Center).
Christian.Muller@busoc.be

B.USOC manages several space experiments on the ISS and other platforms at the benefit of the science investigators who have proposed these instruments. One of these projects is the SOLAR package, monitoring the solar spectrum since February 2008. The operations of SOLAR, normally foreseen for less than two years, have been extended and are now planned to last until 2017. B.USOC has a mandate to preserve the data and distribute it to the Principal Investigators who then derive scientific products which are published and archived in science database. SOLAR is used as a case study for the ICT FP7 project PERICLES (http://pericles-project.eu/). PERICLES aims to ensure that digital content remains accessible in an environment that is subject to continual change. This can encompass not only technological change, but also changes in semantics, academic or professional practice. PERICLES will take a 'preservation by design' approach that involves modelling, capturing and maintaining detailed and complex information about digital content, the evolving environment in which it exists, and the processes and policies to which it is subject. PERICLES represents a way for B.USOC of not only preserving the data and documentation of SOLAR but to transform this collection into a living archive. It is planned to include the products derived by the scientists and the related metadata generated by the science teams. In a further stage, in relation with the evolution of the mission duration itself and changes in operation procedures (for example since 2012, SOLAR measures full solar rotations near the solstices), the science teams can envisage higher level products and develop them from the newly reorganised archive.

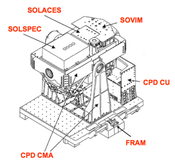
Partners

- 1 King's College London KCL UK: ccordinator
- 2 University of Borås HB Sweden: IT (large fashion and textile data base)
- 3 CERTH-ITI CERTH Greece: IT
- 4 Dotsoft DOT Greece: IT
- 5 Georg-August-Universität Göttingen UGOE Germany: IT (scientific library)
- 6 University of Liverpool ULIV UK: IT
- 7 Space Application Services SpaceApps Belgium: IT
- 8 Xerox XEROX France: IT
- 9 University of Edinburgh EDIN UK: IT
- 10 Tate Gallery TATE UK: media case
- 11 Industrious Media IM UK: IT
- 12 B.USOC (IASB BIRA) BUSOC Belgium: space case

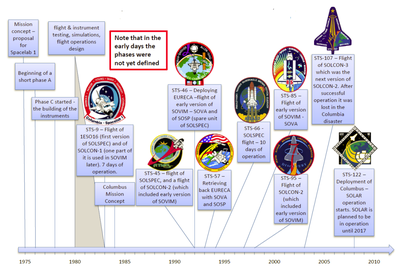
Two partners are essentially data providers, while all the others are essentially IT.

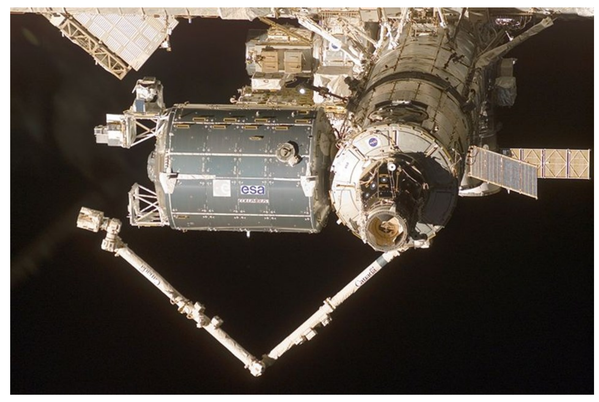
SOLAR

- The SOLAR payload is built from three complementary space science instruments that measure the solar spectral irradiance with an unprecedented accuracy across almost the whole spectrum: 17-3000 nm. This range carries 99% of the Sun's energy emission. Apart from the contributions to solar and stellar physics, knowledge of the solar energy flux (and its variations) entering the Earth's atmosphere is of great importance for atmospheric modeling, atmospheric chemistry and climatology. The three instruments are:
  - **SOLSPEC** (Solar Spectra Irradiance Measurements, developed by CNRS,France and IASB/BIRA, Belgium)
  - **SOL-ACES** (Auto-Calibrating Extreme Ultraviolet and Ultraviolet Spectrophotometers, developed by the Fraunhofer Institute, Deutschland)
  - **SOVIM** (Solar Variable and Irradiance Monitor, jointly developed by the Observatory of DAVOS, Switzerland and the Royal Meteorological Institute, Belgium)



SOLAR HISTORY





Why preserve solar data?

- Total solar irradiance and spectral solar irradiances are the main energy source of the earth system.
- Their short and long term variations influence the earth climate.
- We have verified data for only the last 30-40 years, the space age.
- Anything longer comes from proxies.

What are the data?

- Engineering documents
- Operation documents
- Documents generated during the operations.
- Telemetry
- Level 0 data transmitted to the PI
- Level 1 data and all scientific products generated by the PI's, associated metadata.
- Data published by the PI's.

