esac

European Space Astronomy Centre
Camino Bajo del Castillo s/n
Urb. Villafranca del Castillo
28692 Villanueva de la Cañada
Madrid
Spain
Tel. (34) 91 813 1100
Fax (34) 91 813 1139
www.esa.int

PSA ARCHIVING GUIDE FOR EXTERNAL DATA PRODUCERS

Prepared by **Diego Fraga Agudo, Daniela Coia, Emmanuel Grotheer**
Reference **ESDC-PSA-TN-0009**
Issue / Revision **1 / 1**
Date of Issue **08/07/2019**
Status **READY**
Document Type **TN**
Distribution **-**

European Space Agency
Agence spatiale européenne

Original - https://issues.cosmos.esa.int/socciwiki/display/PSA/ESDC-PSA-TN-0009+PSA+Archiving+Guide+for+External+Data+Producers

APPROVAL

PSA Archiving Guide for External Data Producers	
Issue 1	**Revision** 1
Author Diego Fraga Agudo, Daniela Coia, Emmanuel Grotheer	**Date** 08-07-2019
Approved by	**Date**
B. Merin — Bruno Merín	21/02/3030
S. Besse — Sebastien Besse	10/01/2020

CHANGE LOG

Reason for change	Issue	Revision	Date
Ready for review	1	0	03/11/2018
Updates based on PSA User-Group comments	1	1	03/03/2019
Updates about the Guest Storage Facility	1	1	08/07/2019

CHANGE RECORD

Issue 1	**Revision** 1		
Reason for change	**Date**	**Pages**	**Paragraph(s)**
First version	03/11/2018	All	All
Updates based on PSA User-Group comments	03/03/2019	multiple	2.3 ; etc
Updates about the Guest Storage Facility	08/07/2019	7-8	2.4

Table of contents:

1. INTRODUCTION

1.1 Purpose and Scope

The current document gives guidelines for the preparation and submission of data to the Planetary Science Archive (PSA) directed to potential external data producers not familiar with the PSA.

Within this document "**external data producer**" refers to a data producer who is not directly involved in the relative instrument team of the experiment generating the initial data. The term will also apply to producers that might be directly involved in instrument teams but that generate higher level data, i.e. beyond those standard mission products, usually raw and calibrated, to be delivered to the PSA for the mere reason of participating in an ESA (European Space Agency) mission.

For producers directly involved in ESA missions producing standard mission data, parts of this document are not applicable.

1.2 Applicable Documents

Ref.	Document	Version
AD1	PDS3 Planetary Data System Standards Reference, JPL D-7669, Part 2	V 3.8, February 2009
AD2	PDS4 Planetary Data System Standards Reference, JPL D-7669, Part 2	V 1.8.0, March 2017 or latest
AD3	PSA PDS3 Archiving Guide, ESDC-PSA-TN-0009	V 1.0 or latest version
AD4	PSA PDS4 Archiving Guide, ESDC-PSA-TN-0002	V 2.0, August 2017 or latest
AD5	PSA Product User Guide template, ESDC-PSA-TPL-0001	V 1.0, November 2018

Table 1 Applicable documents

1.3 Reference Documents

Ref.	Document	Version
RD1	Abbreviations and Acronyms, ESDC-PSA-TN-0003	Latest version
RD2	PDS4 Concepts	V 1.8.0 (6 of April 2017) or latest
RD3	The PDS4 Data Provider's Handbook	V 1.7.0 (April 2017) or latest
RD4	A.J. Macfarlane, *Improving accessibility and discovery of ESA planetary data through the new planetary science archive*, Planetary and Space Science, 2018	Versionless
RD5	S. Besse, et al., ESA's Planetary Science Archive: Preserve and Present Reliable Scientific Data Sets, Planetary and Space Science, 2018	Versionless

Table 2 Applicable documents

1.4 Abbreviations and Acronyms

See document [RD1].

