Planetary and Space Science xxx (2017) 1–13

Contents lists available at ScienceDirect

Planetary and Space Science

journal homepage: www.elsevier.com/locate/pss

ROSETTA: How to archive more than 10 years of mission

Maud Barthelemy [a,*], D. Heather [a], E. Grotheer [a], S. Besse [a], R. Andres [a], F. Vallejo [a], T. Barnes [b], L. Kolokolova [b], L. O'Rourke [a], D. Fraga [a], M.F. A'Hearn [b], P. Martin [a], M.G.G.T. Taylor [c]

[a] ESA/ESAC, ES-28691, Urb. Villafranca del Castillo, Madrid, Spain
[b] SBN/PDS, University of Maryland, College Park, MD, 20742-2421, USA
[c] ESA/ESTEC, Keplerlaan 1, Noordwijk, The Netherlands

ARTICLE INFO

Keywords:
Comet
Long term archive
International mission
PSA
PDS
Rosetta

ABSTRACT

The Rosetta spacecraft was launched in 2004 and, after several planetary and two asteroid fly-bys, arrived at comet 67P/Churyumov-Gerasimenko in August 2014. After escorting the comet for two years and executing its scientific observations, the mission ended on 30 September 2016 through a touch down on the comet surface. This paper describes how the Planetary Science Archive (PSA) and the Planetary Data System – Small Bodies Node (PDS-SBN) worked with the Rosetta instrument teams to prepare the science data collected over the course of the Rosetta mission for inclusion in the science archive. As Rosetta is an international mission in collaboration between ESA and NASA, all science data from the mission are fully archived within both the PSA and the PDS. The Rosetta archiving process, supporting tools, archiving systems, and their evolution throughout the mission are described, along with a discussion of a number of the challenges faced during the Rosetta implementation. The paper then presents the current status of the archive for each of the science instruments, before looking to the improvements planned both for the archive itself and for the Rosetta data content. The lessons learned from the first 13 years of archiving on Rosetta are finally discussed with an aim to help future missions plan and implement their science archives.

1. Introduction

Rosetta was a Cornerstone Mission of the ESA Horizon 2000 programme. Launched in March 2004, it arrived at its final destination, comet 67P/Churyumov-Gerasimenko (CG), in August 2014. During its 10-year cruise to the comet, Rosetta observed multiple targets, including close encounters with the asteroids Steins and Lutetia, becoming the first European mission to encounter these primitive objects. After arrival at 67P/Churyumov-Gerasimenko, the comet nucleus and coma were studied in detail, both from orbit with the Rosetta spacecraft as well as in-situ with the 'Philae' lander. Rosetta was the first mission to encounter and then follow a comet as it passed by the Sun, and was also the first mission to land a probe on a comet nucleus. At the end of September 2016, the operational mission ended when Rosetta touched down on the surface of the comet.

The scientific data returned by this mission are vital to improving our understanding of the complex and changing nature of comets as they travel around the Sun, and of the formation and evolution of the Solar System. The responsibility to preserve and disseminate these data lies with the mission's science archive (Section 2), ensuring that the data and information from Rosetta are stored in a scientifically useful format, and made available to the science community for years to come.

Rosetta was an international mission between ESA and NASA, carrying a total of 25 instruments, 3 of which were provided by the US. As part of the international agreement between the two Agencies, a complete copy of all Rosetta science data is required both at ESA's Planetary Science Archive (PSA) and at NASA's Planetary Data System (PDS). Sections 2.1 and 2.2 describe the PSA and PDS systems in further detail.

The complex nature of the Rosetta mission presented many challenges to the implementation of the archiving systems, and the specific setup of the archive for Rosetta is discussed throughout Section 3 of this paper. With 25 science instruments from multiple countries, two spacecraft, two space agencies, and many official bodies and workings group within the archiving systems, there was a need on the mission to clearly identify the responsibilities of each party within the process, and this is described further in Section 3.1.

The basic process followed by the Rosetta archive is discussed in Section 3.2, outlining the flow of the data from the instrument teams to

* Corresponding author.
E-mail address: maud.barthelemy@esa.int (M. Barthelemy).

http://dx.doi.org/10.1016/j.pss.2017.07.004
Received 27 May 2016; Received in revised form 16 June 2017; Accepted 12 July 2017
Available online xxxx

the final archive. To help guarantee usability and readability of the science data, Rosetta follows the so-called PDS3 Standards, a set of norms and requirements that aid the understanding of scientific data holdings in the long term. A critical part of the process is related to the technical validation of the data to ensure it complies with these PDS Standards as well as any additional requirements from the PSA and PDS archiving systems. The validation processes and tools developed for use on Rosetta are discussed in Section 3.3.

As well as the technical validation, all data delivered by Rosetta must undergo an independent scientific evaluation. This is achieved via a science peer review process, fully supported by both the PSA and PDS teams, with review meetings held in Europe and in the US. The full review process is described in detail in Section 3.4.