USOC – User Support & Operations Center

- In 1998, ESA Manned Space Program board decided to adopt a decentralized infrastructure for the support of European payloads on-board the International Space Station (ISS). This concept was based on operating multiple User Support and Operations Centres (USOCs), each assigned to supporting a majority of tasks related to the preparation and in-flight operations of European payloads.
- The USOCs are based in national centres distributed throughout Europe.
- Depending on the tasks assigned to a USOC, they have the responsibility of a Facility Responsible Centre (FRC) or Facility Support Centre (FSC).

Columbus five internal payload racks at launch (courtesy ESA)

USOC – User Support & Operations Center

- All these USOCs use common software and hardware infrastructure to be securely connected to the Columbus Control Centre (Col-CC). Two kinds of datastream are exchanged between USOCs and Col-CC, **telemetry** and **telecommands**.

Engineering Documentation

- ESR : Experiment Science Requirements, drafted by ESA project Scientists detailing the requirements from the Principal Investigator (actually for SOLAR this ESR was only drafted when the payload was already operational for several years). Generally this is document is part of phase A.
- For SOLAR, in the Engineering DB there are 923 documents from the preparation phases 0 to D, these include:
  - Specifications and design documents
  - Acceptance data packages
  - Test plans
  - Acceptance, Functional and other Test Reports
  - Instrument/PL ICDs
  - Other ICDs (Interface Control Document)
  - Assessment reports
  - Safety data packages
  - Design Reports
  - CIDL (Configuration Item Design List)/ABCLs
  - User/Operations manuals
  - Certifications
  - Thermal analysis
  - Verification Control Document
  - ...

Operations documents

- **PODF: Payload Data File** is a XML file, providing a pre-defined sequence of commands and checks to operate the payload. For payloads interfacing Columbus every ground operator and crew member needs to follow a PODF to execute a payload activity.
- **Operations Manual**: B.USOC specific guidelines and procedures for the on console operations.
- **BUSOC wiki** (https://virtualweb.busoc.be/): the BUSOC wikipages provide an overview of the roles and responsibilities, guidelines on Increment preparation, use of the Predictor tool etc..
- **eRoom**: the eRoom is a Documentum installation that serves as a repository of interface control documents, user manuals, specifications, technical notes, procedures, protocols, minutes, presentation, reports, operational products, databases and emails.
- **Minutes of Meetings**: on a weekly basis there are meetings focused on the preparation and on-going operations. These can be internal only or with external parties from the Flight operations or the scientists

Documents generated during operations

- **OSTPV (Onboard Short Term Planner Viewer)** – OSTPV is a web application that displays timeline information about ground and onboard procedures, schedules and activities.
- **MDB** – The Mission DataBase contains a machine-readable description of the telemetry, including the size of various parameters sent in telemetry packets and their interpretation from binary to engineering values.
- **YAMCS** is a mission control system. It allows to acquire *telemetry* and dispatch *telecommands*. It can parse the telemetry stream and store it. YAMCS is an open source software developed by SpaceApps and BUSOC.
- **Telemetry**: Telemetry is data organized into packets that is sent by the payload to the control centre. Three main different kind of telemetry can be considered:
- **Housekeeping**
- **Health and status**
- **Science**

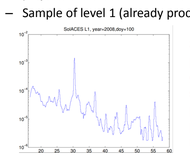
Data generated during operations

- **Telecommands**
- **Auxiliary data**
- **CEFN Tool** (– Columbus Flight Notes (CEFN)
- **Console Logs** – Console logs are short timestamped
- **SOLAR Mission Tool Light** – the SOLAR Mission Tool light was the precursor of the Predictor Tool.. Currently it is still used for short and long term handover information and to gather inputs for the BUSOC optimization board.
- **Checklist** - the checklist remains on console and provides the operator a list of routine tasks to be performed during his/her shift
- **Predictor Tool**: The SOLAR Predictor tool automatically fetches and parses information from a series of sources relevant for SOLAR Operations
- **Anomaly tracking tools**: two tools.
- **Planning tools**: again two tools

An exhaustive and detailed list would fill several poster spaces.

Science data

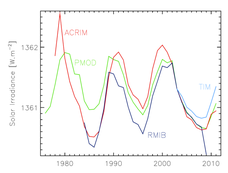
- The science data is accepted as raw data, and then **processed** to include the information from the calibration curves. It may later be further processed to include other calculations of the scientists.
- The raw data is usually referred to as **level 0** data. The data processed with the calibration curves information is called **level 1** data. However, in different projects the exact definition of the levels may change.
- Sample of level 1 (already processed) SOLACE level 1 data as graphs:



Science data generation is a scientific process by which the science team generates a products, reuse of the data should produce updated and new products. All the steps used in the original process should thus be preserved as metadata.

