Annex 9: PV 2013 paper

A Data e-Infrastructure for the International Space Station

Luigi Carotenuto [1], Rainer Willnecker [2], Patrick Hambloch [2], Sébastien Barde [3], Olivier Belbis [3], Jean-Christophe Malapert [3]

[1] *Telespazio*
Via E. Gianturco 31, 80146 Napoli, Italy
EMail: luigi.carotenuto@telespazio.com

[2] *DLR*
Portz-Wahnheide, Linder Höhe, 51147, Cologne, Germany

[3] *CNES*
18 Avenue Edouard Belin, 31400, Toulouse, France

ABSTRACT

CIRCE (Cooperative International space station Research data Conservation and Exploitation) is a coordination action promoted by Telespazio, DLR and CNES, co-funded by the European Commission. Its goal is fostering cooperation with the international partners of the International Space Station (ISS) and it aims at designing a roadmap for the establishment of a European e-infrastructure geared towards long-term preservation and exploitation of scientific data produced on board the ISS.
Together with experts from research and industry and representatives of the Space Agencies involved in the ISS program, the project is elaborating on the status of technologies, as well as policies and regulations involved in the exchange of data in the scope of scientific exploitation of ISS.
The paper describes the vision, the challenges and the next steps that are necessary in establishing such an e-infrastructure. Special attention is given to knowledge representation for data searching and data access policy, as key factors for data exploitation.

Keywords: data e-infrastructure, International Space Station, data preservation, data exploitation, knowledge representation, data policy

INTRODUCTION

The International Space Station (ISS) is the most complex and powerful laboratory for research in space; its utilization concerns scientific disciplines ranging from Life to Physical Sciences as well as technology developments.

Presently the data produced by experiments on board the ISS and other microgravity platforms are delivered to the research teams that conceived the experiments; the research teams analyze the data and the results are condensed in the set of articles published by the research team on the experiment subject. Once the analysis is completed, data is stored in local archives, becoming rapidly scarcely visible to the scientific community; moreover its preservation in the long term is not ensured, as well as in particular the preservation of the huge amount of information that is generally needed to interpret the data. Therefore, any use of the data for other research purposes is generally prevented.

On the other hand, space data is an important scientific asset because:

- space experiments demand a huge amount of resources, and consequently also successful experiments are scarcely repeatable
- space experiments are selected through a rigorous peer review; therefore space data is expected to have a relevant scientific content
- space experiments concern many scientific domains (from life to physical and material sciences)

Consequently, use of space data should be maximized, due to its potential contribution to the development of knowledge and applications for the citizens in many domains. The preservation and utilization of space data in the long term after the completion of the space experiments is a pre-requisite for maximizing the return of space missions. To this purpose a data e-infrastructure is needed to support a systematic preservation and utilization of data produced by space experiments on ISS and other space platforms.

As first initiative aimed at approaching the above challenges for pursuing the preservation of scientific space data, the European User Support & Operation Centres (USOC) with other research institutions and companies have promoted the ULISSE project [1-7]. ULISSE ("USOC knowLedge Integration and dissemination for Space Science Experimentation") is a research project, funded by the European Commission within the first Space Call of the Seventh Framework Programme (FP7). The general goal of ULISSE is improving preservation, valorization and exploitation of the data produced across multiple domains by European scientific experimentation in space on the ISS and other space platforms.

The wide range of scientific disciplines involved in space experimentation and the international scale of the ISS programme pose some challenges for the realization of a data e-infrastructure. ULISSE has explored the main issues of data preservation for ISS, identifying possible solutions and tools and implementing a Demonstrator which provides the main services for data access and use integrating a set of distributed resources.

ULISSE has provided a very fruitful experience and the results of the Demonstrator have proven the feasibility and usefulness of an e-infrastructure to interoperate the ISS distributed repositories in accordance with the applicable legal constraints.

The ULISSE experience has also pointed out that the establishment of an operative data e-infrastructure would require a close cooperation among the main ISS stakeholders. To respond to this need, Telespazio, DLR and CNES are promoting CIRCE (Cooperative International space station Research data Conservation and Exploitation), a coordination action co-funded by the European Commission within the 7th Framework Programme (FP7). Final goal of CIRCE [8] is developing a common vision for the ISS data e-infrastructure identifying a roadmap for its implementation.