2. THE PLANETARY SCIENCE ARCHIVE (PSA)

2.1 PSA Introduction

The Planetary Science Archive (PSA) is the central repository of the European Space Agency (ESA) for scientific and engineering data returned by ESA's planetary missions or multi-agency missions in which ESA participates. Products from multi-agency missions in cooperation with partners like for example NASA, JAXA or Roscosmos may be archived by the agency leading the mission, although eventually archived or duplicated in the PSA.

Currently, the PSA archives raw, calibrated and higher level data. and is part of, and physically located at, the European Space Astronomy Centre (ESAC).

The main goals of the PSA are:

- Preserve in the long term the data collected by ESA planetary missions
- Ensure the data's short and long term usability by means of standardization, documentation and simple file formats
- Provide open and free access to all published data

The web address of the PSA is psa.esa.int, where it is possible to find an online graphical user interface with search capabilities, documentation and a link to an FTP repository.

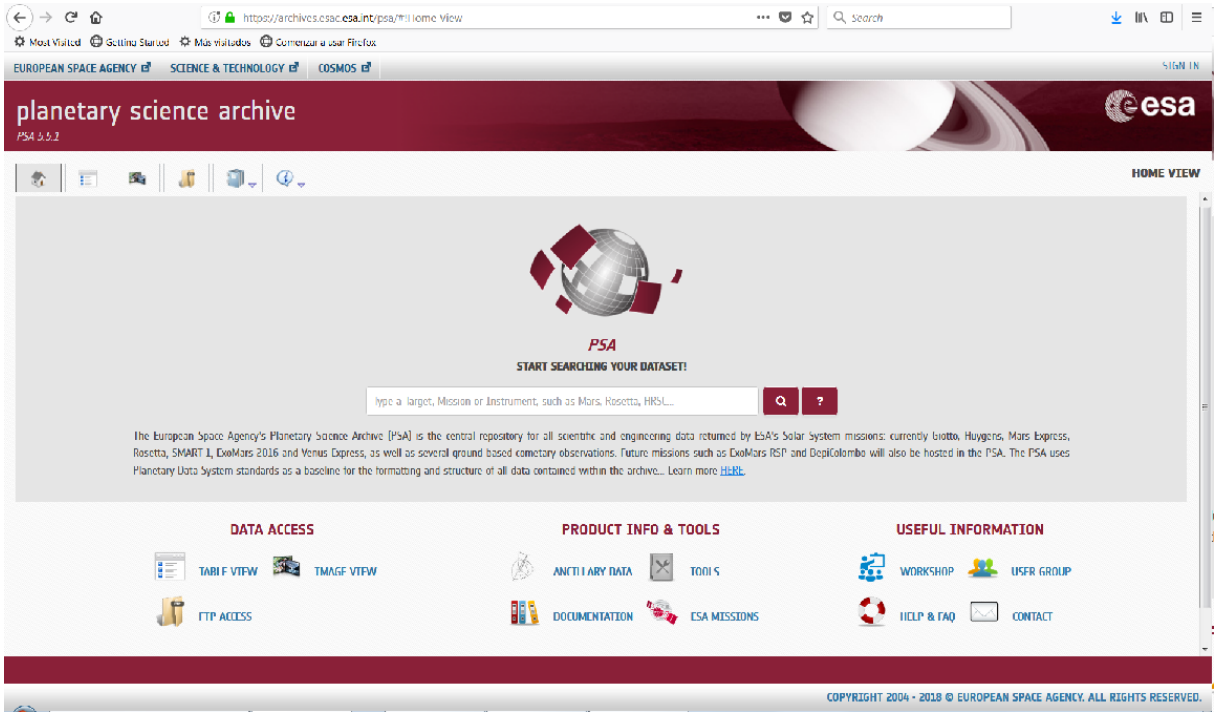

Figure 1 Planetary Science Archive web interface

The search interface is adequate for searching and downloading specific products. It allows users to query via different filters such as mission, target, instrument type, instrument, acquisition time, wavelength, or processing level. Using the CQL query language it is possible to do more complex searches filtering by a pre-selected set of metadata.