Which rules are considered by solar scientists for accepting the release of scientific data?

- **Peer review: publication in the refereed litterature.**
- **Acceptance by international bodies: COSPAR.**
- **The final objective might vary with time, coming from an initial objective: the solar constant to short and long term variations of solar output. (in 2013 some radiative models still use a parameter called solar constant)**



The comparison process requires intimate instrument knowledge and needs the scientific team to be present and cannot be reproduced when the scientific institutes have disappeared: the data should be archived to be reprocessed without the support of the PI.

Questionnaire on Data Evolution and Preservation

Your opinion on this preservation process, allowing the evolution of data on scientific observation, is important to our consortium. A questionnaire is available at this meeting from myself (B.USOC) and Rani Pinchuk (Space Applications Service).
In addition to the questionnaire, we are interested to interview several attendees. If you agree to share with us your requirements to such data preservation system, please contact us!

Conclusions

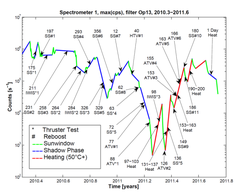
- The PERICLES project lasts until 2017 and an inventory of the data sources for the SOLAR project has already been done.
- The deliverable will be a tool allowing to use the record as an archive and to reuse it for further studies of the earth and solar environment and climate.

Scientists point of view

- The scientists group is divided into two sub groups – the scientists who work on a specific space experiment and the other scientists.
- Scientists working on space experiments are usually not interested in the very long term. They are interested in publishing papers about their findings, and are maybe careful to keep the data from competitor scientists. Currently there are no incentives to take the long-term view for the data. They consider the published science data as the final record.
- few new initial requirements for the longer scale:
  - Allowing to use the data together for further research.
  - Correlate the data from different missions.
  - Availability of both raw data with information of the instrument for proper interpretation for similar research.
  - Availability of processed data for future research – the data must be easily accessible for future research. For example, providing a searching tool allowing to enter some search criteria (for example some key words related to solar activity, or a time frame) and then the data of interest is displayed for download.

An example: Degradation and mitigation

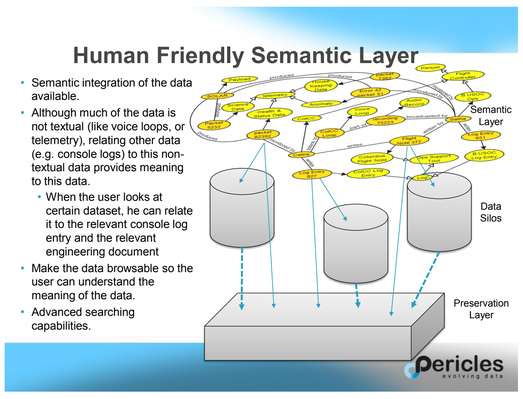
- Example: contamination of detectors of SOL-ACES
- The transmission of the spectrometer is restored by heating to 50°C. The filter wheel has been analysed as non contaminated. **The correction process needs both attended operations and an instrument science team.**
- The analysis of this and other contamination incidents on the ISS is still in progress.



The history of corrections and mitigation procedures is essential for data reuse.

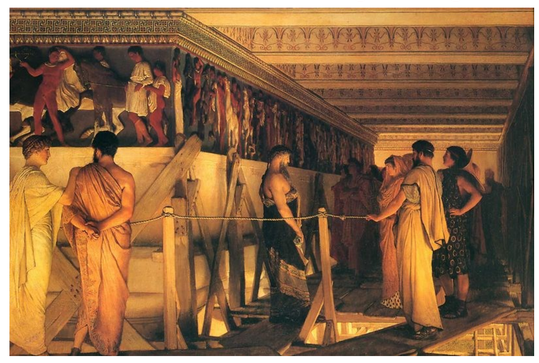
Human Friendly Semantic Layer

- Semantic integration of the data available.
- Although much of the data is not textual (like voice loops, or telemetry), relating other data (e.g. console logs) to this non-textual data provides meaning to this data.
  - When the user looks at certain dataset, he can relate it to the relevant console log entry and the relevant engineering document
- Make the data browsable so the user can understand the meaning of the data.
- Advanced searching capabilities.



An example of the developments under progress in PERICLES to integrate the different sources of data.

Final delivery: a product outliving its designers



Acknowledgments
The authors wish to thank the European Space Agency (ESA) (Directorate of Human Spaceflight) and the Belgian Science Policy Office (BELSPO) (ESA Prodex and other programs) for SOLAR funding.
PERICLES is funded under FP7 ICT Call 9, and addresses work programme objective ICT-2011.4.3 Digital Preservation and we are grateful to the European Commission for its support.

COST Action ES1005
TOSCA - Towards a more complete assessment of the impact of solar variability on the Earth's climate