

PERICLES: a knowledge management programme applied to solar data from International Space Station-Columbus

Christian Muller¹ and the PERICLES consortium

¹B.USOC (Belgian Users Support and Operation Centre), Royal Belgian Institute for Space Aeronomy, Avenue Circulaire, 3, B-1180 Brussels, Belgium
email: Christian.muller@busoc.be

Abstract. The FP-7 (Framework Programme 7 of the European Union) PERICLES project addresses the life-cycle of large and complex data sets to cater for the evolution of context of data sets and user communities, including groups unanticipated when the data was created. Semantics of data sets are thus also expected to evolve and the project includes elements which could address the reuse of data sets at periods where the data providers and even their institutions are not available any more. This paper presents the PERICLES science case with the example of the SOLAR (SOLAR monitoring observatory) payload on International Space Station-Columbus.

Keywords. preservation – reuse – space data – earth data – automation

1. Introduction

PERICLES (Promoting and Enhancing the Reuse of Information throughout the Content Lifecycle exploiting Evolving Semantics) is an FP-7 project started on February 2013. It aims at preserving by design large and complex data sets. PERICLES is coordinated by Kings College London, UK and its partners are University of Bors (Sweden), CErTH- ITI (Greece), DotSoft (Greece), Georg-August-Universitt Gttingen (Germany), University of Liverpool (UK), Space Application Services (Belgium), the Xerox company France and University of Edinburgh (UK). Two additional partners provide the two case studies: Tate Gallery (UK) brings the digital art and media case study and B.USOC (Belgian Users Support and Operations Centre) brings the space science case study. B.USOC supports experiments on the International Space Station and is the curator of the collected data and operation history. The B.USOC operation team includes B.USOC and Space Applications Services personnel and is thus ideally configured to participate in this project.

B.USOC has control over the entire data chain between the instrument in flight and the distribution of the data to the scientific teams which generate the science products.

2. The earth and space science use case

As a first test of the concept, B.USOC has chosen to analyse the SOLAR payload flying since 2008 on the ESA (European Space Agency) Columbus module of the ISS. Solar observation data are prime candidates for long term data preservation as variabilities of the solar spectral irradiance have an influence on earth climate. Long series of solar observations have also an important astrophysical significance as they help constrain the stellar models. The paradigm of these observations has already changed a lot in the last fifty years from a time where scientists were aiming at determining with high accuracy the solar constant which was the total solar energy per surface unit received at the top of the

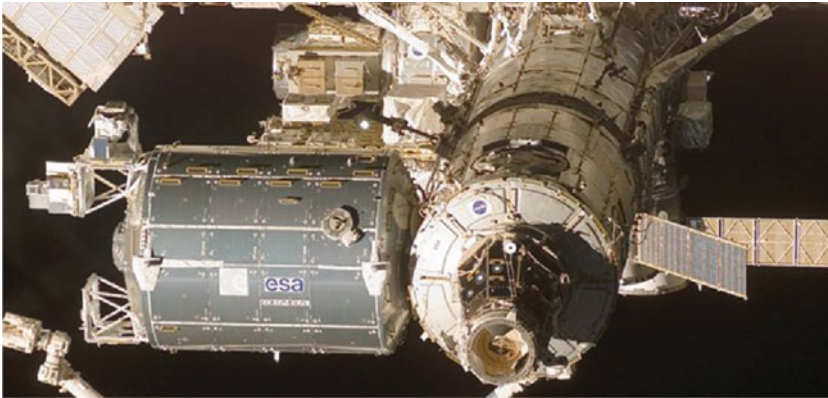


Figure 1. The SOLAR package on the outside of the ESA Columbus module center left of the picture (NASA, National Administration for Space and Aeronautics, document).

earth's atmosphere to the present situation where the same quantity is known as the total solar irradiance and has been shown by thirty years of space observations to vary of about one tenth of a per cent in synchronism with the solar cycle. Right now, larger variations have been detected at UV wavelengths but their effects on climate and atmospheric chemistry are still a matter of scientific discussion. SOLAR is especially interesting as a full data inventory could be performed from proposal level to the eventually recovered instrument. In this paper, the present stage of PERICLES will be described in the context of space operations and data handling, other cases in earth and space sciences accessible to B.USOC and the Royal Belgian Institute for Space Aeronomy are used in PERICLES demonstration, as will be examples available at NASA and other space agencies.

ESA data policy requires that all data manipulation is performed within the restricted area of B.USOC and are thus only available to B.USOC and SpaceApps personnel authorized in these areas, this has led PERICLES to use more open data sets as the final data of SOLAR archived by the scientists themselves.

3. Pericles specificities

Models are used in PERICLES as means to describe the existing Digital Ecosystems (DEs) and to enable an approach termed Model Driven Preservation, where ontological models are used to understand the impact of external changes on reuse of complex digital objects. PERICLES therefore provides the means to produce a structured, lightweight representation of the current ecosystem. DEs are composed using basic entities such as digital objects, technical services, policies, processes and user communities. We then provide techniques to validate the correct implementation of policies and to perform change management.

