PSA User Group Meeting 2023-02

Date: 15-16th February 2023

Meeting room: D052, ESAC

Virtual participation: Teams

Participants:

- Mark Bentley
- Beatriz Sanchez-Cano
- Daniela Coia
- Gianrico Filacchione
- Emmanuel Grotheer
- Tanya Lim
- Ernst Hauber
- Sonia Fornasier
- Antonio Genova
- Antonio Garcia Munoz
- Jan Soucek
- Paolo Tortora
- Andrea Nass
- Anni Maattanen
- Bruno Merin Martin
- Guido De Marchi

1. WEDNESDAY 15TH FEBRUARY

Time	Presenter and Title	Notes
08:30	Bus pickup outside Hotel Exe Moncloa	
09:30	Arrival, welcome and logistics	
09:35	Around the room (and online!) introductions	
10:00	Bruno Merín: Current status and future plans of the archives	- Beatriz notes that platform housekeeping can contain useful information, and we should try to archive/make public these data. - Ernst confirms that this was a resolution of the outgoing group - Mark comments that this is under discussion for missions also beyond planetary, but is hampered by issues of proprietary knowledge and insufficient documentation. - Ernst asks if ESA recognises the effort needed to maintain legacy mission data - Guido responds that they do, but there are budgetary constraints to consider, and that the User Group has a role to play in deciding priorities based on the community needs. - Mark comments that technically PSA is set up for longer term preservation of data from legacy missions, but that the real bottleneck is human resources for validating new data from legacy missions. - Beatriz asks if we can update e.g. errata files when we find issues with legacy data. Mark confirms that this is possible.
10:10	Ernst Hauber: Report of the outgoing user group	- Ernst presents the report of the outgoing group - the group tried to balance meetings at ESA with other institutes - Guido comments that we are constrained now to one physical meeting per year - but we can organise virtual meetings, and potentially combine with conferences etc. - Several mission and instrument-specific topics were discussed which corresponding reccomendations - in many cases these still need to be followed up on ☐ Mark Bentley to compile a list of these and update the new user group on status and possible implementation - The topic of platform HK was further discussed - especially important for the ExoMars rover where primary science can be done with HK - it was mentioned that ESA was trying to get funding to archive Chandrayaan-1 data

→ THE EUROPEAN SPACE AGENCY

		☐ Mark Bentley to find out what this entailed and where we are now
		• In discussion of the GSF it was commented that UPWARDS and PLANMAP data are not yet available
		o Mark adds that UPWARDS data were physically delivered, but need further processing (they were supposed to be in PDS4, but are clearly not).
		o Emmanuel adds that the delivery of UPWARDS was right at the end of the contract and it was not possible to follow up to make changes/improve the data afterwards
10:45	Guido De Marchi: The Archive Research Visitors Programme	• Guido presents the steady decrease in numbers of planetary applications for the ARVP
		o whilst the total numbers remain constant
		• Whilst this could simply be low number statistics, several possible solutions were raised to increase the number of applications
		• Beatriz wondered if we could promote this better through instrument teams
		• Andrea suggests we could propose some topics of interest, especially for lower level applications (although we would have to discuss how to rank these)
		• Antonio Genova wonders if we could go back to previous proposers to submit again after feedback
		o as long as we are not lowering the bar too much
		• Gianrico wonders if it is related to the cycles of planetary missions (cruise vs ops etc.)
		o are many of the people who might apply working on other things?
		• Ernst asks if we could extend this to master students
		o some of the possible issues with timing in a masters project were discussed
		• Guido adds that the application website proposes that a professor could also prepare an application for a student to take up
		o but are we then assessing the student or the professor?
		• Antonio Genova in the reporting (e.g. presentation) it would be interesting to add the name of the mission whose data is proposed to be used
11:15	Coffee	
11:45	Active mission archive reports • MEX • ExoMars TGO • BepiColombo	MEX • Mark asks if the ASPERA recommendations made by the outgoing UG are included in the EXPRO contracts mentioned
		o currently not all UG recommendations are included, but should be in a later CCN
		• Beatriz:MARSIS sub-surface data are being worked on at JPL and will be delivered to PDS
		☐ Emmanuel to check if these will also be delivered to PSA

