ESA's Planetary Science Archive (PSA): Ensuring The Long-term Usability Of Data

PV 2009 Conference,
1-3 December, ESAC

David J. Heather [1],

M. Barthelemy [2], N. Manaud [2], S. Martinez [2], H. Metselaar [2], M. Szumlas [2],

J. Kissi-Ameyaw [1], and the PSA Development Team [2]

[1] (ESA/ESTEC), [2] (ESA/ESAC)

dheather@rssd.esa.int

Outline

❑ **Definition and Aim - what is the PSA?**

❑ **Standards Used - The PDS**

❑ **Planetary archiving at ESA - PSA top-level processes**
- Archive Development
- The Review Cycle

❑ **Verification and validation procedures**
- The PSA Dictionary
- Checking syntax and consistency (PVV and PVS)

❑ **Linking to the User Interfaces**
- Query definition - the User Interfaces
- Searching and Retrieving Data

❑ **Current data and future Plans**

PSA Definition and Purpose

The _Planetary Science Archive_ (PSA) is the initiative, the setup, the process and the implementation to preserve data from ESA's spacecraft to planetary bodies, as well as supplementary information acquired in labo...

- **All PSA data are compliant with NASA's Planetary Data System (PDS) Standards**
  - This is the first step in ensuring that the data delivered to PSA are as widely usable as possible and will remain

Within the PDS Standards, data are organized into Volumes and data sets, collecting together observations of similar type, processing level, and/or from a specific mission phase or observation campaign.

JPL D-7669, Part 2

ROOT
— ARREADME.TXT
— ERRATA.TXT*
— VOLDESC.CAT

DOCUMENT | CATALOG | LABEL | SOFTWARE | CALIB | GEOMETRY | INDEX | DATA | EXTRAS

DOCINFO.TXT | CATINFO.TXT
CATALOG.CAT**
MISSION.CAT
INSTHOST.CAT
INST.CAT
DATASET.CAT
PERSON.CAT
REF.CAT | LABINFO.TXT
INCLUDE FILE 1
INCLUDE FILE 2 | SOFTINFO.TXT | CALINFO.TXT | GEOMINFO.TXT | INDXINFO.TXT
INDEX.LBL
INDEX.TAB | LABEL FILE 1
DATA FILE 1
LABEL FILE 2
DATA FILE 2
LABELED DATA FILE 1
LABELED DATA FILE 2
LABELED DATA FILE 3
INCLUDE FILE 1 *
INCLUDE FILE 2 * | EXTRINFO.TXT

xxxxINFO.TXT Required for each non-data subdirectory if present
* Optional
** Individual catalog files are preferred, or they may be combined in a single CATALOG.CAT file.

The PDS Standards

➤ **All PDS compatible data formats conform to a unified standard format incorporating documentation, calibration, and raw and processed data files.**

➤ **The format distinguishes:**

  ➤ Meta-data Files / Volume/Dataset Description Files

  ➤ **These files are included on the distribution medium instances and are parameters of the data collection, for example: the list file. The volume description file, detail the organization of the) data is files are called "catalog objects" or just "catalog" files.**

  ➤ **At the intermediate level one finds which describe the intended to be an entire, self-contained archive (for example a set of DVDs or a ZIP file. The volume description file, the type of archived (in terms of directories, etc.) and any additional documentation and software which might be included.**

```
VOLDESC.CAT

PDS_VERSION_ID            = PDS3
RECORD_TYPE               = STREAM
RELEASE_ID                = 0001
REVISION_ID               = 0004
OBJECT                    = VOLUME
  VOLUME_SERIES_NAME      = "MISSION TO MARS"
  VOLUME_SET_NAME         = "MARS EXPRESS ASPERA3 NPI DATA"
  VOLUME_SET_ID           = "SE_IRF_IRFK_MEXASP_3000"
  VOLUME_NAME             = "VOLUME 1: MARS EXPRESS ASPERA3
                             EDR NPI DATA"
  VOLUME_ID               = "MEXASP_3100"
  VOLUME_VERSION_ID       = "Version 1"
  PUBLICATION_DATE        = 2005-01-31
  VOLUMES                 = 1
  MEDIUM_TYPE             = "ONLINE"
  VOLUME_FORMAT           = "ISO-9660"
  DATA_SET_ID             = "MEX-M-ASPERA3-2-EDR-NPI-V1.0"
  DESCRIPTION             = "
Data from the Neutral Particle Imager (NPI) instrument of the
Analyzer of Space Plasmas and Energetic Atoms experiment, 3rd edition
(ASPERA-3), aboard the Mars Express spacecraft.  The ASPERA-3
Principal Investigator is Dr. Rickard Lundin, Swedish Institute of
Space Physics (IRF), Kiruna, Sweden.  The primary source for ASPERA-3
data is the ESA Planetary Science Archive (PSA)."

  OBJECT                    = DATA_PRODUCER
    INSTITUTION_NAME        = "SOUTHWEST RESEARCH INSTITUTE"
    FACILITY_NAME           = "N/A"
    FULL_NAME               = "N/A"
    ADDRESS_TEXT            = "6220 CULEBRA RD., SAN ANTONIO,
                               TX  78238"
  END_OBJECT                = DATA_PRODUCER

  OBJECT                    = CATALOG
    ^MISSION_CATALOG        = "MISSION.CAT"
    ^INSTRUMENT_HOST_CATALOG = "INSTHOST.CAT"
    ^INSTRUMENT_CATALOG     = "ASPERA3_INST.CAT"
    ^DATA_SET_CATALOG       = "ASPERA3_NPI_EDR_DS.CAT"
    ^PERSONNEL_CATALOG      = "PERSON.CAT"
    ^DATA_SET_RELEASE_CATALOG = "ASPERA3_NPI_RELEASE.CAT"
    ^REFERENCE_CATALOG      = "REF.CAT"
    ^SOFTWARE_CATALOG       = "ASPERA3_SOFTWARE.CAT"
  END_OBJECT                = CATALOG
END_OBJECT                = VOLUME
END
```

The PDS Standards

- **The lowest level is that of the data themselves.**
  - Each file will have a PDS label either appended to the beginning of the file or, more often, in a separate but proximate file. "One file, one label" is the general rule.

