PDS Peer Review for raw radio science Voyager data at Jupiter

Data Provider: Richard Simpson (Radio Science Advisor, Ring-Moon Systems Node).
Science Reviewers: Yohai Kaspi (Weizmann Institute of Science), Paolo Tortora (University of Bologna).
PDS Reviewers: Richard Chen (Engineering Node), Lyle Huber (Atmospheres Node).

In the following document, the reviewer comments are numbered, with the data provider action (and any subsequent comment) indicated by "DP action:" (and "**DP comment:**"); the responses from Ring-Moon Systems Node (Mia Mace, Mitch Gordon, and Matt Tiscareno) are noted in *blue italics inline*.

rchen #1: add lid_references to VG1J context products that appear in product labels to parent (*e.g.*, collection and bundle) labels:

DP action: Added lid_references to collection and bundle labels as requested.

RMS: <Context_Area> has been added to bundle_*.xml, collection_data.xml, collection_document.xml (and refer to rchen #4).
Confirmed that <Context_Area> present in collection_browse.xml, collection_calib_freq.xml, collection_context.xml, collection_data.xml, collection_document.xml, collection_geometry.xml in each bundle, and bundle_*.xml

rchen #2: Create subdirectories to divide the files, aiming for no more than 20 products in each subdirectory (per a DPH recommendation)

Most future scientists handling these data will want to concatenate the samples so they can be processed coherently. That the original samples are currently blocked into 54 files of nominally 16.92 MB is a historical artifact of the 1979-era magnetic tape technology. The samples are, in fact, continuous across product boundaries with two exceptions. Future users will likely create one file of 16*16.92 MB from the "ingress" occultation products and one file of 35*16.92 MB from the "egress" occultation products. There is an unexplained 20-minute gap between ingress and egress in the Voyager 1 data. There are also three files with results from pre-encounter tests. Separating the files into subdirectories hinders coherent processing rather than helping — especially if the 35 egress products are separated into two (or more) subdirectories.

Because of the gap in the Voyager 1 data during the deepest part of the occultation, the Voyager 1 products could be divided into pre-encounter tests (3), early occultation products (16), and late occultation (35) products. But there is no gap during the Voyager 2 occultation, so any division of products in that case would be based on totally arbitrary criteria.

DP action: None. The 54 labels can be viewed using a command like `ls *.xml`. Separating coherently acquired samples into arbitrarily defined subdirectories makes science data analysis more cumbersome. PDS4 sets no requirements for physical storage of data, which is at the discretion of curating nodes. If the curating node wishes to use subdirectories, it can do so; the data provider believes the current structure is better for users.

RMS: R Simpson's argument is valid and no action is required.
In the DPH, under Section 10 ('Assembling the Whole Archive'), the following is stated: "Start assembling your archive by building a directory tree to hold the pieces. In particular, if you have more than about 20 data products, consider how to divide your many data files and their labels into subdirectories. The directory tree usually mimics the bundle and collection hierarchy, but it doesn't have to." There are 326 files and 248 files (and no subdirs) in the Vgr1 and Vgr2 data collections, respectively (although there are only 54 and 41 product XMLs, each product is 6 files). So R Chen's suggestion sounds reasonable. However, R Simpson's point about usability holds.

DP comment: Most users would prefer to work with the ASCII translations — the record headers (for housekeeping information), the S-band samples, and the X-band samples. The [ODR files] are in an archaic format [8-bit byte MSB integers and floats]; the ASCII translations, although bigger files, will be much easier for users to read into contemporary software. Also, extracting the record headers, the S-band samples, and the X-band samples into separate time-ordered files means that analysis can be constructed on an entire occultation ingress or egress following a simple 'cat' command.
Although it's true that the [ODR] files are in alphanumeric order, that is not the same as time order. Some of the original data will be out of time order if concatenated based solely on file names. The ASCII translations have file names constructed using date and time, so they will be in time order.

rchen #3: Labels list possible values header/trailer tables; in fact, other values appear.

DP action: The list of expected values has been removed from labels. The listed expected values are identifiers for prediction files which are not included in the archive. See also rchen #12.

