SCIENTIFIC VALIDATION IN AN OPERATIONAL ARCHIVE – EXPERIENCE FROM BEPICOLOMBO

M.S. Bentley[1,3], T. Cornet[2,3], S. Martinez[3], A.J. Macfarlane[4,3], R. Moss[5,3] [1]HE Space for ESA (mark.bentley@esa.int), ESAC, Camino Bajo del Castillo s/n, Urb. Villafranca del Castillo, 28692, Villanueva de la Cañada, Madrid, Spain, [2]Aurora for ESA, [3]European Space Astronomy Centre (ESAC), Camino bajo del Castillo, s/n Urbanización Villafranca del Castillo, Villanueva de la Cañada, E-28692 Madrid, Spain, [4]Serco for ESA, [5]Vega Telespazio for ESA

Introduction: Missions delivering PDS4 data to the ESA Planetary Science Archive (PSA) [1] use an operational approach. Rather than collecting and preparing an archive of data for a given period (e.g. months), data products are produced and archived regularly. The challenges of ensuring technical and scientific validation, as well as data integrity and completeness, in such an operational archive are discussed here.

Data processing for BepiColombo: For the ESA/JAXA BepiColombo mission, raw products are generated from telemetry after each ground-station pass, using either software developed by the Science Ground Segment (SGS), or code provided by the instrument teams and integrated using Docker. Higher level processing pipelines are *required* to use PDS4 products as inputs to ensure traceability through the different processing levels. After reception by the SGS, the key product meta-data are registered in a database (e.g. LID, instrument name, start/stop times, etc.) and the products stored in the mission repository. At the same time, these products are sent to the Planetary Science Archive (PSA) for ingestion.

Benefits and challenges: The operational archive concept has the advantage that usable data products are generated very soon after the reception of telemetry. As a consequence, the BepiColombo science team is much more likely to use the archive products for their analysis, rather than generating a separate science product, and leaving archive product generation until later in the data life-cycle, as in some previous planetary missions. This gives a first level of validation simply because many more users are rapidly inspecting the data. This is amplified, in the case of BepiColombo, by the use of the Quick-Look Analysis (QLA) System, described below.

On the other hand, additional complexity is also introduced due to several aspects: (a) several versions of a product can be generated, some of which are incomplete - this must be carefully managed and communicated; (b) data received with high latency may appear "missing" from the archive in completeness checks; (c) data products can and do evolve during the mission and the archive and documentation must carefully track these changes. Indeed it is expected that products evolve through several versions of the PDS information model during the mission lifetime.

Of course the operational nature of the archive is mainly for the benefit of the instrument teams – by the time data are publicly released, they should be properly technically and scientifically validated.

The Quick-Look Analysis system: One of the primary benefits of the operational archive is that near-real-time visualisation of the data products can be performed, for engineering, science and data-quality purposes. The Quick-Look Analysis (QLA) web application developed by the SGS includes generic functionality for checking pipeline logs, telecommand history, telemetry events etc. It also includes instrument-specific visualisations customised to the instrument team's requirements, and built on top of the PDS4 archive products. This allows SGS and instrument team users to rapidly check their instrument health and data quality and start to understand their science data.

Collaborative functions include sharing of specified data with the entire science team, and a Science Analysis Forum (SAF) which centralises and preserves scientific discussion. This combination is designed to enhance the collaborations necessary to successfully fulfil the science goals of BepiColombo, which requires a data-driven strategy for (re-)planning due to the short mission duration.

SAF discussions can be initiated directly from the QLA by choosing a region of interest in a plot, image, spectrum or other visualisation. The forum post will automatically include a snapshot of the data visualisation, a link back to the QLA, and a set of derived metadata including the ROI, spatial and temporal coordinates etc. This enables the scientific discussion to be placed into context and allows search in both time and space.

Archive data validation: There are several ways in which the archive products are validated, at different points in their life-cycle. To begin, all generated products are checked for technical validity using the SGS-developed PDS4 Packager. This wraps the NASA

PDS validate tool along with pre-configured dictionaries, enforces the dictionary stack used for each product, and adds additional mission-specific checks (for example – do the product start time and the specified mission phase correspond to each other?).

The second level of validation is performed by the SGS and instrument teams who regularly check the key housekeeping and science data via the QLA. This allows a human inspection of both data quality, completeness and scientific validation. Although this is not performed completely systematically, it provides a powerful way to find anomalies and interesting events. When technical issues are found – which could be related to the instrument or the data processing – issue reports are raised and tracked to resolution.

Completeness checks are also performed periodically. Because the SGS hosts a copy of all telemetry packets, ingested into a database, it is straightforward to answer questions such as “do we have PDS4 products for each date when science packets were received”.

More detailed checks, taking into account the telecommand history and knowledge of the instrument operations are being considered. These would compare telecommand parameters and instrument modes to generated products to ensure that the correct number and type of products are produced. Figure 1 shows an example of a dashboard combining shared plots from various instruments active during the Earth flyby in April 2020. Since these data are plotted directly from archive products, immediate scientific validation is possible.

Currently it is planned to add more instrument data visualisations and a first version of the SAF during 2021, when the second Venus and first Mercury flybys will take place.

References:

[1] Besse, S. et al. (2018) Planet. Space Sci. 150, 131-140.

Figure 1: A snapshot of a QLA dashboard showing interactive plots of data acquired during the BepiColombo Earth flyby. All data are read “on the fly” from PDS4 products.