- **The LBL file must contain a full description of the data object(s)**
  - Allows for all PDS products to be read by existing standard software
  - Allows for long-term data usage with clear description of data to allow users to read files with their own software


```
PDS_VERSION_ID                = PDS3

/* FILE DATA ELEMENTS */
RECORD_TYPE                   = FIXED_LENGTH
RECORD_BYTES                  = 10420
FILE_RECORDS                  = 60291
LABEL_RECORDS                 = 2

/* POINTERS TO DATA OBJECTS */
^IMAGE_HEADER                 = 3
^IMAGE                        = 4

/* IDENTIFICATION DATA ELEMENTS */
FILE_NAME                     =
"H0887_0000_ND2.IMG"
DATA_SET_ID                   = "MEX-M-HRSC-3-
RDR-V2.0"
DETECTOR_ID                   =
MEX_HRSC_NADIR
...
PROCESSING_LEVEL_ID           = 2
RELEASE_ID                    = 0012
REVISION_ID                   = 0000
...
OBJECT                        = IMAGE
INTERCHANGE_FORMAT            = BINARY
LINES
                    = 60288
LINE_PREFIX_BYTES             = 68
LINE_SAMPLES                  = 5176
SAMPLE_TYPE                   = MSB_INTEGER
SAMPLE_BITS                   = 16
BANDS                         = 1
BAND_STORAGE_TYPE             =
BAND_SEQUENTIAL
MAXIMUM                       = 206
MEAN                          = 798813
MINIMUM                       = 23
STANDARD_DEVIATION            = 23.3757
END_OBJECT
```

Image reader or PDS reader

Planetary Archiving at ESA

❑ **Top-level processes and procedures followed by PSA**

- o Consultation during data product, documentation and data set development – typically through 'Feedback Document'

- o Independent reviews co-ordination and support

- o Handling the science data delivered by the PI teams, completing checks and validation on all data prior to ingestion
  - o Documentation checks
  - o PVV
  - o QBTool
  - o PVS
  - o Spot checks

- o Data sets proprietary access management after ingestion

❑ **The overall concept for archiving is documented in an official 'Archive Plan'**

❑ **Typically produced in line with the 'Archive Conventions' document that outlines all of the mission-wide keywords, values and concepts to be used by all instruments.**

- Written by the PSA in close coordination with the lead scientist
- All top-level aspects of the archive are defined; allow for inter-instrument data analyses.
- Instrument co-operatives can also be highlighted.

❑ **Once the top-level documents are delivered, the teams put together the Experimenter to Archive Interface Control Document (EAICD)**

- Details the archive for the given instrument.
- This is subject to rigorous internal validation. Once agreed internally, this document is placed in to the first step of the PSA's review procedure.
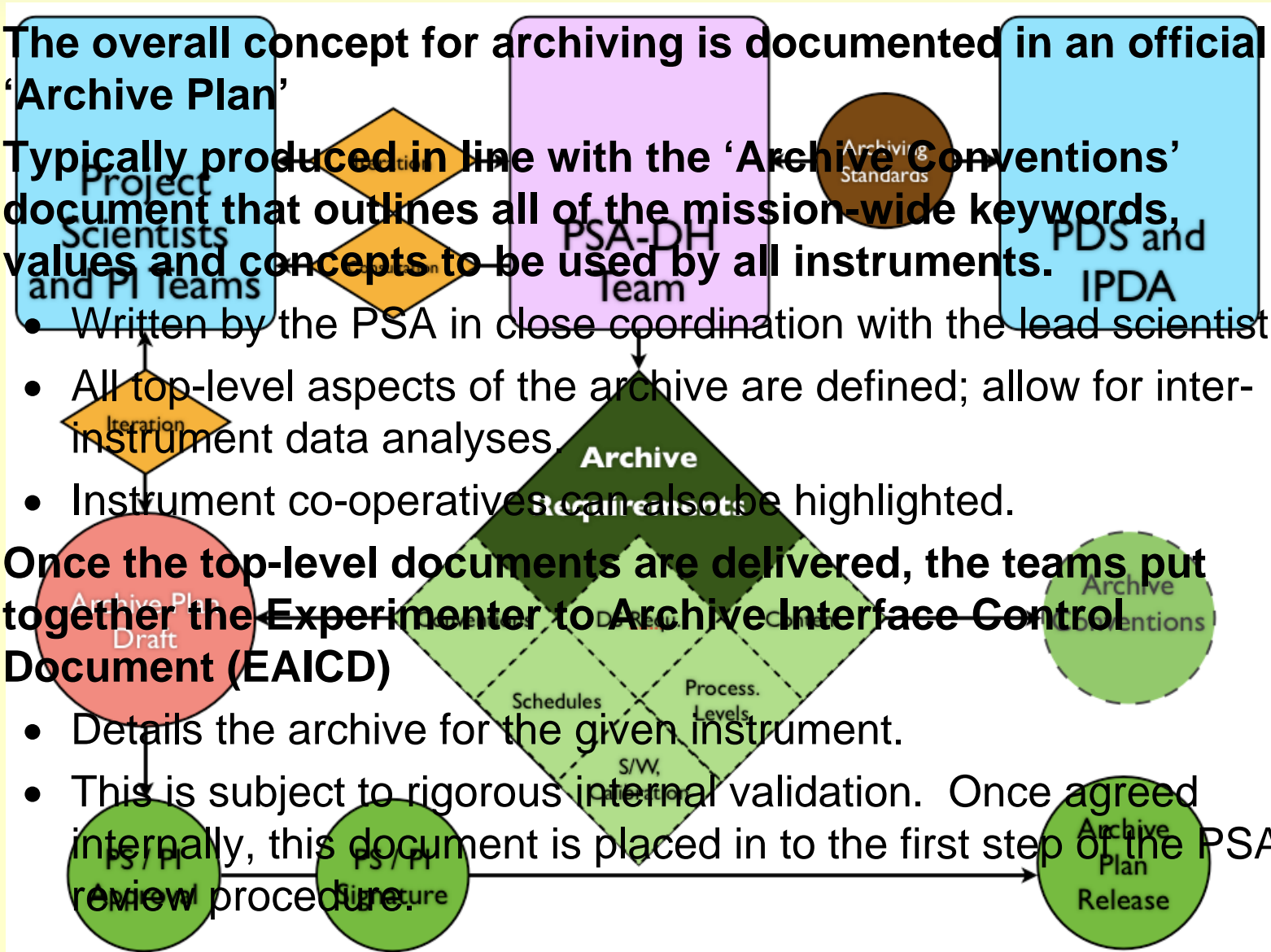
Review Lifecycle
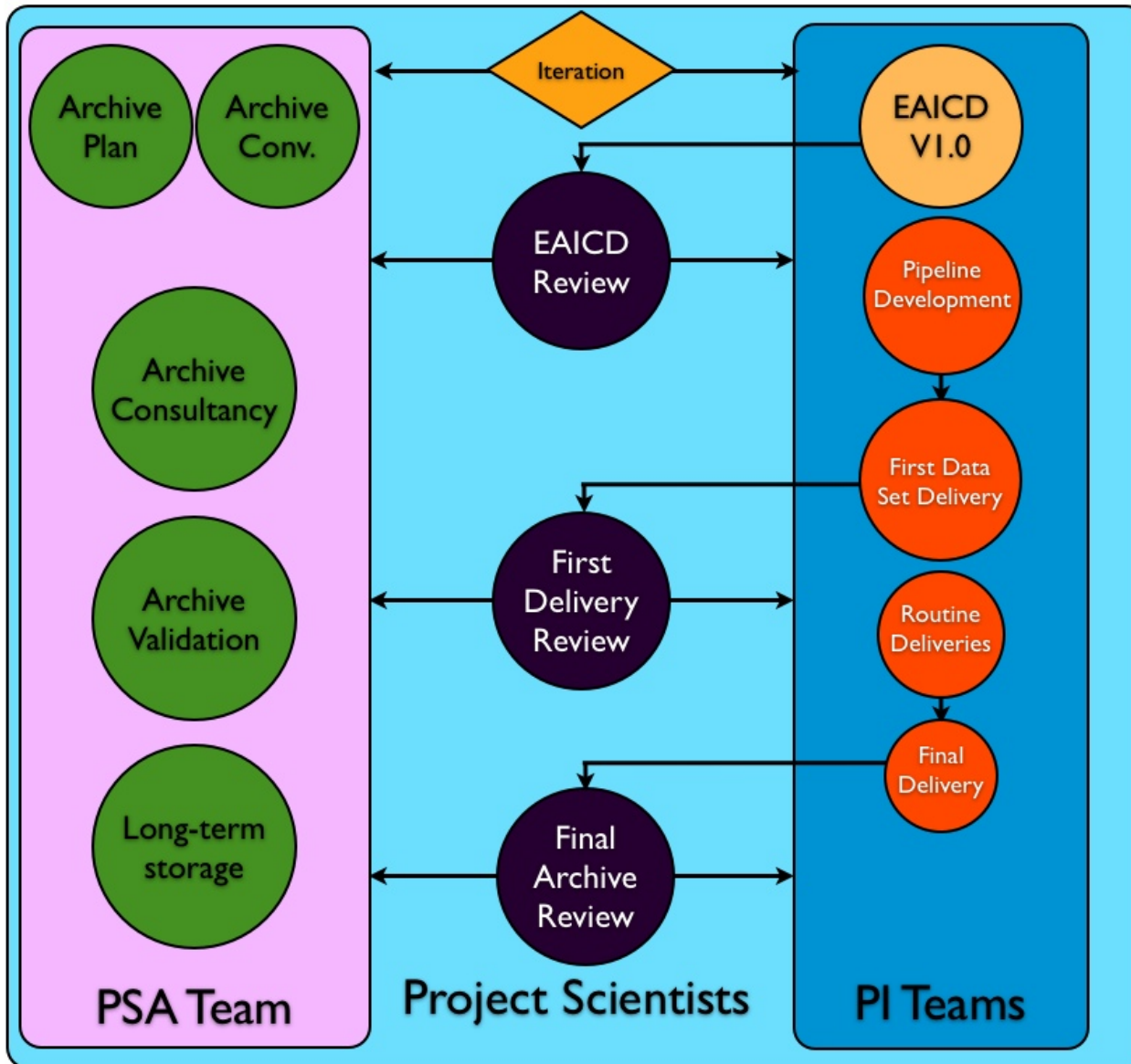
❑ **Typically, reviews have been split into three main phases for the planetary missions**