For bulk downloads it is more appropriate to use an FTP client connected to the public PSA FTP server using the following address and credentials (See Table 3).

PSA public FTP	
Host	ftp://psa.esac.esa.int
User	anonymous
Password	Not needed
Directory	/pub/mirror

Table 3 PSA public FTP credential

It is also possible to browse the FTP server using a web browser.

Apart from the unlimited free access to the data through an online GUI (Graphical User Interface) or FTP, other services and tools that are provided to the science community are:

- A help desk (reachable through the contact form in the web)
- Documentation
- Machine access to the data through an API including two protocols:
  - PDAP (Planetary Data Access Protocol)
  - EPN-TAP (EuroPlaNet - Table Access Protocol)
- GIS (Geographic Information System) capabilities. This is under analysis and will be available in the future

More about the PSA can be read in documents [RD4] and [RD5].

2.2 Data Accepted from External Producers

Apart from data coming from instrument teams directly involved in ESA's missions or missions with ESA participation, PSA archives some scientific data from "external producers".

Acceptable types of data include:

- Scientific data relevant for missions archived in the PSA such as professional and amateur Earth-based observations supporting the analysis of mission data. A more specific example are data from Earth-based telescopes observing comet 67P, which was the target of the Rosetta mission.
- High level science products: Products derived from scientific data that provide additional scientific value to mission data sets and are relevant to other scientists. An example of this would be mineralogical maps of a planet.
- Recalibration of already archived products sets demonstrating a better or alternative calibration or data processing.

The proposed scientific data must be relevant to the science exploitation of planetary ESA missions and must be observational data or derived from it. The submission of results from models that are not based on observational data is not encouraged.

Please contact the help desk if you think you have relevant products according to the criteria explained above.

PSA will provide consultancy support to prepare and submit the data and appropriate credits will be given to the data producer.

2.3 What is the PSA Expecting from the Data Producer

We expect the following from the data producers:

- A contact point: The name, email and contact details of the person in charge of interacting with the PSA.
- Time: You will need some time dedicated to prepare the data set, interact with PSA and potentially modify the data set before the release.
- Usually we expect a PDS (Planetary Data System) compliant data set: It is not enough to provide the data, we need a PDS compliant data set, this means:
    - The data files have to be in a PDS compliant format
    - The data must be documented in a PDS compliant way
    - A directory structure compliant with PDS
      For more information on the PDS standards see chapter 3 "PDS Standards Introduction".
- Usually we expect you to follow the PSA guidelines: PSA establishes its own guidelines within the PDS framework. These guidelines can be found in [AD3] for PDS3 and [AD4] for PDS4.
- Raw and calibrated data: Usually it is expected to receive both raw and calibrated data. Nevertheless for high level data products there can be exceptions where the raw data is not provided (a clear case would be if the high level products are derived from data already in the PSA archive).

In a case by case basis, data not conforming to the PDS standards is also accepted. See section 2.4 Guest Storage Facility (GSF) for more information.

In all cases the dataset must be sufficiently documented, complete and self-contained to allow for scientific exploitation without the need for interaction with the instrument team, obtaining additional calibration software (excluding general purpose data analysis software) or calibration data. All the support data and software should be included.

2.4 Guest Storage Facility (GSF)

Delivering in a PDS compliant form has benefits such as taking full advantage of the PSA services. It also standardises how data is documented and structured, sets a common metadata framework and makes the products readable by PDS aware software.

However, the PSA recognises that PDS rules might be an obstacle to deliver data to the PSA. For that reason the Guest Storage Facility (GSF) has been established. The GSF accepts data that is not PDS compliant offering more flexibility to the data producers. The only rule that PSA will enforce for data to be stored at the PSA is the existence of a Product User Guide (PUG) [AD5].
When needed, the PSA-UG will be conducted to evaluate the relevance of the dataset for the GSF.