RMS: The expected values have been removed from <File_Area_Observational> :<Table_Character>:<Record_Character></Field_Character> :<description> and <File_Area_Observational> :<Table_Binary>:<Record_Binary>:<Field_Binary>: <description> for the label files flagged by R Chen; i.e.:

data/vg1j_63mr_1979064t123001.xml

data/vg1j_63mr_1979064t123321.xml

data/vg1j_63mr_1979064t123641.xml

data/vg1j_63mr_1979064t184300.xml

data/vg1j_63mr_1979064t184620.xml

data/vg1j_63mr_1979064t184940.xml

data/vg1j_63mr_1979064t185300.xml

data/vg1j_63mr_1979064t185620.xml

data/vg1j_63mr_1979064t185940.xml

(And the last value on each record in the .tab still has value "JOSE" or "VG23").

rchen #4: Add references to context products in document labels (then propagate to document collection and bundle, if not already there).

DP action: Done. Context products have been added to document product labels, propagated to the document collection label, and then propagated to the bundle product label (as needed).

RMS: <Context_Area> has been added to: mission.xml, sis_vgj_rs_1.0.xml, vg1j_hga.xml (and collection_document.xml and bundle_*xml, as noted in rchen #1).

rchen #5: Correct http references to https.

DP action: Done.

RMS: Yes, http references fixed for all noted –
document/mission.xml
document/sis_vgj_rs_1.0.xml
document/vg1j_hga.xml
calib_freq/vgr_uso.xml

rchen #6: document/mission.xml differs from document/mission.xml in the Voyager 2 document collection only in its LID; both point to the same document. Please converge on one LID, which should be primary in exactly one collection.csv.

I think I understand the comment, but I don't understand the problem. mission.{pdf,txt,xml} is a primary member of the Voyager 1 document collection. There is a mission.{pdf,txt,xml} which is a primary member of the Voyager 2 document collection. Each could be tailored to its spacecraft, though that has not been done so far. These are not context products, so I'm puzzled by the request for a conformity that was never intended.

DP action: None.

RMS: The mission.pdf files in the Voyager 1 and 2 document collections are identical; the collection_document.csv for Voyager 1 lists:
P,urn:nasa:pds:voyager1_rss_jupiter_raw:document:mission::1.0
and the collection_document.csv for Voyager 2 lists:
P,urn:nasa:pds:voyager2_rss_jupiter_raw:document:mission::1.0

There should be two distinct products with separate copies of the {pdf,txt} files even if currently the contents are the same. When they operated, NASA referred to them as separate and distinct missions. These documents could easily diverge, for example, if someone wanted to include targets in the documents. There should be only one LID assigned to a product, but a product is a label plus the file(s) it describes. There is no restriction on having two separate products (labels with different LIDs) with identical files. It is fine to not include secondary member references to the mission documents and it's clear that the Vgr2 mission document is not listed as a primary member of the Vgr1 bundle (and vice versa).

DP comment: Cross referencing the mission document(s) might only help if they were different; since the document was written in 1977 and is the preprint of a paper published before the encounters, I see no reason to make changes. Not done, per your allowance.

rchen #7: document/sis_vgj_rs_1.0.xml differs from document/sis_vgj_rs_1.0.xml in the Voyager 2 document collection in its LID …; both point to the same document. Please converge on one LID, which should be primary in exactly one collection.csv.

Same response as above. This Software Interface Specification (SIS) was developed to cover both Voyager Jupiter bundles. It was given different LIDs to allow for diverging content as the bundles were built; but content has not diverged. The documents are primary members of their respective document collections; it should not matter to users whether they have identical content.

DP action: None.

RMS: As above.

DP comment: I think cross referencing the SIS documents [Vgr1 SIS as secondary member of Vgr2 document collection, and vice versa] would be more confusing than helpful to users.

rchen #8: Correct LID bilateral references.

DP action: Done; the products should point to each other.

RMS: Yes, geometry/vh008a.xml now has an <Internal_Reference> to urn:nasa:pds:voyager1_rss_jupiter_raw:geometry:vh008b, and geometry/vh008b.xml an <Internal_Reference> to urn:nasa:pds:voyager1_rss_jupiter_raw:geometry:vh008a

rchen #9: add lid_references to VG2J context products that appear in product labels to parent (*e.g.*, collection and bundle) labels:

DP Action: Added lid_references to collection and bundle labels as requested.

RMS: <Context_Area> has been added to bundle_*.xml and collection_document.xml. But should there be a <Context_Area> in collection_data.xml?

DP comment: Added Context_Area to three collection labels in each bundle.

RMS: Confirmed that <Context_Area> present in collection_browse.xml, collection_calib_freq.xml, collection_context.xml, collection_data.xml, collection_document.xml, collection_geometry.xml in each bundle, and bundle_*.xml

rchen #10: browse/vg2_radio_ing.xml and browse//vg2_radio_egr.xml have LIDs which do not match the LIDs in collection_browse.csv.