Through its many years of operations, Rosetta observed multiple targets (planets, asteroids and 67P-Churyumov Gerasimenko), and the instrumentation evolved as more time was spent in space. Data within the archive must therefore reflect these changes over time with corresponding updates to the calibrations and metadata describing instrument performance. Other missions such as NASA/ESA's Cassini/Huygens (Witasse et al., 2007) and NASA's New Horizons have faced similar problems, and having direct collaboration and support from experts at NASA's PDS has therefore been key to overcoming many of these issues. Outside of the challenges presented by the nature of the mission itself, changes to the archiving infrastructure, validation tools and the archive standards within the lifetime of Rosetta have also been difficult to manage. These and other specific issues experienced by the implementation of the Rosetta archive are discussed in detail in Section 3.5.

At the time of writing, Rosetta is in the post-operation phase, mainly dedicated to the final population of the archive with complete and enhanced data from each instrument. A detailed description of the current content of the Rosetta archive for each science instrument is provided in Section 4.

Details of the current activities being undertaken as part of the post-operations and archive enhancement are provided in Section 5, with significant effort being put into populating the archive with the best data possible before the instrument team support is cut and the specific support for the mission from ESA ends. When the post-operations phase ends in late 2019, the Rosetta Archive will become a so-called ‘legacy’ archive, where the data themselves are preserved and are effectively static, so it is vital to ensure the best data are in the archive at that time.

The paper closes with a summary of the lessons learned from the archiving on Rosetta (Section 6), outlining the many challenges that have been faced and suggesting approaches for future missions to follow in order to either avoid or mitigate the impact of similar issues within their implementation.

2. Archiving of planetary mission data

Archiving is an integral part of any science mission. In the long-term, the legacy of any mission will be the science results derived from the data returned, and it is therefore critical that those data are preserved and made available to the scientific community both now, and in the long-term, after the instrument teams disband.

Preserving data on a disk is straightforward, as is making that disk's contents available to the community. The objective and challenge for the long-term archive is not simply to store the data and knowledge over the course of a mission and judge its value for today, but to try and ensure that it will still be useful to scientists several years from now. Today's software and data formats are not going to be the same as those used in years to come, so there is a need to preserve not only the data itself, but also clear descriptions and documentation concerning the instrument operations, data structures, calibrations, ancillary data, geometry and scientific status of the archive content.

Preserving the data in a central, Agency-level resource such as the PSA and PDS rather than leaving individual instrument teams to preserve their own data also ensures long-term maintenance and accessibility, as ESA and NASA will commit to funding the science archives long after the mission has ceased to operate.

For all of the above reasons, archiving is a core activity on all ESA science missions, and the PSA was designed in response to these needs for the planetary science community.

2.1. The European space Agency's Planetary Science Archive (PSA)

The need for ESA to establish a science data archive for planetary exploration missions stemmed from the approval of Rosetta as the first cornerstone of the Horizon 2000 programme. This became the starting point for the definition of the PSA, with the view that it should become mandatory for all future ESA planetary missions to use it to store and disseminate their science data.

For Rosetta, all science and engineering data must be fully prepared by the instrument teams, and follow an external validation and review process before being included in the archive. If any change is required to the data after this, a new version must be prepared outside of the archiving systems, and will follow the same process before being added. No data pipelines are running within the PSA archives, and all data preparation and calibration is completed external to the systems. Note that for ESA planetary missions, each instrument team is funded by their national space agency for the instrument development, operations, science analysis and archiving.

This so-called ‘long-term’ archiving approach was selected for the PSA from the outset and the systems were set up to manage the archive and its content accordingly. For these purposes, ESA followed the recommendations from the CCSDS (Consultative Committee for Space Data Systems) mentioned in the OAIS (Open Archival Information System) (OAIS, 2012).

The decision to develop a centralized archive for all of ESA's planetary missions has allowed the PSA today to serve the planetary community as a whole with all of our data (e.g. Giotto, Mars Express, Venus Express, Huygens, SMART-1, and Rosetta) from a single access point (Cosmos.esaa). The core objective of the PSA is the provision of a centralized repository that will preserve and deliver peer-reviewed data products from all of ESA's planetary missions to the scientific community. In order to do this, and meet the requirements for a long-term science archive outlined in the previous section, the PSA needs to be much more than a physical data store. The PSA is in fact the entire initiative, the setup, the process and the implementation to preserve data from ESA's spacecraft to planetary bodies, as well as supplementary information such as calibration files and supporting observations from ground-based observatories.

It was decided at an early stage that the PSA would follow NASA's Planetary Data System (PDS) standard (PDS, 2003) as a baseline for defining the structure and format of the data sets and their content. At the time of archive definition, the PDS was already a well-known and understood standard within the planetary science community, allowing for archiving, description and preservation of data. Use of this standard also ensures cross-compatibility with all other PDS compliant data. The direct support from the Small Bodies Node of the PDS on Rosetta was a great help in setting up the PSA according to the Standards and in ensuring the content of the archive is compliant.

The PSA team, consisting of scientists and archiving experts, provides consultancy to all of the data producers throughout the entire archiving process. This begins as soon as an instrument is selected, working with the instrument team to define a set of data products and data set structures that will be suitable for the long-term archive. Consultancy is also provided to support the instrument teams with their documentation and knowledge management. One of the most important documents is the Experiment to Archive Interface Control Document (EAICD), which provides a full description of the data sets being archived by an instrument, including a summary of the data pipeline, the structure of the data products, data sets and any calibration or geometry information provided. This document should be the starting point for anybody not familiar with the instruments when trying to understand a data set.