 o EAICD review: Internal review of documentation

 o First Delivery Review: Independent review of first data delivered for release to public.

 o For long 'multi-flyby' missions (e.g. Rosetta), reviews are held at each major milestone / flyby.

 o Final mission archive Review: Independent review of full mission archive, preparing for deep archiving.

❑ **Data are NOT released in PSA until the independent review is fully closed.**

Basic Data Set Lifecycle
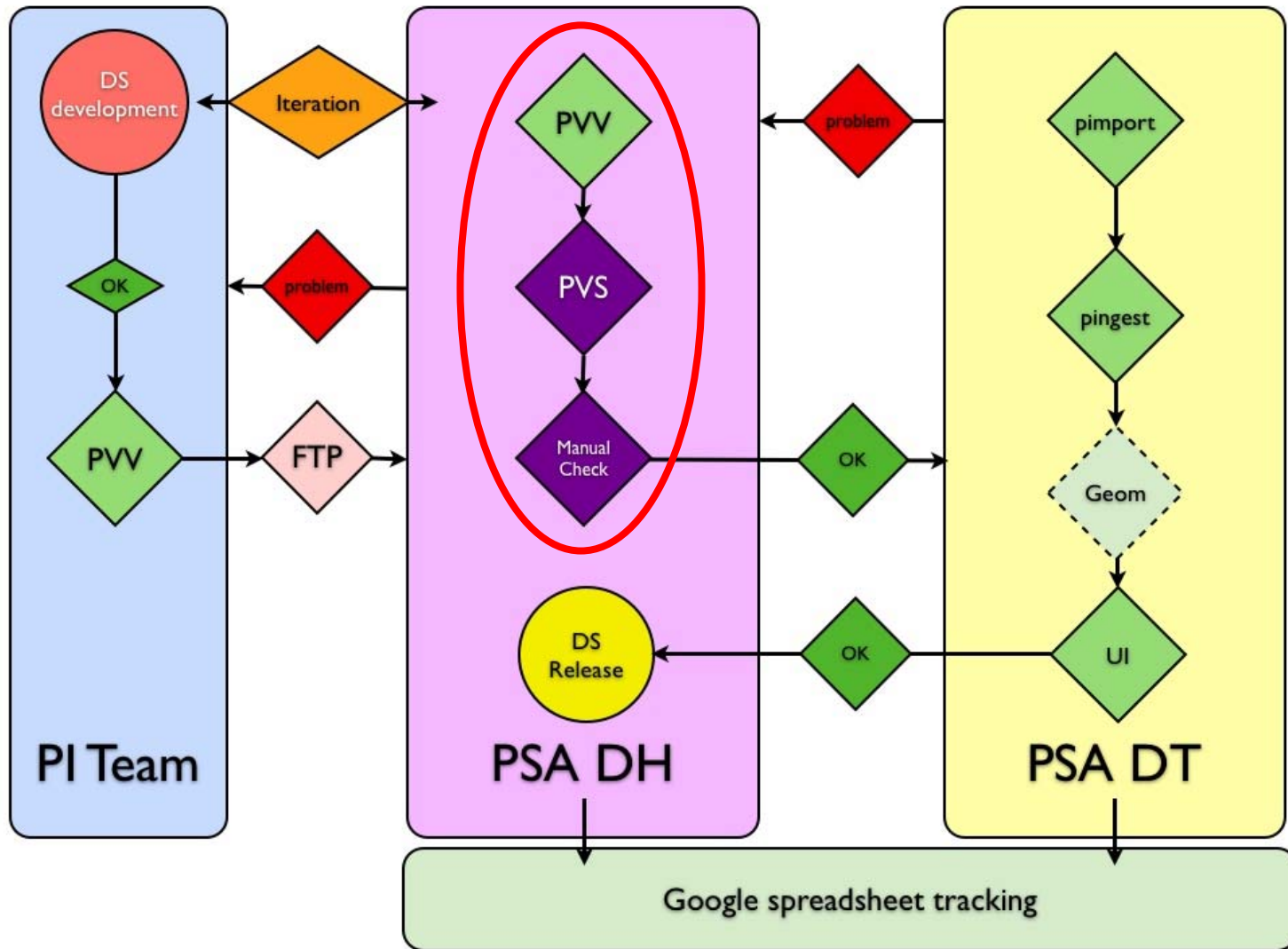
Standard Validation Tools /Checks

❑ **All data delivered to the PSA are subjected to a set of rigorous validation procedures.**

- *PVV Syntax validation:* ensures syntax and the presence of all required keywords. No data can be ingested without full PVV compliance.