		TGO
		    • Antonio asks if any radio science data from TGO will be archived
		    • Tanya replies that currently not - however, support can be given should this be necessary
		
		BepiColombo
		    • Discussion of what we mean by "GIS" - here we distinguish two things:
		        o map view (i.e. GIS functionality)
		        o GIS-ready products
		    • Gianrico raises some use cases for SIMBIO-SYS
		        o Mark says that these are covered by recent discussions on archiving of the target database and allowing search for products intersecting/enclosed by these targets
		    • Andrea comments that semantic meta-data are also important here

13:00	Mark Bentley: Intro to the PSA:	Mark reports on the current status of the PSA and future plans
13:30	Lunch	

15:00	Discussion - Future archives	JUICE
		    • pipelines are being prepared for those instruments active during NECP
		
		CLPR
		    • David reports on 3 flight opportunities for PROSPECT-related hardware
		        o 1 - full PROSPECT payload should fly end 2026/early 2027
		            ▪ European Payload, archive scientist support by David Heather
		            ▪ no formal agreement for funding of archive development, but not excluded (fully HRE)
		        o 2 - PITMS ion trap mass spectrometer (Open University) to fly on first Astrobotic mission
		            ▪ launch expected Q2 this year
		            ▪ very short operations expected (low archive data volume)
		        o 3 - JAXA would like to fly payload on LUPEX mission with ISRO
		            ▪ 2025-2026 timeframe
		            ▪ longer lifetime than 2 (rover mission)
		            ▪ JAXA is the archive authority here
		    • In general HME has money from the last ministerial for archiving, but unclear how they will spend this
		
		Comet Interceptor

- Mark comments that the archiving situation for CI is a bit different than for other planetary missions in that the data processing will be done by an external entity. The BUSOC will integrate and run pipelines to produce products, also higher levels ones fusing data from different instruments, and then deliver to PIs.
    - several people expressed some concern over this and care will be needed to ensure that any contracts are written in a way that lets us collaborate in a productive manner.

ExoMars rover
- Essentially we need to wait for the funding situation to be clarified to see when/how to resume archiving
- Mark comments that we may need to phase this to avoid starting too soon and having to re-develop later, but also we need to provide minimum support to the ROCC if they continue their activities (unclear)
- Tanya reports that some work was being done at the end of last year, but now we are at full stop
    - indeed Tanya is fully booked out on other projects, and there is a big problem with knowledge transfer

EnVision
- Mark reports that the PSA and archiving in general were presented to the science study team recently, and that EnVision seems like a fairly "classical" PDS4 mission in terms of the archive

MSR
- Ernst reports on a meeting with Andrea and Gerhard Kminek last week on MSR programmatics
- The first samples to be returned are already at the first depot, delivered by Perserverance
    - deemed to have high scientific value
- The MSR *programme* covers all hardware **after** M2020 (Perserverance), including
    - sample retrieval lander
    - Mars ascent vehicle
    - Earth return orbiter
    - sample receiving facility
- Mark asks which hardware is European - ERO and fetch rover?
    - Ernst answers that the fetch rover is no longer part of the plan
        - instead it is assumed that Perserverance will still be operating to bring samples to the lander
        - derivatives of the Inguinuity lander are also being considered if P is dead