Starting from the ULISSE experience, CIRCE addresses all the main aspects that need to be defined for an ISS data e-infrastructure, as:

- metadata structure, contents and management
- knowledge representation and standardization of ontologies/taxonomies
- semantic technologies for knowledge searching and data mining
- data elaboration and post-processing for feature/event recognition, modelling and validation
- data policies, preservation of intellectual property rights and promotion of cooperation for data exploitation
- user authentication, authorization and accounting, in compliance with the applicable data policies
- integration and management of distributed resources, ensuring their interoperability
- identification of operative processes to feed a data e-infrastructure compatible with the processes and rules of the European Ground Segment for ISS

The above topics are analysed also with the support of external experts of the relevant technological fields, with the purpose of consolidating the scenario and requirements for the ISS domain, assessing the technology status and identifying possible gaps to be filled with future research actions.

CIRCE started on July 2012 and will end on December 2013; a first workshop has been organised in February 2013 with the participation of representatives of space agencies, scientific communities, information technology providers. A second workshop is under preparation before the conclusion of the project.

This paper discusses the present achievements of the first workshop, highlighting in particular the challenges related to knowledge representation in a multi-disciplinary environment and to policies for accessing and using data, and presenting the future prospects on the subject of ISS data preservation.

REPRESENTATION OF INFORMATION

Data searching requires the availability of information about the scientific context of the available data; suitable scientific ontologies are needed; their design and evolution has to sustain several functionalities. Firstly scientists not involved in space projects should be able to find the space data even without any information about space projects; secondly scientific ontologies have to promote the data cross-feeding among different scientific domains. Moreover, interoperability with other scientific and service infrastructures and data centres has to be ensured also through proper standardization of semantics and metadata.

The wide range of disciplines addressed by research on ISS makes the representation of the related information a hard challenge. Any representation model would in fact be able to describe very diverse contents ensuring at the same time the interoperability of the many distribute repositories and resources that produce and manage the information.

During the ULISSE project, two data models were available: one derived from on a XML Schema (ISO 19139-2), implemented through the Geonetwork application [9], and another one based on semantic technology (topic maps) [10-12].

The XML schema (ISO19139-2) defines a set of finite elements to be filled by the content administrator where topic maps defines a semantic solution, which can describe the full knowledge of a content by addition of semantic vocabulary.

Usually, getting a general consensus from the community upon any standardization of the knowledge about a domain within a XML schema is a hard work. Thus, standardize in one unique XML schema the full knowledge of several domains is extremely complex.

On the contrary, semantic technologies do not have any limit about describing a knowledge because the technology allows to add easily new semantic vocabulary.

During ULISSE it was difficult to keep in synchronize the content of these two data models based on two different technologies. However, merging XML schema and semantics was a very positive experience. In fact, once a core of XML schema is identified, metadata can be described at a high level. Then, it is possible to integrate this schema in the semantic technology. In this way, a core data model is available via semantic technology. Moreover, the data model can be extended easily via semantic technology by each content administrator to fully describe the knowledge about the domain at any extent. It is therefore suggested to define a core data model, to be used for all disciplines, to be extended using specific ontologies and semantic technologies.

On the other hand, the ULISSE Metadata Standard was developed tailoring the ISO 19139 schema that was conceived to describe observational data. Due to its rigidity, this schema cannot provide a detailed description of space experiments data. It is therefore suggested to drop off the reference to the ISO 19139 standard and to develop a new schema.

The core of this information is the description of the experiment dataset (in terms of contents, formats, files, storage location, access rules, etc). Also other general information about the experiment are included into the metadata standard, in order to provide a uniform structure for searching keywords; this information concerns the investigators team members (names, affiliations, etc), the space platform, mission and instrument used for the experiment, the scientific discipline, the sponsor agency, references or links to related publications, etc. The ULISSE schema takes the experiment as the univocally determined resource that provides the reference for identifying the entire dataset produced by the experiment itself. However, this approach fits with the research in Physical Sciences (where experiments indeed represent isolated projects), but does not describe properly the research in Life Sciences (where, in Human Physiology for example, different datasets could be produced on different subjects in different missions using the same instrument, with similar or different protocols).

It is therefore suggested to adopt the dataset as the reference, univocally determined resource, labeled with a unique identifier and linked to its specific metadata and to the applicable topics belonging to a more general ontology of the domain.

DATA POLICY

Access to data requires an adequate data policy, suitable to protect intellectual property while not impeding data exploitation.

Presently no common policy for data usage is available in Europe. Consequently no standard procedure has been established yet for accessing science data and interested users must identify reference persons for requesting data on a case-by-case basis. The establishment of a common policy for data use would help removing this relevant bottle-neck for data exploitation.