- *PVS:* completes more qualitative checks and ensures keyword values are consistent with the data themselves.

- *Manual Checks:* a series of spot-checks are made on all data deliveries to ensure data can be read and are useable. These vary greatly between the different instruments delivering data.

❑ **The validation requirements are drawn from all archive documentation**

- These are input primarily via the 'PSA Dictionary' and are then verified automatically by the PSA Volume Verifier (PVV) and PSA Validation System (PVS) tools.

PSA Dictionary

- **Dictionary is at t**
- **PSA maintain the**
  - Built up from t
  - Mission specifi
- **'Labels' are defin**
  - Specify the rec
    archive, missio
  - PVV checks ag
  - Ensures that a
    level, all data l
- **Maintenance of t**
  - PSA and PDS
  - PSA work in cl
  - Regular mergir

PVV and PVS Validation

- **PSA Validation**
  - Check ... datasets ... this PSA ...
  - Syntax ... correlation of all
  - information ... PDS ... and date and
  - Trans... Runs...
  - Output ... that provide a
  - list of ... Intern...
  - Critical ... ATA.TXT file
- **Distribut... validate ...** practically when
  - instru... ... data sets before ...

```
[dheather@ssols02 MEX-M-HRSC-3-RDR-V2.0_NOW]$ pvv verify -Dwarn=off -DmaxMemory=2048m -Ddictionary=3.0 -Drelease=73

------------------------------------------------
                Verifying Entire DataSet
------------------------------------------------
***********************************************************
Checking Structure of Dataset
***********************************************************
Checking catalogs of Dataset
***********************************************************
[VERIFY ][E][02.038][0001] End of odl file reached, missing END statement
(CATALOG/REF.CAT, 0)
    Malformed odl file missing END keyword required to mark end of ODL
    part of file, end of file reached.
[VERIFY ][E][02.040][0002] Expected keyword missing for this file type
(CATALOG/DATASET.CAT, -1)
    Keyword definition for LABEL_REVISION_NOTE missing from
    CATALOG/DATASET.CAT. File =CATALOG/DATASET.CAT, line -1
[VERIFY ][E][02.040][0003] Expected keyword missing for this file type
(CATALOG/SOFT.CAT, -1)
    Keyword definition for LABEL_REVISION_NOTE missing from
    CATALOG/SOFT.CAT. File =CATALOG/SOFT.CAT, line -1
***********************************************************
Checking data products of Dataset
***********************************************************
***********************************************************
Checking documents of Dataset
***********************************************************
***********************************************************
Checking browse products of Dataset
***********************************************************
***********************************************************
Checking geometry index of Dataset
***********************************************************
***********************************************************
Checking software products of Dataset
***********************************************************


EXECUTION FAILED

Total time: 1 minute 37 seconds
```

ID	Status	Check			
		Total	Pending	Fail	Error
RELEASE_ID=0002, REVISION_ID=0000	not valid	224	0	24	0

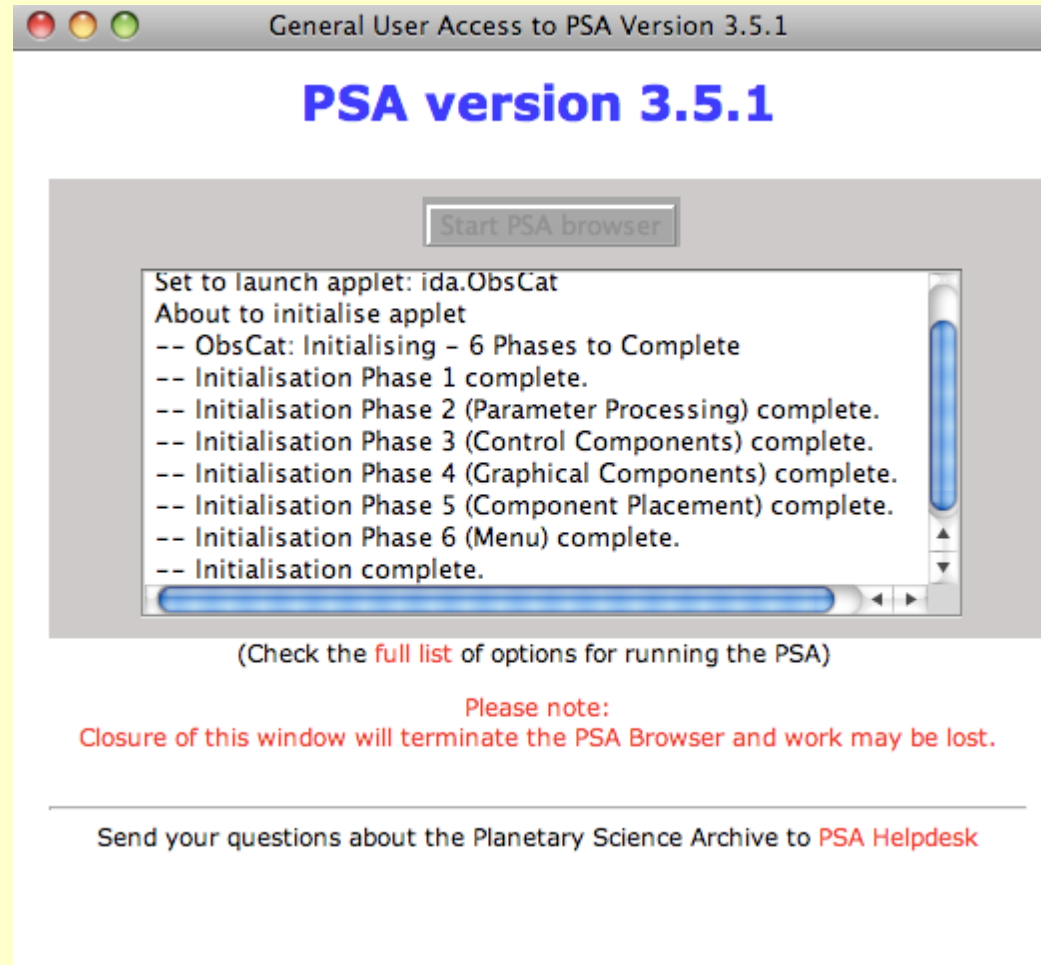# The User Interface Link

PSA is a client-server application. The user runs a JAVA application locally on her/his desktop.

http://www.rssd.esa.int/psa

❑ All parameters that can be queried must be available for ingestion
  • from a label or associated file
  • Query definition must be determined quite early
  • PSA ensure that all query parameters are included in the labels defined within the PSA Dictionary.

