Rosetta Review, Day 1

9 October 2017

Attending:

Gerbs Bauer

Karl Hibbitts

Rudy Frahm

Tilden Barnes

Mark Perry

Silvia Protopapa

Emily Law

Ludmilla Kolokolova

Laurie Feaga

Sasha Kempf

Anne Raugh (recording)

Via Webex:

Laurence O’Rourke

Dave Heather

Kevin Walsh

“Hans-Herbert”

“Durand”

Thierry Semon

Patrick Martin

Matt Taylor

Larry begins with a summary of previous review results and the state of recommendations made. Highlights:

* The recommendation regarding geometry and SPICE has been acted on. Implementation is ongoing. Whether SPICE data might be delivered and when is not clear. Many agreements with teams now include enhanced geometry. [Is this limited to new data deliveries?]
* “Missing” data (published but not included in archive) have been attributed to pipeline “glitches”. These are being chased down individually.
* SREM data will be ingested in Spring 2018.

**VIRTIS**

(VIRTIS team and European reviewer Frederic Schmidt join Webex)

Frederic Schmidt’s review – one major RID for IR data that appear to be missing (but could not have been taken) and will be addressed by documentation.

Silvia’s presentation. The HISTORY object issue was raised again, but this has yet to be investigated. The bad data type for the side planes is a known ongoing issue the team fails to correct but rather documents it in the ERRATA.TXT file. The discrepancy between Silvia’s calibration and the published calibration was shown. F. Schmidt attributes this to an error in the calibration pipeline having to do with the saturation subroutine. It affects quite a lot of data, unfortunately, and was discovered only shortly before review. A sample of data from the revised pipeline will be supplied to the reviewers. The herringbone pattern also found by Silvia is also a known issue with pipeline geometry. The team plans to deliver corrected versions for this as well. The 428-432 length difference is because somewhere during the calibration pipeline 4 bins are removed. This needs to be added or emphasized in the calibration file.

The team notes there is no bad pixel map because bad pixels are flagged dynamically as part of the pipeline processing. The team will add a description of the pixel flagging/saturation routine to the documentation.

**COSAC**

Martin Rubin presents. He checked the RIDs raised by previous reviewers (not available for this review). Some negotiation over outstanding documentation (or possibly label) updates and how possible it is to provide the conversion factor requested. Significant outstanding RID from before involving geometry discrepancies between what’s in the label and what is output by SPICE. This is being worked. No new RIDs filed by M. Rubin – but clarifications on some existing RIDs.

Mark Perry presents. He filed no RIDs partly from insufficient time to thoroughly analyze the data. (Note that he was looking for calibration documents in the /CALIB directory, which is understandable but not what the directory is for. It’s not really any data preparer’s job to teach end-users how to navigate PDS3, but I wonder what’s in the CALINFO.TXT file in these empty directories.) He suggests that removing duplicated timing columns in the data set [not sure I heard that right] would make it much smaller and easier to manage for end-users. Larry notes there is precious little funding for changing pipeline output at this stage of the game, but they take the note.

**PTOLEMY**

Matt Rubin’s presentation – again reviewing the status of previous RIDs. Similar issues to COSAC and they will be handled similarly. The difference between the level 5 spectrum ad the published level 5 data is due to improved calibration. This will be documented.

Mark Perry’s presentation. No new RIDs. He notes there were some discrepancy between documents included and those listed in the docinfo.txt file, for example, but he assumes they will be cleaned up in final delivery preparations.

**SESAME**

Arnaud Masson’s presentation. Once again the “missing data” (compared to published data) problem arises. The morning discussion concluded this might be a documentation issue – that the 350 number is the upper limit and varies depending on the observational circumstances (I think – hard to hear). But that doesn’t actually account for the observations actually listed in the paper, yet not appearing in the archive. This will be investigated.

A. Masson notes time is only given in instrument time, not UTC. Larry and Hans-Herbert note the data are generated by SONC. This is slightly complex, because there are issues in correctly translating the instrument clock to UTC. This will be discussed and resolved as far as possible next week in Europe.

He also notes there are published plots that are not reproducible with the given information in the dataset. This appears to be documentation issues, but will be investigated. (This involves the PP side of the instrument, who were not directly represented on the WebEx.) There is also a plot of “missing” data again.

Kevin Walsh’s presentation. One non-RID comment – there are no “TAFT” data files, as expected. Martin Knapmeyer notes the foot temperatures recorded are all in the SEQP file. The documentation should be updated for that.

Once again – “missing data” corresponding to a published plot. Note there is no RID on this one specifically, but there are other RIDs in reference to “missing” published data. The team notes some of this “missing” data would be in the REBOUND phase, and thus not included in the data set under review.

Hans-Herbert notes the document K. Walsh listed as “missing or misnamed” does in fact exist by the given name but was inadvertently omitted. It will be added back in.

There is one DIM data point, but there is apparently some confusion over whether this is in the level 3 or some other level data set (it is calibrated). It should be in the Level 3 data, so this will be addressed.