The GSF hosts high level products derived from ESA missions data and other complementary data relevant for ESA missions. The standard mission data, defined here as the data that all missions must deliver to the PSA (usually raw and calibrated data) cannot be hosted in the GSF but shall be archived instead in the PSA in the normal way and following PDS standard.

The GSF at its current form takes the shape of an FTP location. Data in the GSF are not available in the PSA web interface and cannot benefit from all the services of the PSA.

The link to the GSF FTP is: ftp://npsa01.esac.esa.int/pub/mirror/Guest-Storage-Facility/
The link to the GSF general pages is: https://www.cosmos.esa.int/web/psa/psa_gsf

Figure 2 shows the split of the PSA FTP into ESA Missions FTP where the data is fully PDS compliant and data in the GSF where the data is not (or not fully) PDS compliant.

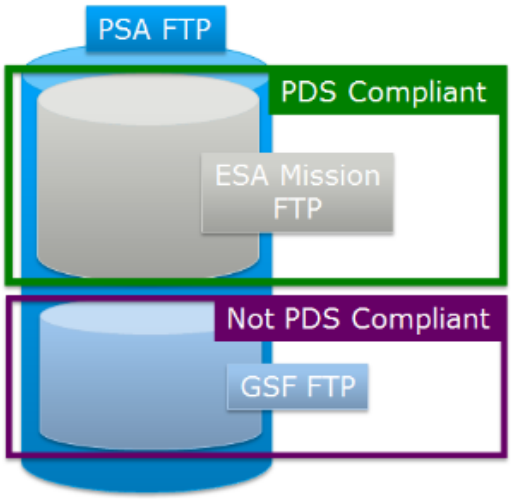

Figure 2 ESA Missions FTP Vs. GSF FTP

2.5 Life Cycle of Data Sets before Public Release

The typical steps to release data to the PSA is summarised below:

1. Initially, the potential data producer can contact the PSA team, or the PSA team could also approach the potential data producer. PSA will consult the PSA user group to evaluate the relevance of the dataset for the archive.
2. The data producer and the PSA team will then discuss if the data are of interest for the PSA and some details about documentation, format etc.
3. The data producer will deliver the data to the PSA by FTP (credentials will be provided)
4. PSA will analyse the delivery following two types of validation:
    a. A validation against the PDS Standards and PSA guidelines (unless they are waived in a case by case basis)
    b. A scientific validation, which can be done using one of the following two methods:
        i. An external peer review by expert scientists (organised by PSA).
        ii. A peer reviewed publication, where the production and calibration of the data are discussed, can be accepted as a proof of sufficient scientific quality.
5. Depending on the results of the previous steps, additional iterations between PSA and the data producer might be needed.
6. Re-delivery of the products might be needed in some cases.
7. The data set is ingested into the PSA as protected data (non public). The producer can review how the data are presented in the PSA.
8. After obtaining the permission from the producer the data set is released publicly.

Steps 2 to 4 aim at lasting no more than two months and steps 6 to 8 aim at lasting not more than one month. This timeline is indicative; it could be shorter or longer depending on the PSA resources and the availability of the data producer.

2.6 PSA Contact Information

There is a "Contact Us" section in the PSA web site (psa.esa.int) with updated contact information.

The PSA User-Group is reachable via psa-ug@sciops.esa.int.

By e-mail PSA is reachable via psahelp@cosmos.esa.int.

The postal address is:

Planetary Science Archive
European Space Astronomy Centre (ESAC)
Camino Bajo del Castillo s/n
Urb. Villafranca del Castillo
28692 Villanueva de la Cañada, Madrid

3. PDS STANDARDS INTRODUCTION

3.1 First introduction to the PDS Standards

PDS can have two meanings:

1. An organization linked to NASA that archives NASA mission data. It is the PSA equivalent in the USA.
2. The standards promoted by PDS.

We use the term PDS in this document with the second meaning.