- The location and instrumentation of the sample receiving facility is not yet decided
    - there is also a question if Europe will have a counterpart here
- The NASA-ESA MoU includes mirroring of both flight data and sample return analyses at PDS and PSA
    - this will include analogue data, which could start soon!
- The sample analysis community is often different from the flight mission teams
    - and these people have little knowledge of archiving of PDS4
- There are 3 key documents coming up (see https://pds-geosciences.wustl.edu/missions/mars2020/returned_sample_science.htm)
    - Initial sample report
        - what is known about the samples at the time they are packaged
        - already on PDS and contain basic information on sample acquisition
        - these documents will not be changed
    - Sample dossier
        - more comprehensive, including links to papers
        - *according to the above link, currently in preparation*
    - Sample record inventory
        - a living document, appended with additional (meta)-data opver time
        - e..g during transfer to Earth and even after landing
- A Mars Sample Return Archiving Working Group is planned to be formed soon
    - about ~15 people in total
    - **the UG is asked to consider what representation is required from Europe**
        - **e.g. reps from PSA, laboratories, museums etc.**
        - **Gerhard would like an answer to this ASAP since the group should start this year**
        - **the proposal is to hold a virtual meeting of the UG before summer**
    - this group will likely conclude if archiving is to be done in PDS4 or not
- Gianrico asks if ESA has experience in other areas that might be relevant
    - Arnaud comments that HREDA has some relevance
- Mark will meet with Gerhard next week at CAB/INTA and confirm that the user group will organise a meeting as above
- ☐ All to brainstorm possible European delegates to the MSR archiving working group

16:00	Cross-archive collaboration Jos De Bruijne: Solar System Objects in GAIA DR3 Arnaud Masson: Heliophysics archives	GAIA - Jos presents SSO data in GAIA and eventually Euclid archives - large dataset of asteroid orbit, morphology, spin orientation, rotation period etc. - There are interesting planetary data in astronomy archives, but mostly they are not data one would want to combine with PSA, so ingesting them wouldn't make sense - but we can certainly try to point users to related data in other archives - DataLabs seems like the logical place to demonstrate how to use these - e.g. a GAIA/Euclid SSO notebook in the PSA DataLab - Gianrico strongly supports this idea since these are vital data for small bodies research - Sonia comments that the asteroid spectroscopy data have not yet been used in many publications - Ernst asks if we have lessons learned from the astronomy user group - Jos proposes that we offer a standing invitation to the chairs of all UGs to other groups sessions Heliophysics - Arnaud comments that helio and planetary have obvious synergies and suffer from an artificial split into difference archive areas - DataLabs again seems the logical place to put together some examples of how to combine helio relevant data from different spacecraft (e.g. SOLO and Bepi)
16:30	Coffee	
17:00	DOIs	- Mark outlines the current strategy for DOI issuance (dataset level for PDS3, data collections for PDS4) and shows the planetary DOI holdings: https://www.cosmos.esa.int/web/esdc/doi/planetary - Currently we have several things we want to do, that are not yet implemented: - issue DOIs on arbitrary collections of data (e.g. those used in a given analysis/paper) - this needs an API to ESRN for ESA DOIs and dynamic landing page generation on ESDC side - issue DOIs for documents - currently a manual process with CrossRef and a different contact at ESRIN - collections for supersets (e.g. ALL data from Rosetta/OSIRIS) - Andrea points our FAIR digital objects: https://fairdo.org which could be very useful in this context ☐ Mark Bentley to look into FAIR digital objects and how they could fit into the planetary workflow

→ THE EUROPEAN SPACE AGENCY

		• Jan asks how one can handle many DOIs in the long term
		
		   o Arnaud answers that in helio this is indeed tricky - raw data have a static DOIs and the latest calibration also. If/when data are replaced/updated with a later calibration, and data already pointed to by an existing DOI are also kept/
		
		• Anni asks about derived data (not only raw/cal)
		
		   o since this is a requirement when submitted e.g. to JGR - many authors use Figshare, Zenodo etc.
		
		   o Mark answers that we issue DOIs for the GSF datasets also, and any PDS4 derived data collections will also get a DOI
		
		• Anni asks if there is a way to search for DOIs?
		
		   o Mark answers that indeed PDS have a service for this, which is easier because PDS DOIs have their own prefix in DataCite. We have a TAP service recording all of the DOIs, mission, instrument etc. and adding a search on top of this should be straightforward.