General User Access to PSA Version 3.5.1

PSA version 3.5.1

Start PSA browser

Set to launch applet: ida.ObsCat
About to initialise applet
-- ObsCat: Initialising – 6 Phases to Complete
-- Initialisation Phase 1 complete.
-- Initialisation Phase 2 (Parameter Processing) complete.
-- Initialisation Phase 3 (Control Components) complete.
-- Initialisation Phase 4 (Graphical Components) complete.
-- Initialisation Phase 5 (Component Placement) complete.
-- Initialisation Phase 6 (Menu) complete.
-- Initialisation complete.

(Check the full list of options for running the PSA)

Please note:
Closure of this window will terminate the PSA Browser and work may be lost.

Send your questions about the Planetary Science Archive to PSA Helpdesk

PSA Interface

PSA supports 4 different views:
- Query
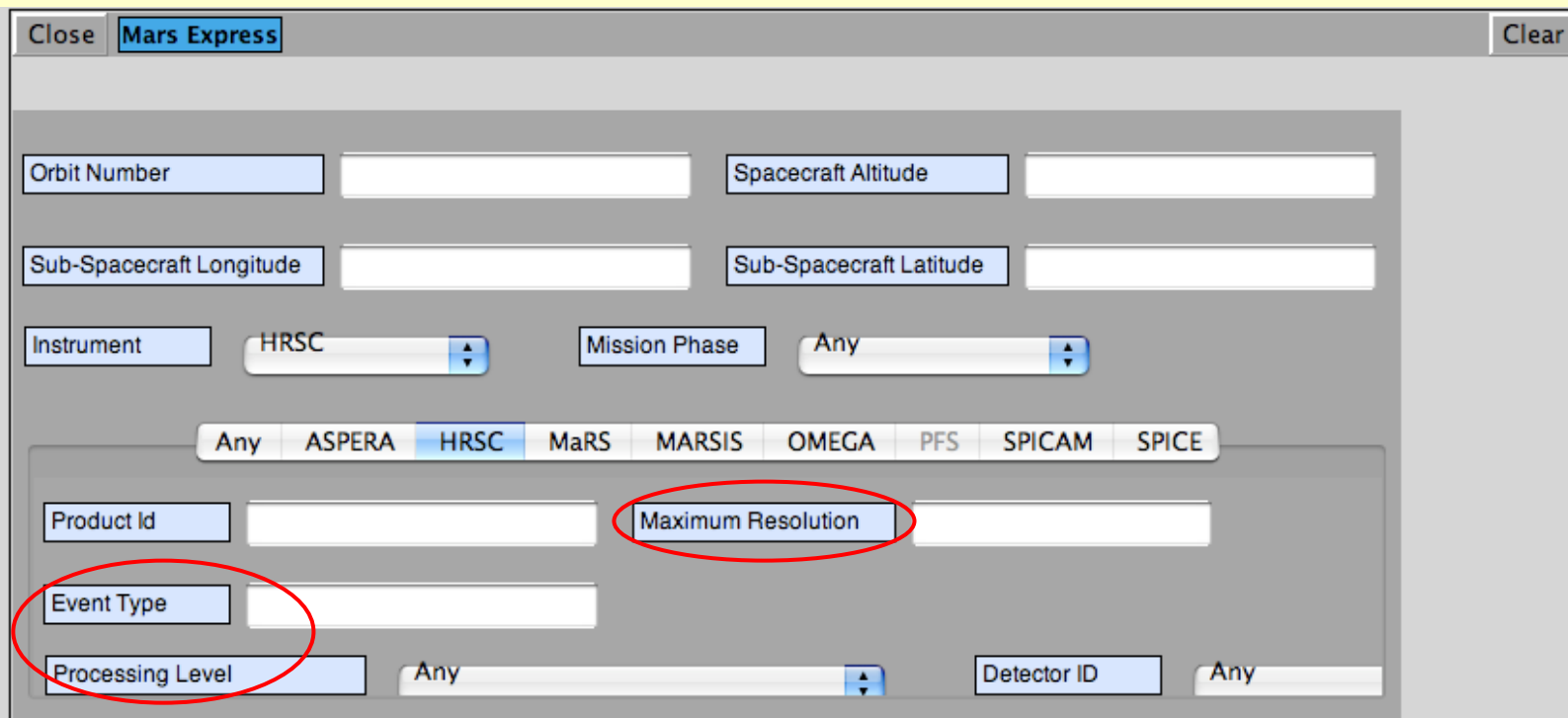- Latest Results
- Delivery Basket
- Login/Register
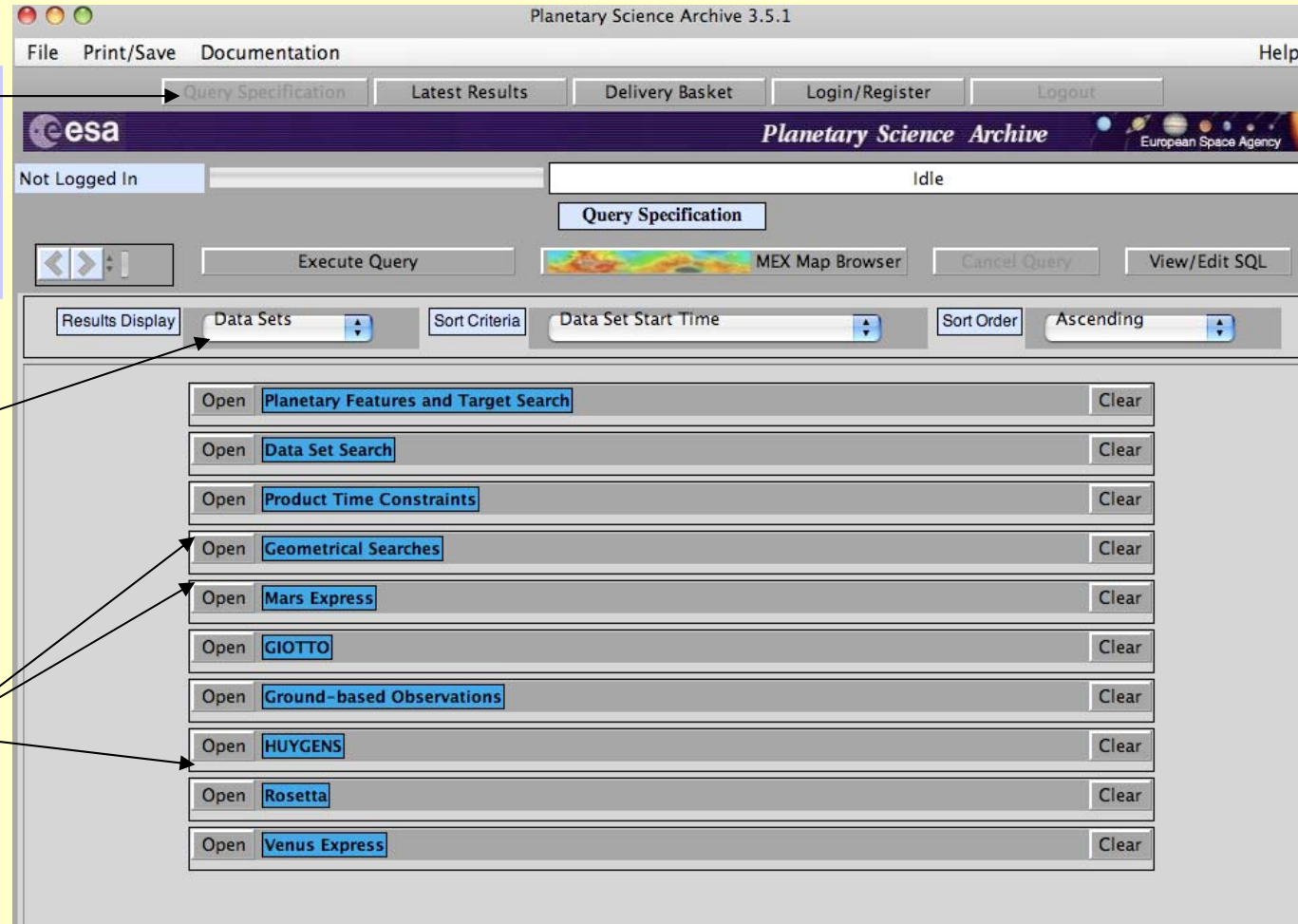
Selection of Result Display Options