CIRCE has performed a review of practices and policies adopted by the Space Agencies, whose results are here summarized. Normally space data is closely related with other data generated on ground (as calibration data, reference data, data from other ground-based experimentations) that normally belong to scientists. The approach for raw space data is different for various Countries:

- in Germany DLR assigns the property of scientific data (including raw data) to the investigators which are committed to publish results within two years
- in France CNES has adopted a data preservation policy and has developed the SERAD service for creating, maintaining and making available to users archives of the CNES space missions after their completion. Metadata is publicly available; data is accessed upon CNES authorization (also with permission of investigators)
- in Italy ASI retains the property of data produced by researches funded by the Agency; in that case ASI collects publications (to be authorized by the Agency) into a digital library. In all other cases the applicable rules assign the property of data to receiving entities
- ESA has issued a data policy [13] that states the priority of science team of the experiment for the analysis of data; release and use of data is restricted to the science team for a defined period (named restricted period) that usually is one year long after the completion of the experiment but can be prolonged upon motivated request. After this period other users may request access to data; data release has to be authorized by ESA after an evaluation of user motivations for data use, taking into account user nationality and the opinion of the science team of the experiment.

The data policies scenario depicted at the first CIRCE workshop on ISS data exploitation is summarized in the following Table 1.

Space Agency	**Agency projects**			**Third Parties projects**		
	Data Property	Central archiving	Distribution to users	Data Property	Central archiving	Distribution to users
DLR	investigators	NO	NO	investigators	NO	NO
CNES	Agency	YES	upon authorization	TBD	NO	NO
ASI	Agency	NO	NO	TBD	NO	NO
ESA	Agency / investigators	YES	upon authorization	TBD	NO	NO
CSA	investigators	NO	NO	TBD	NO	NO
Roscosmos	Agency / investigators	YES	upon authorization	TBD	NO	NO

Table 1: Main features of space data management established by Space Agencies

It is worth noting that also scientists indicated the need of authorizing access to data. In fact, scientific users require credits for the data authors, while space agencies require the verification that data use is compliant with laws, ethics and other possible constraints. The need of a common policy has been confirmed at the first CIRCE workshop. In fact, a common data policy would ease the management of data increasing the scientific return of space missions. The ESA data policy [13] may provide a useful guideline on the matter; however, a relevant upgrade is required, in order to promote and sustain a cooperative approach to data and

knowledge share and exploitation.

Therefore, the above requirements impose the identification and accounting of data users. even if this functionality has not been explicitly required by the users themselves. A recent study on Authentication and Accounting Infrastructures [14] has provided requirements for the field that appear applicable to an ISS data e-infrastructure too. Being a potential barrier for data share and cooperation, user authentication and authorization should be performed at international level; a single harmonized international infrastructure for user authentication and authorization would be advisable to support the implementation of data policies in the context of a data infrastructure.

At the same time, the need of a policy requires also the establishment of a governance for data preservation and dissemination. A governance board, acting on behalf of the owners of data and intellectual property in general, should oversee to the processes of feeding and managing the repositories for preservation, updating the contents and the infrastructure, and managing user accounting and access to data. The board would monitor that the above tasks would be performed in compliance of the ISS Ground Segment rules and constraints; moreover, the board would also ensure a proper reporting on preserved information and its use to the stakeholders.

CONCLUSIONS AND FUTURE STEPS

The current roadmap identifies a set of partial objectives which contribute to the achievement of the final goal: enabling the user communities to make the maximum use of space data in the long term. These partial objectives are:

- Ensure as soon as possible the archiving of operational data produced daily by the ISS experiments (the huge amount of telemetry data that include raw science data, system and command history, videos and voice generated during the execution of a space experiment)
- Development of a data e-infrastructure enabling users to find, get access to and exploit space data
- Preservation of data, meta-data and knowledge in the long term

The implementation of the roadmap is based on the following drivers and guidelines:

- Achievement of a common vision of data preservation among ISS partners, in terms of data policy, common services, functionalities and processes
- Science-driven definition of objectives, services and functionalities of the data e-infrastructure, to meet the needs of the user communities
- Best use of available funding sources through coordinated programmes (within European Union, space agencies and national programmes)
- Maximum exploitation of the existing national and international assets

In accordance with the above concepts, the following main steps have been identified and grouped in three implementation phases.