**CIVA**

Colin Snodgrass’s presentation. He notes that the “bounce image” that was release to the public is still not delivered in the dataset. It also doesn’t show up in the “Timeline” slide showing the overview of when data were taken by the various instruments. There seems to be an issue with SONC not being able to decompress the image from its current format. The CIVA team note that the image needs some special handling in order to get it into a format that matches that of all other CIVA-P images. The team will work the issues. Some discussion followed regarding what is, or might be, delivered as part of each of the phases where, so far, no data have been delivered. (In some cases, CIVA was not on.) Sounds like ESA/PSA need to work out some moderately detailed issues, including whether all check-out data will be archived. But that’s another review…

Brief discussion regarding providing additional and more consistent calibration levels (I/F, albedo, etc.) for the various datasets. Essentially, the problems that need to be solved in this respect are sufficiently complex that this effort represents substantial improvement, rather than correction, of what is being archived at this stage.

Two difference calibration pipelines have produced results that are difficult to directly compare – one pipeline produces images rotated by 90 degrees relative to the others, e.g. This may be hard to address, but PSA will try.

Karl Hibbitts’ presentation. The question about the stereo pairs seems to be that they really are a stereo pair of images of the comet surface, with very low dynamic range. Yves Langevin says the camera 7 images looks suspicious, even then. Perhaps an issue with the normalization?

Karl notes the Level 3 data (Camera 4) appears to show vignetting. His plot “Divining the nature of the correction” raises questions. Yves L. explains that the upper blue range indicates pixels that are “saturated beyond recovery”. There is a file that explains the relationships shown in this plot, but apparently its visibility needs to be raised.

**ROSINA**

Andrew Morse’s presentation. Minor and editorial RIDs only.

Mark Perry’s presentation. He notes that his first-submitted RIDs will need revising, since they were based on a lack of access to the calibration documentation at the level 3 point when he was working them (first). The team (Kathrin Altwegg) refuses to do that because of the duplication of an important document that can diverge in subsequent development [which is a valid reason]. There is also a user guide in preparation that would likely mitigate the issue in future. [Larry and PDS understand the issues, so we’ll work out a reasonable long-term solution offline.]

He notes there appears to be a lack of significant digits for small masses, but K. Altwegg notes these *are* the significant digits for those small masses. (Apparently manual calculation might be able to do better, but not the automatic processing provided by the pipeline.)

**RPC-IES**

Rudy Frahm’s presentation. The allegedly ITAR-controlled documents will be removed (someone said they came from ESA?!?). SBN can help with PDF/A conversion for the mangled documents if needed.

Andrew Walsh’s presentation. Previous RIDs have been implemented. There is another mysterious potential readpds error (Missing valid “PDS\_VERSION\_ID” keyword). This should be investigated. [Can we get a failure sample? Did PDS3 validation flag this?] The reviewer had a number if issues with the ESA GUI to the data, that Larry has noted. Also, there is a potential manpower shortfall for producing higher-order data products. Larry and Gerbs will work out a formal recommend to, hopefully, encourage support for that.

After noon – SBN only Session

**GIADA**

Sascha Kempf’s presentation. Data are *very* well documented and organized. No RIDs submitted because no problems discovered.

Emily Law notes the version numbers are not consistently changed in the LABEL\_REVISION\_NOTE, where the dataset catalog file for V1.0 mentions a V1.1, which doesn’t exist. SBN can fix this, although there is a RID filed.

Certification status: Certified

**COSIMA**

Sascha notes some previous RIDs were rejected, but the issues raised then are still issues for potential users. In particular, confusing terminology; monolithic distribution; little useful documentation on calibration; more caveats than statements; etc.

A. Raugh [the author of these notes] had to leave to investigate a major hardware failure, but the discussion regarding the non-usability of the data continued, with some discussion regarding reproducibility of published results, providing reasonable guidance for this sort of data, etc. Sascha notes that the data are technically sound, just not usable by anyone outside the team.

(recorder out of the room for ~ 45 minutes).

*Certification status: Certified*

***NAVCAM***

*Question regarding FITS in EXTRAS/ rather than DATA, but this is up to the mission. We can serve FITS files at SBN though, as long as the archival form is saved. This will inconvenient, but possible.*

*Certification status: NOT Certified*

***MIDAS***

(joined in progresss)

Michael Hecht notes that the IDL code provided (already) doesn’t run. There’s no reason why it should, but it’s yet another argument for not using code as documentation.

Note that readpds was able to read the data object, and NASAView could to a point, but the display is not the form discipline scientists expect. Still needs some substantial documentation improvements. He responded that the IDL code, while not compatible with the version he attempted to use, is at least well-documented and could serve as an illustration of the algorithm or coding model.

Mike Kelley notes there is a scaling factor included in the object definition.

Certification status: Reviewer considers the data certifiable – the primary liens are on documentation completeness.

**ALICE**

Joshua Kammer’s presentation. Four minor RIDs on documentation.

**RPC-ICA**

Rudy Frahm’s presentation contains a laundry list of significant documentation issues, including some items left as an exercise for the user to provide – inappropriately.

(recorder steps out briefly as RID review continues)

The only “calibrated” data supplied is all zeroes. From the other data, the browse data in the archive shows data where the published results show nothing.

Certification status: Level 3, no – no data was provided to review. Level 2 – will wait on tomorrow’s discussion.

**RPC-MIP**

Rudy Frahm’s presentation. Rudy notes the RIDs already in the system for this review mention level 5 data, but he did not find level 5 data. Actually, these data were an optional part of the review.

In the data files, Rudy notes duplicate records and instances when time stamps move backwards. This is a show-stopper.

Emily notes some labels were missing the spacecraft clock start/stop times.

Certification status: Not certifiable.