17:30	Close	
20:30	Social dinner	Txirimiri Ferraz https://txirimiri.es/#txferraz Calle de Ferraz, 38, 28008 Madrid

→ THE EUROPEAN SPACE AGENCY

2. THURSDAY 16TH FEBRUARY

Time	Presenter and Title	Notes

09:30	Discussion - Evolution of the GSF
	
	• type and scope of deliveries
	• resources for validation
	• meta-data requirements
	    o Product_External
	• future database, UI, API etc.
		https://www.cosmos.esa.int/web/psa/psa_gsf
		
		• **Mark** presents the current GSF and highlights the main problems we have right now:
		    o lack of resources to do good validation (one of the conditions to avoid "polluting" the long term repository with poor data)
		    o the manual nature of the beast
		    o lack of meta-data, discoverability etc.
		      ▪ it is definitely not "FAIR"
		• **Gianrico** asks if the User Group could/would be asked to review data for the GSF
		    o **Mark** thanks Gianrico for the suggestion and indeed this would be most welcome, especially where we lack expertise in house for a given data type.
		• **Gianrico** asks if our DOIs are indexed by ADS?
		    o **Guido** answers not yet - discussions are ongoing for the heliophysics side and things should move forward, slowly, in the next year or so
		    o **Mark** adds that in particular the small bodies node of the PDS is pushing this for PDS4 DOIs
		• **Mark Bentley** mentions the new PDS4 product type "Product_External" which will allow the PDS4 information model to be used to describe non-archival data (i.e. data in a repository or outside of a PDS4 archive)
		    o but we would still need an easy way to generate XML labels etc. for users
		    o and a dynamic way to generate DOIs and create landing pages etc.
		☐ **Mark Bentley** to draft some users stories for a basic GSF "archive" based on Product_External
		• **Guido** mentions that the astronomy User Group have suggested repatriation of data from various e.g. EU projects

		- o the UG has reviewed H2020 projects and ranked them, and contracts raised to do the data repatriation - o do we need to consider something similar, in particular for large planetary projects? - Mark mentions that the GSF is not very visible in its web presence, but this is/was partly by design - o Ernst comments that this was partly done to avoid being overwhelmed with requests
10:00	Discussion - Legacy missions - maintenance and updates - available resources	- Mark summarises the situation - as already pointed out by Bruno we soon will have more legacy missions than current, and updates become difficult - PSA is in decent shape from the technical side, since we have common formats and data are stored in one archive - o but hard on the archive scientist side to validate new data - o and mission teams tend to dissipate very quickly after the end of a mission - Very old data do not typically result in many new papers - o e.g. Giotto, one in the last few years - Beatriz suggests that the User Group recommends the extension of the post-ops phase - Guido indeed for Rosetta we knew we needed a longer post-ops phase - and "enhanced archiving" contracts were very useful here - o on the other hand we have TGO, which as a HRE-funded mission has zero planned post-ops - Beatriz expresses concern over the model with external parties doing data generation (e.g. Comet Interceptor) and comments that we need to be very careful on the interface here - Antonio asks if the budget in EU projects includes money to make data available? - o Guido - yes, but never enough, and if not careful can get used for other things - o Ernst the steereing committees for such projects should really take this onboard and a good data management plan is needed from the start

→ THE EUROPEAN SPACE AGENCY

10:30	Discussion - PDS3 → PDS4 migration
	• summary of previous discussions
	• plans for 2023
		• Mark discusses the need for converting old data into the modern format, in order to take advantage of modern tools and services
		• Beatriz asks how we can better link to tools like pdr and other software to work with data
		  o Mark indeed this is tough - we have a long standing discussion pending on how to link from the archive to external resources, so watch this space
		• Gianrico asks how ESA is represented in PDS
		  o Mark adds that we have representation under the IPDA banner in the DDWG (data design working group) and on the CCB
		  o in addition we are in other working groups of relevance (DOIs, tools/software etc.)
		• Jan could we consider a survey to understand better HOW people use the data?
		  o Mark that would be useful - we know there is a drop-off in satisfaction after the download, but it's hard to get good data after that point