The PDS standards define among other things:

1. A framework for the distribution of files in folders (the file structure of the archive).
2. A way to describe the format of the data providing considerable flexibility for the actual data format.
3. A way to document the data using metadata (labels).

Regarding point 2, it is important to note that PDS, in many senses, is not a data format. Indeed many formats are accepted as long as the format is properly declared using metadata in a PDS defined way. Any format is acceptable as long as the internal structure of the file can be fully described using the PDS rules. For example the data can consist of files containing ASCII tables where each row is a different measurement and each column a different physical parameter. PDS does not specify if the columns of this table have to be of a fixed or variable width or comma separated or use other separator or even no separators, but it does define a set of metadata where parameters such as the number of rows, whether it is comma separated or fixed-width etc. are defined. This metadata language defined by PDS is very flexible so that it can describe most formats defined by the producers.

PDS defines an "Observational Product" as one or more data file(s) accompanied by one PDS "label" with metadata. The label can be "detached", i.e. stored in its own file or "attached", i.e. included in the same file as the data. The recommendation is to always use detached labels for observational products.

The label contains metadata that can be classified in three categories attending to its purpose:

- Metadata to declare (define) the format of the data:

  The metadata in the label fully describes in a PDS specific way the format of the data so that the data is readable by anyone or any software designed to parse the label file. Rather than defining a format, PDS standards define the way in which the format is declared. For any binary or ASCII data stored in any format, such as tables or arrays, it will likely be possible to declare the format using PDS rules. If this is the case, one can create the label file to convert the data into PDS compliant data.
  For example these metadata could indicate that the data is an image represented as a binary array with 1024 rows and 512 columns. In PDS3 (PDS version 3) the number of rows and columns would be indicated as *LINES = 1024* and *LINE_SAMPLES = 512*.

- Metadata to document and contextualize the data:

Labels include metadata not only to define the format of the data but also to document the data, including for example textual descriptions of the context, acquisition and processing.
For example in PDS3 a short description of a data set is given as in this example: DATA_SET_TERSE_DESC = "Data from the Micro-Imaging Dust Analysis System of the ROSETTA Orbiter collected in the ROSETTA EXTENSION 2 mission phase."

- Metadata to enable search capabilities:

  Labels include metadata that are needed to document the data as described above, but some metadata can also make the data searchable via different interfaces. Examples include instrument name, mission, and acquisition date, etc.
  Observational products are not the only files to need a label, most files (e.g. documentation) also require labels.
  PDS standards not only define the requirements for the products as summarised above, they also set requirements for the documentation that must be provided. Additionally a framework is defined for structuring files into folders, the naming conventions, unique identification of items and versioning.
  Further constraints and guidelines compatible with the general standards are set at agency level, in this case ESA. These constraints are included in [AD3] for PDS3 and in [AD4] for PDS4 standards.

3.2 PDS3 vs PDS4

There are currently two well differentiated versions of the PDS Standards: PDS3 and PDS4.

PDS3 was the PDS version in use for many years thus many of the data found in the PSA currently are in this format. The last missions to use the PDS3 standard are Mars Express and Rosetta, whereas new missions use PDS4 (e.g., Exomars, BepiColombo). PDS4 is a major review of the standards with significant changes to adapt PDS to modern technologies, standards and practices.

One of the major changes is that PDS3 uses ODL language whereas PDS4 uses the more modern and widely recognized XML standard as the language for the label. Table 4 summarizes some other characteristics and differences between PDS3 and PDS4.

	PDS3	PDS4
Label language (metadata)	ODL (Object Description Language)	XML (eXtensible Markup Language)
File organization: top level structure	Volumes and data sets (*)	Bundles (*)
File organization: lower level structure	Directories within the data set (*)	Collections (*)
Text document encoding	7-bit ASCII PDF/A is accepted de facto though it is not in the standard.	UTF-8 PDF/A Note: 7-bit ASCII is a subset of UTF-8
Metadata terminology 1	"Objects"	"XML Tags" that represent PDS4 "Classes". The instances of the classes are PDS4 "Objects".
Metadata terminology 2	"Keywords"	"XML Tags" that represent PDS4 "Attributes".