These principles will be applied to the use cases first in a data preservation and reuse approach with the objective of producing long term series of data using present and past experiments in order to determine trends and variations in parameters. The PERICLES appraisal tools will determine which parts of the data flow should be preserved to design a realistic long term data preservation policy. This first part will be proposed to space agencies which have a mandate for Long Term Data Preservation (LTDP). LTDP has an important scientific value as it constrains models of the evolution of the phenomena observed from space. It has also an essential societal value as it validates the models of climate change and thus increases the confidence in the forecasts of temperature and sea

level. SOLAR is a set of three instruments operated by B.USOC. The three instruments are described in a common publication (Schmidtke *et al.* 2006).

4. Pericles tools: policy editor, sheer curation and appraisal

PERICLES has developed several tools, the restrictions imposed by ESA HRE (Human and Robotic Exploration) on manipulation of the data for new objectives make the demonstration of these tools much easier with EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) data for which the distribution and reuse is explicitly agreed by the data provider. Precisely, the first tool to be considered is the policy editor which formulises the evolution of data policy.

The Policy Editor is a web-based tool that assists in the creation of a consistent and complete set of policies and constraints. It is based on selecting, instantiating and using policy blueprints from predefined sets of policy templates. These sets can be defined for example per subject, per industry or organisation, or on any other grouping that is deemed relevant. The blueprints can be designed once by experts, encapsulation best practices, and then (re)used across multiple projects. Policy blueprints can be hierarchically constructed and can contain variables. Using the Policy Editor, these variables are given a value and all other instantiated policies that refer to the same variable are updated with that value, reducing Data Management Plan preparation time and ensuring consistency. The fields are either provided as free text, or can be selected from a list of options originating from a data provider. The content and its structure of the instantiated policies depends on the appropriate policy model that the Policy Editor is configured with. It describes both the data and the way the data are produced, termed Contextual Information. This can include data, software and workflows. It is proposed to use the OAI-ORE format (Open Archive Initiative - Object Reuse and Exchange User User Guide[†]), which defines standards for the description and exchange of aggregations of web resources, to package such information. The OAI-ORE format can be used first for defining an OAIS Information Package.

In PERICLES, we refer to evolving collections of independent resources as a Digital Ecosystem (DE). Models of Digital Ecosystems are based on an abstract upper ontology, the Linked Resource Model (LRM). At its core the LRM defines an ecosystem by means of the constituent entities and the dependencies between them. The focus of the LRM is on dealing with change, which can be governed by change management policies incorporated into the model. The LRM can then be specialised to specific application areas such as space science by incorporating lower-level domain ontologies.

Sheer curation describes a situation in which curation activities are integrated into the workflow of the researchers creating or capturing data. Sheer curation approaches can greatly reduce the amount of effort that researchers need to undertake in annotating and preparing experiments for preservation. In our cases, it correspond to SOLAR where B.USOC is facility responsible facility.

The PERICLES Extraction Tool, PET (Corubolo *et al.* 2014) is a tool for the automated extraction of contextual information from the environment of digital objects or experiments. Once the tool has been configured its operation is largely transparent to the user. The PET tool can be applied in many situations, such as the running of scientific experiments or simulations from a PC. This capturing of metadata of the environment, together with context data, can be helpful to characterise user implicit knowledge. The PET was tested in B.USOC for UV solar data acquired by a ground based network of

[†] <http://www.openarchives.org/ore/1.0/>

the Royal Belgian Institute for Space Aeronomy, ESA data policy did not allow us to test it in the operational area of ISS (International Space Station) SOLAR.

Appraisal and selection process designates specific datasets to be preserved as it is unfortunately unrealistic to preserve all the data flows relevant to a space experiment. Since these data may not be known precisely in advance, it will be needed to specify appraisal criteria. Automation of the appraisal processes would ensure objectivity and a possible reuse of the data by scientists with different objectives than at experiment design. In PERICLES Deliverable D5.2 (<http://pericles-project.eu/deliverables/53>), we produced a comprehensive catalogue of appraisal criteria and began an investigation of their automation.

5. Conclusions

The systematic data preservation programs for earth and space science data become more and more a necessity as global change must be determined accurately. This process constrains the models for future evolution and thus moves from the scientific field to the societal one. The PERICLES process provides the tools towards a more dynamical data preservation.

6. Acknowledgments

This work was supported by the European Commission Seventh Framework Program under Grant Agreement Number FP7601138 PERICLES.

References

- Corubolo, F. Eggers, A. G., Hasan, A., Hedges, M., Waddington, S., & Ludwig, J. 2014, *A pragmatic approach to significant environment information collection to support object reuse*, IPRES proceedings.
- Schmidtke, G., Frhlich, C., & Thuillier, G. 2006, *ISS-SOLAR: Total (TSI) and spectral (SSI) irradiance measurements*, Advances in Space Research ,Advances in Space Research , 37, 255-264.