Preparatory Phase

Main tasks are:

- Assessment of user requirements for exploitation of space data
- Achievement of a standard definition of data description (meta-data) able to unite interoperability and differences among science domains, according to user needs
- Definition of a common policy for data share and use, able to unite preservation of intellectual property rights and promotion of data and knowledge share
- Inventory of the existing operational data produced by experiments already performed on ISS
- Preparation of meta-data for the existing operational data
- Upgrade of the existing repositories within the ISS ground segment for the temporary storage of operational data
- Temporary storage of existing operational data and meta-data in the repositories of the ISS ground segment

Implementation Phase

Main tasks are:

- Preservation of the future operational data, preparing meta-data, formatting data for the temporary storage in the repositories of the ISS ground segment
- Definition of system architecture and system requirements for data exploitation
- Detailed design and implementation of the data e-infrastructure for space data exploitation
- Population of the knowledge base (meta-data for data exploitation) according to the user needs
- Management of data dissemination and use in accordance with the data policy, establishing an infrastructure governance board

Long term preservation Phase

Main tasks are:

- Upgrade of the existing ESA data centres for the long term storage of space data
- Archiving of operational data, related meta-data and knowledge base contents for long term preservation
- Management of long term archives, upgrading storage technologies and data formats to ensure content accessibility
- Management of data dissemination and use in accordance with the data policy, maintaining and updating the data e-infrastructure

From the above roadmap, it is evident that the establishment of a data e-infrastructure for the exploitation in the long term of the ISS scientific data will demand the support and active involvement of all stakeholders of space research: European Commission as funding entity and policy maker, space agencies, as funding and management entities, ICT companies and experts to contribute to the implementation, scientific communities to tailor services in accordance with their needs. This list is not exhaustive; the identification of all needed parties and the promotion of coordinated initiatives on the subject are the ambitious objectives of the CIRCE project. Results will be made available through the project web site [8].

REFERENCES

[1] - ULISSE project page: http://www.ULISSE-space.eu/

[2] - E.A. Kuijpers, L. Carotenuto, J.-C. Malapert, D. Markov-Vetter, I. Melatti, A. Orlandini, R. Pinchuk: "Collaboration On ISS Experiment Data And Knowledge Representation", 63rd International Astronautical Congress, Naples, IAC-12-D5.2.11 (2012)

[3] - L. Carotenuto, A. Berthier, M. Bernat, V. Lapuerta, R. Pinchuk, D. Pasquariello, G. Minei, C. Albanese: "The ULISSE contribution to education and outreach", paper IAC-12-E1.7.11, 63rd Congress of Int. Astron. Feder., Naples (2012)

[4] - "Laboratory Science with Space Data", D. Beysens, L. Carotenuto, J.W.A. van Loon, M. Zell (Eds), Springer (2011)

[5] - L. Carotenuto: "ULISSE – Usocs knowledge integration and dissemination for space science experimentation", in "Let's Embrace Space", European Commission DG Enterprise Editor, 340-349 (2011)

[6] - E.A. Kuijpers, L. Carotenuto, C. Cornier, D. Damen, A. Grimbach: "The Ulisse Environment For Collaboration On ISS Experiment Data And Knowledge Representation", 61st International Astronautical Congress, Prague, paper IAC-10-D5.2.2 (2010)

[7] - ULISSE Promotional Portal: *http://ulisse.medes.fr*

[8] - CIRCE project page: *http://www.CIRCE-space.eu/*

[9] - http://geonetwork-opensource.org/

[10] - D. Damen and R. Pinchuk: "Creating topic maps ontologies for space experiments", Proc. TMRA 2009 conference, Leipzig, p. 3-12 (2009)

[11] - G. Sören-Schwartze, R. Willnecker, S. Sous: "Data and Information Management of ISS Payload And Experiment Data", 62nd International Astronautical Congress, Cape Town, SA, paper IAC-11-D5.2.8 (2011)

[12] - D. Damen and M. Patriksson: "Putting Topic Maps to REST", Proc. TMRA 2010 conference, Leipzig, p. 9-18 (2010)

[13] - "Human Spaceflight Data Policy", ESA/PB-HSR(2004)29, Rev. 2, Corr. 1

[14] - "Advancing Technologies and Federating Communities", AAA Study Final Report (2012) http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/aaa-study-final-report.pdf

The authors gratefully acknowledge that the activity leading to the present publication has received funding from the European Union Seventh Framework Programme (FP7 / 2007 – 2013) under the grant agreement nr. 313058.