Query Panel are ANDed when querying the database

PSA Interface
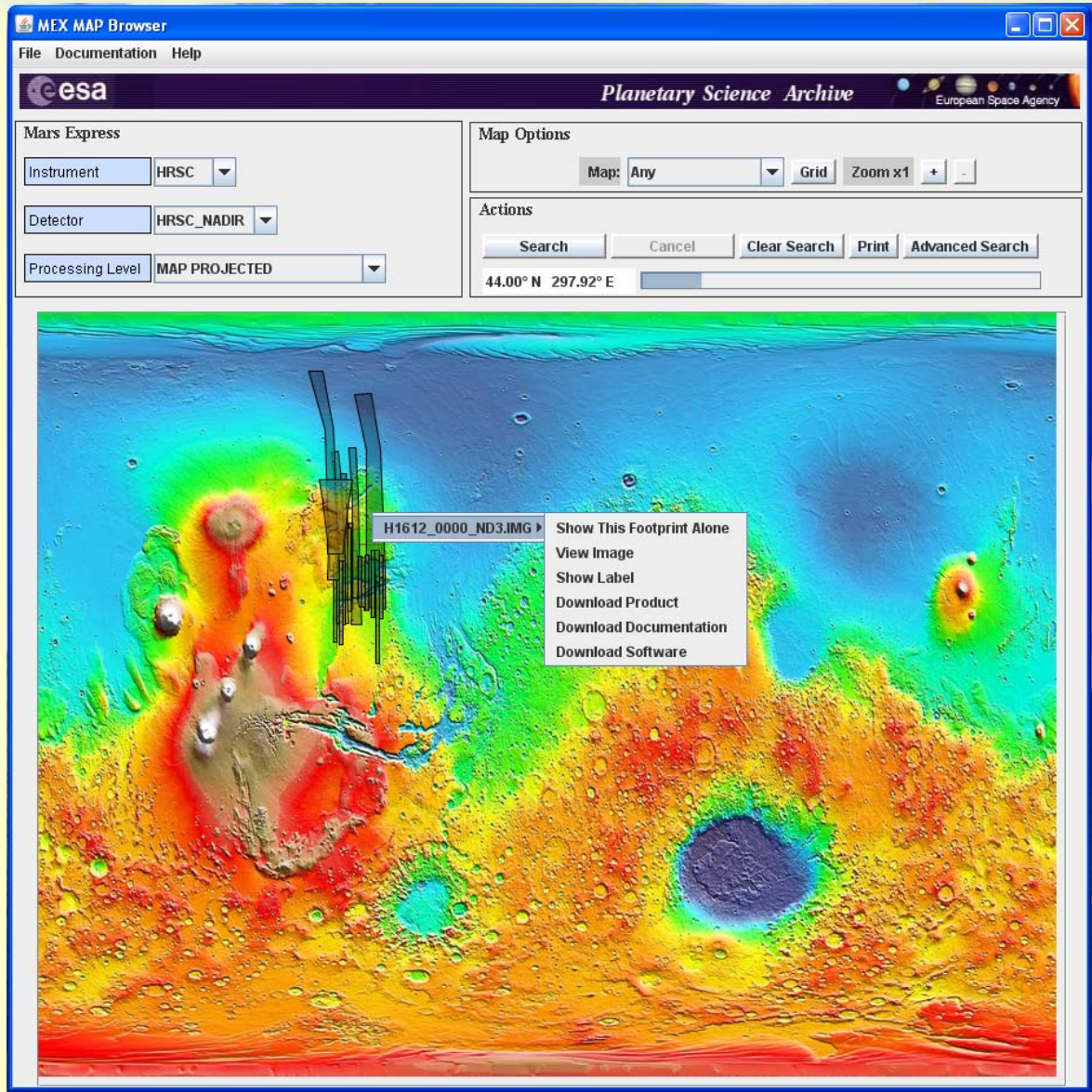
Mars Express Specific Panel:

- Individual instrument query details

- In this example, HRSC requires that the 'Maximum Resolution', 'Event Type', 'Processing Level ID' etc. keywords are present in labels or associated files

- This will be entered into the PSA Dictionary for the HRSC label.