Delivery mechanism	Originally designed for delivery in physical media (e.g. floppy disks or CDs), later used for online delivery.	Designed from the beginning for online delivery.
Versioning	At data set level	At file level

Table 4 PDS3 Vs PDS4 main differences

(*) There is not an exact one-to-one relation between the PDS3 concepts of volumes, data sets and directories and the PDS4 concepts of bundles and collections. For example in the implementation of PDS4 standards at ESA, there is one single bundle per mission whereas there could be many volumes and data sets per mission in PDS3.

3.3 Which PDS Version to use?

PDS4 is gradually replacing PDS3 as the standard for planetary data. In general the data provided to the PSA are recommended to be in PDS4.

For data that supports or complements data from the PDS3 missions, choosing PDS3 has the advantage of homogeneity with the existing data whereas PDS4 has the advantage of being the modern standard. In this case PSA accepts both PDS3 or PDS4 and it is left to the discretion of the producer to choose one over the other.

4. PDS3 SPECIFIC GUIDELINES

producers of PDS3 data must read the [AD3] *PSA PDS3 Archiving Guide*, which is the ESA guide for PDS3 targeting external data producers (the meaning of external in this case is explained in section 1).

The PDS3 standard is described in [AD1] *PDS3 Planetary Data System Standards Reference*. This document can be found on the PDS web site at https://pds.nasa.gov/tools/standards-reference.shtml

5. PDS4 SPECIFIC GUIDELINES

Producers of PDS4 data must read the [AD4] *PSA PDS4 Archiving Guide*. This document gives an introduction to the PDS4 standard and in addition is the document defining the PSA / ESA specific requirements and guidelines.

The following documentation provided by PDS is also available:

[RD2], *PDS4 Concepts*: A highly recommended document for getting started with the PDS4 standard.

[AD2], *PDS4 Planetary Data System Standards Reference*: This is the document defining the standard. It will make reference to the PDS4 Information Model (IM), which defines the PDS4 structure and to the PDS4 Data Dictionary Data Base (DDDB) that is the fundamental reference for definitions of classes and attributes. The IM and the DDDB are also part of the standard.

[RD3], *The PDS4 Data Provider's Handbook*: A cookbook guide with step-by-step instructions for developing an archive. It is written from the PDS / NASA perspective but it is mostly applicable to ESA.
All these PDS documents can be found at the PDS website at
https://pds.nasa.gov/pds4/about/portal.shtml

6. AVAILABLE TOOLS

There are a number of tools written in different languages that are useful for writing, reading, validating or visualizing PDS products or PDS data sets / bundles. Some are standalone programs whereas others are libraries to be used by programmers.

The PSA information about available tools can be found at:
https://www.cosmos.esa.int/web/psa/tools

The PDS information about available tools can be found at:

https://pds.nasa.gov/tools/about/

Finally SBN (the Small Bodies Node which is part of PDS) maintains a webpage with some tools designed by them. These tools are, unless otherwise specified in their page or documentation, applicable to any planetary data and not only to small bodies:
https://pdssbn.astro.umd.edu/tools/software.shtml

An important tool provided by PSA is PVV (PSA Volume Verifier). This command line software is used for validation of PDS3 data sets. All PDS3 data sets must be verified with this tool before submission to the PSA.

In turn, PDS4 data must be validated with the Validation Tool by PDS:
https://pds.nasa.gov/tools/about/validate/

It is worth mentioning that to generate PDS4 labels "manually" any standard XML editor can be helpful, such as for example Oxygen XML (www.oxygenxml.com) or EditiX (www.editix.com), just to mention two. For automatic generation of PDS4 label files there are libraries for all major programing languages to handle XML that can be integrated into processing software (pipelines).

END OF DOCUMENT