11:15	Coffee	

11:45	APIs (programmatic interfaces)
	• Current APIs
	  o PDAP, EPN-TAP
	• Future changes
	  o deprecate PDAP
	  o add support for PDS API (in progress)
	• Support for other APIs?
		• Mark reviews the current APIs
		  o PDAP is a legacy API and all IDPA members have agreed to deprecate this
		  o EPN-TAP provides a subset of meta-data compatible with other EPN (Europlanet) data providers for interoperability. Currently it allows download of products (zipped).
		  o In the coming months we are starting to ingest data into the PDS registry, which will provide support for the PDS API, and querying of arbitrary meta-data
		  o We also want to add a download endpoint which accepts an ADQL query and returns a single package containing matching products
		    ▪ to avoid the issue we have now of having to download multiple zips

12:15	Discussion - DataLabs • introduction and demo • use cases for PSA • plans for the next year	• Mark gave a quick demo of DataLabs o showing Octave (desktop) and Jupyter (web) applications • PSA already has two data volumes mounted in DataLabs

→ THE EUROPEAN SPACE AGENCY

		○ Rosetta ○ other legacy missions - Active missions are more difficult due to the need to separate public and private data - Waiting for DataLabs to support static mounting of volumes when a lab is launched ○ then the TAP services will be updated to provide the path to each label ○ this will allow reproducible data science without having to know the paths etc. ○ i.e. you query, return a set of products, and simply access them as needed - Antonio suggests we would benefit from a way to replicate the DataLabs environments also locally ○ Mark agrees that it would be useful to expose the environment files used
13:00	Lunch	
15:00	Organisation of the user group - meeting frequency - communication channels ○ mailing list ○ Confluence ○ Teams or OpenPlanetary Slack?	- The group discussed whether the community knows about the existence of the user group, which they really should see as their representatives to PSA. Ways to announce the new group and improve this could include: ○ DHAWG mailing lists for current missions ○ Twitter (PSA, ESDC) ○ PSA registered members list ○ PEN - Gianrico asks if and how we were represented at conferences ○ Mark says that PSA in general has a presence at EPSC as part of the ESA booth - Discussing usability and accessibilty of data, Beatriz commented that for some Rosetta data she had been advised *not* to go to the archive, but to use the data on the PI team site, which has better quality ○ all agree that this is not acceptable and we should push to have the best data in the archive ○ it was also commented, however, that the 6-month proprietary period is often not enough for final calibration ○ there has been a tendency for some teams to push "a" version to the archive, knowing it is not the best one

→ THE EUROPEAN SPACE AGENCY

		• Pending actions from the previous user group?
		     o GSF needs close attention
		     o Francesca proposed to host a student workshop in Rome, but this did not happen due to the pandemic.
		     o Paulo had identified problems with radio science data, which have not yet been actioned
		           ▪ suggests that Antonio Genova look at the recommendations and current state of the data to see what we should do.
		  ☐ Mark Bentley to check final recommendations from outgoing group and make sure these are capture and actioned where possible.

16:00	Close	

3.

→ THE EUROPEAN SPACE AGENCY

4. ACTIONS

Description	Assignee
☐ Mark Bentley to check final recommendations from outgoing group and make sure these are capture and actioned where possible.	Mark Bentley
☐ Mark Bentley to draft some users stories for a basic GSF "archive" based on Product_External	Mark Bentley
☐ Mark Bentley to look into FAIR digital objects and how they could fit into the planetary workflow	Mark Bentley
☐ All to brainstorm possible European delegates to the MSR archiving working group	
☐ Emmanuel to check if these will also be delivered to PSA	Emmanuel Grotheer
☐ Mark Bentley to compile a list of these and update the new user group on status and possible implementation	Mark Bentley
☐ Mark Bentley to find out what this entailed and where we are now	Mark Bentley