PSA UI – The Map Interface

Search Button
Select Base Map Type
(MOLA / VIKING...)
Select Instrument /
1. Click and drag to
Detector
select your area of
interest

2. Press 'Search'
Zoom function
and wait for
Access to 'standard' UI
footprints
5. Transfer at any time
to the standard interface
3. Click on desired
for advanced searches.
footprint
Long / Lat information
Your search parameters
will be remembered.
4. Download
directly or view
6. Return to the map
label / docs etc.
browser, you can finish
your search as required.
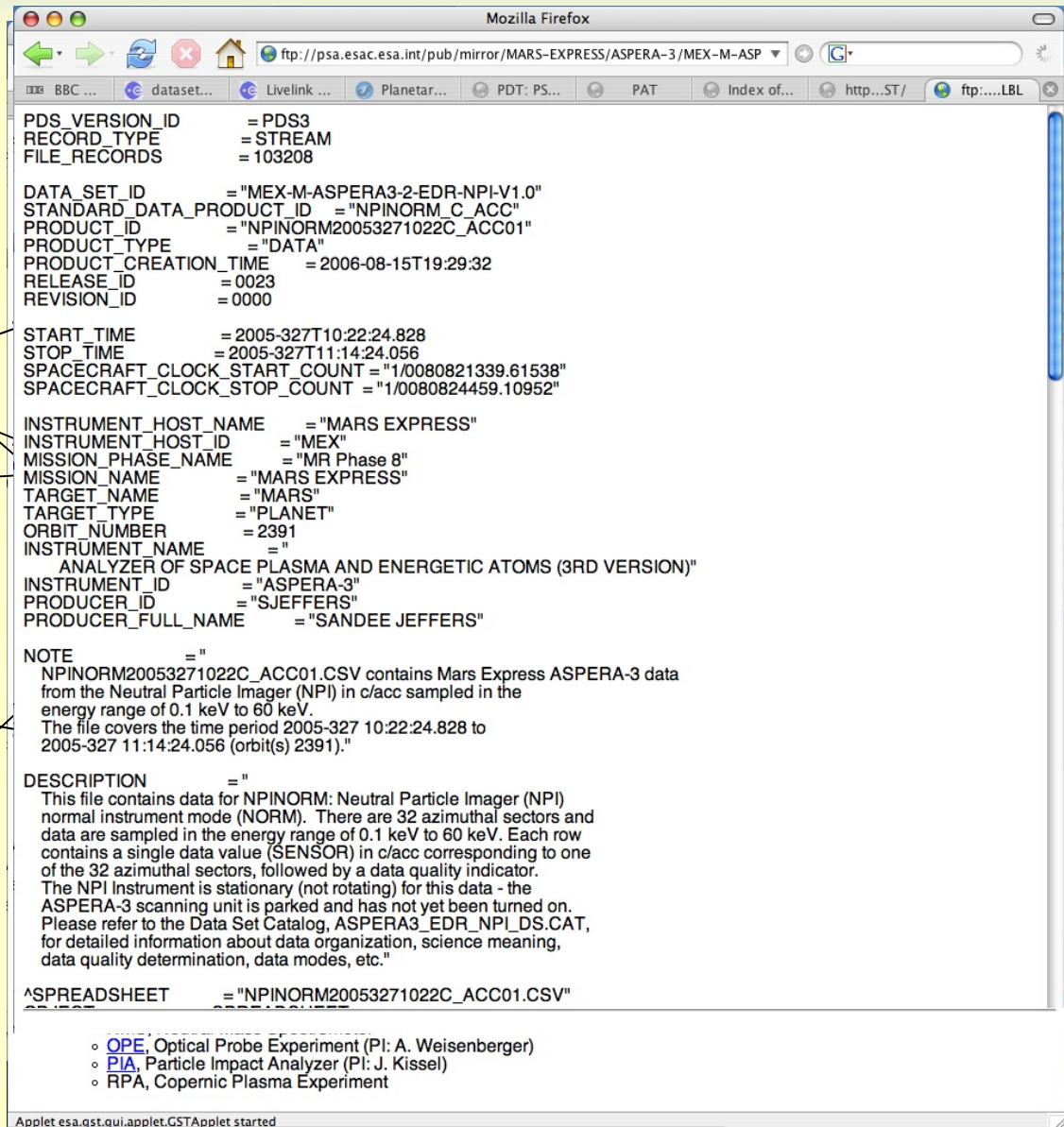
PSA Browser Interface

❑ **Opening Screen**

Click on the instrument that interests you

Select the data set you wish to look at

Browse through the directories to locate the files you want

Right click the product you want to save etc. to bring up the menu

Left click to view directly (where possible)

Mozilla Firefox

ftp://psa.esac.esa.int/pub/mirror/MARS-EXPRESS/ASPERA-3/MEX-M-ASP ▼

BBC ... | dataset... | Livelink ... | Planetar... | PDT: PS... | PAT | Index of... | http...ST/ | ftp:....LBL

```
PDS_VERSION_ID          = PDS3
RECORD_TYPE             = STREAM
FILE_RECORDS            = 103208

DATA_SET_ID             = "MEX-M-ASPERA3-2-EDR-NPI-V1.0"
STANDARD_DATA_PRODUCT_ID    = "NPINORM_C_ACC"
PRODUCT_ID              = "NPINORM20053271022C_ACC01"
PRODUCT_TYPE            = "DATA"
PRODUCT_CREATION_TIME   = 2006-08-15T19:29:32
RELEASE_ID              = 0023
REVISION_ID             = 0000

START_TIME              = 2005-327T10:22:24.828
STOP_TIME               = 2005-327T11:14:24.056
SPACECRAFT_CLOCK_START_COUNT = "1/0080821339.61538"
SPACECRAFT_CLOCK_STOP_COUNT  = "1/0080824459.10952"

INSTRUMENT_HOST_NAME    = "MARS EXPRESS"
INSTRUMENT_HOST_ID      = "MEX"
MISSION_PHASE_NAME      = "MR Phase 8"
MISSION_NAME            = "MARS EXPRESS"
TARGET_NAME             = "MARS"
TARGET_TYPE             = "PLANET"
ORBIT_NUMBER            = 2391
INSTRUMENT_NAME         = "
    ANALYZER OF SPACE PLASMA AND ENERGETIC ATOMS (3RD VERSION)"
INSTRUMENT_ID           = "ASPERA-3"
PRODUCER_ID             = "SJEFFERS"
PRODUCER_FULL_NAME      = "SANDEE JEFFERS"

NOTE                    = "
    NPINORM20053271022C_ACC01.CSV contains Mars Express ASPERA-3 data
    from the Neutral Particle Imager (NPI) in c/acc sampled in the
    energy range of 0.1 keV to 60 keV.
    The file covers the time period 2005-327 10:22:24.828 to
    2005-327 11:14:24.056 (orbit(s) 2391)."

DESCRIPTION             = "
    This file contains data for NPINORM: Neutral Particle Imager (NPI)
    normal instrument mode (NORM). There are 32 azimuthal sectors and
    data are sampled in the energy range of 0.1 keV to 60 keV. Each row
    contains a single data value (SENSOR) in c/acc corresponding to one
    of the 32 azimuthal sectors, followed by a data quality indicator.
    The NPI Instrument is stationary (not rotating) for this data - the
    ASPERA-3 scanning unit is parked and has not yet been turned on.
    Please refer to the Data Set Catalog, ASPERA3_EDR_NPI_DS.CAT,
    for detailed information about data organization, science meaning,
    data quality determination, data modes, etc."

^SPREADSHEET            = "NPINORM20053271022C_ACC01.CSV"
```
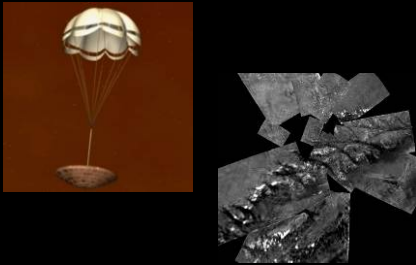
o OPE, Optical Probe Experiment (PI: A. Weisenberger)
o PIA, Particle Impact Analyzer (PI: J. Kissel)
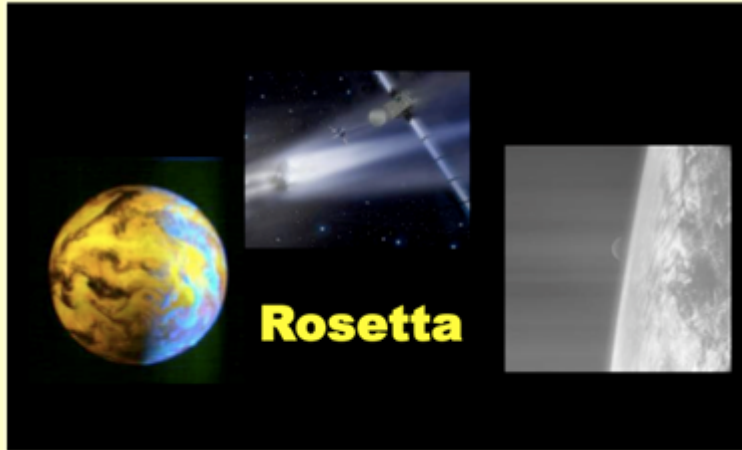o RPA, Copernic Plasma Experiment

Applet esa.gst.gui.applet.GSTApplet started

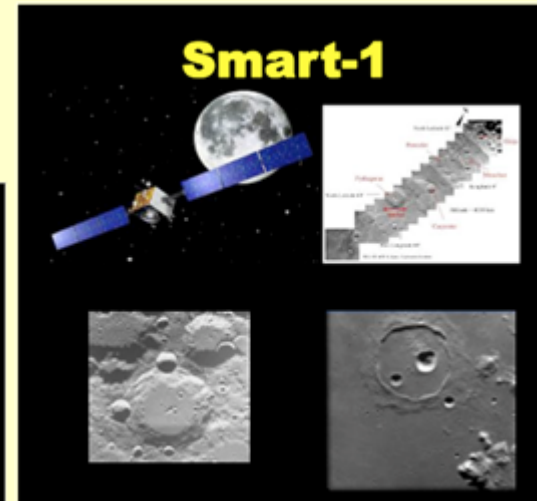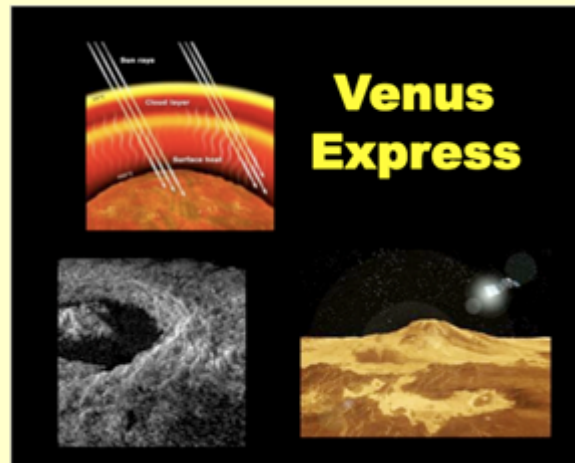# PSA : one archive, several missions

Huygens

Rosetta

Giotto

Mars Express
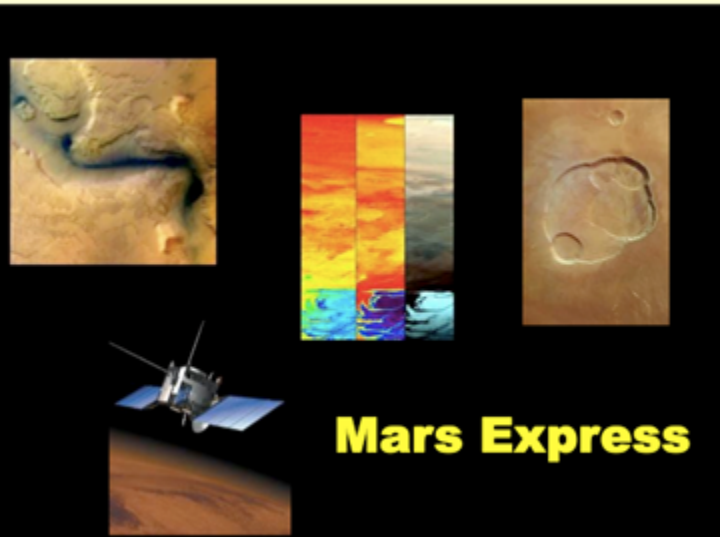
Venus Express

Smart-1

Current Status, Available Datasets

- ➢ GIOTTO data from comet Halley and Gripp-Skellerup
- ➢ Comet Halley ground-based observations (Halley-Watch)
- ➢ Comet Wirtanen ground-based observations
- ➢ Mars Express instrument and auxiliary data (ongoing)
- ➢ Venus Express instrument and auxiliary data (ongoing)
- ➢ Smart-1 instrument and auxiliary data (pending)
- ➢ Rosetta instrument and auxiliary data (ongoing , Peer Reviews in progress)
- ➢ Huygens (complete)
- ➢ BepiColombo data handling and archive support – in preparation
- ➢ ESA supported instruments on Chandrayaan 1 – pipeline development and support from PSA

IPDA and the Future

❑ **The PSA work very closely with experts at NASA's PDS as the Standards continue to develop.**

❑ **Archiving experts from all major countries involved in planetary exploration form the International Planetary Data Alliance (IPDA).**

- One of the main objectives of this group is to try to develop data and archives that are inter-operable.

- Validation of data to allow for this interoperability will be a future requirement in addition to the standard validations already discussed

❑ **The validation system is designed to ensures that all of ESA's planetary data will be available and useful to the community for many years after the mission has ended, and that they will be as compatible with as many other planetary archive data as possible.**

- Time will tell!...