DP action: Corrected the LIDs in collection_browse.csv so they match.

RMS: Validate referencing checks pass, and checked by eye.

rchen #11: Create subdirectories to divide the files, aiming for no more than 20 products in each subdirectory (per a DPH recommendation)

See rchen #2.

RMS: See rchen #2, no action required from data provider.

rchen #12: Labels list possible values header/trailer tables; in fact, other values appear.

DP action: The list of expected values has been removed from labels. The listed expected values are identifiers for prediction files which are not included in the archive. See also rchen #3.

RMS: The expected values have been removed from <File_Area_Observational> :<Table_Character>:<Record_Character></ Field_Character> :<description> and <File_Area_Observational> :<Table_Binary>:<Record_Binary>:<Field_Binary>: <description> for the label files flagged by R Chen; i.e.:

data/vg2j_14mr_1979191t131727.tab

data/vg2j_14mr_1979191t131727.tab

data/vg2j_14mr_1979191t131727.tab

data/vg2j_14mr_1979191t132407.tab

data/vg2j_14mr_1979191t132407.tab

data/vg2j_14mr_1979191t132407.tab

(And the last value on each record (in the .tab) still has value "TEST").

rchen #13: The LID in data/vg2j_14mr_1979191t131727.xml is vg2j_14mr_1979191t131727; but the LID listed in collection_data.csv is vg2j_14mr_1979191t131728.

DP action: Corrected the LID in collection_data.csv.

RMS: Validate tool's referencing checks pass, and verified by eye.

rchen #14: Mismatch between LID in document/sis_vgj_rs_1.0.xl and LID in collection_document.csv.

DP Action: Correct LID in collection_document.csv.

RMS: Validate tool's referencing checks pass, and verified by eye.

rchen #15: Correct http references to https.

DP action: Done.

RMS: Yes, http references fixed for all noted –
document/mission.xml
document/sis_vgj_rs_1.0.xml
calib_freq/vgr_uso.xml

rchen #16: Correct LID bilateral references.

DP action: Done; the products should point to each other.

RMS: Yes, geometry/vu001a.xml now has an <Internal_Reference> to urn:nasa:pds:voyager1_rss_jupiter_raw:geometry:vu001b, and geometry/vu001b.xml an <Internal_Reference> to urn:nasa:pds:voyager1_rss_jupiter_raw:geometry:vu001a

lhuber #1: Remove spurious .DS_Store files in some directories.

DP action: Done.

RMS: Done.

lhuber #2: Add Context collection to bundle.

DP action: Done.

RMS: Done.

lhuber #3: LID inconsistencies between VG2J products and Inventory listings.

DP action: Corrected (see rchen #10, rchen #13, and rchen #14).

RMS: Done.

ptortora #1: The documentation does not explain how to use the data.

There are many ways to "use" these data; it is not within the scope of this SIS to describe the various approaches. Instead, the SIS references a document that explains and illustrates how to extract values from original binary files [3], two documents that explain receiver operation [2,4], and several papers that present results [5-9]. Two references [6-7] describe the atmospheric profiles derived from the Voyager Jupiter data. The request for "specific guidance on the data processing" would be hard to meet; scientist have invested careers developing those procedures. See also the response to ykaspi #5 below.

DP action: None.

RMS: Data provider's response is valid.

ptortora #2: The documentation does not allow identification of the ASCII file which contains the header/trailer data and the ASCII files which contain the receiver samples. Absence of a file naming guide is a major drawback.

File naming is described in SIS section 3.2.1. Section 3.2.1.1 explains that file name extensions reveal the ODR-derived file contents; for example, files with extension 'hdr' contain values from record headers and trailers, and files with extension 'tab' contain translations of binary receiver samples to ASCII. The base file name (before the period) includes a field which distinguishes the 'band' — for example, vg1j_63mr_1979064t184300.txt is the text file with terse field names

(column headings) that accompanies the 'hdr' file. 'm' indicates that the file applies to both S- and X-band data.

DP action: The VG1J bundle_readme.txt now includes an example showing the relationship between an original VG1J ODR (binary) and the derived ASCII files. The same example has been added to SIS section 3.2.1.1. A VG2J example has been added to the VG2J bundle_readme.txt file.

RMS: The additional examples in the bundle_readme.txts are helpful; suggest that they also refer to the SIS directly: "For example, the following six files constitute one PDS4 product (refer to SIS, section 3.2.1, for file naming conventions) …".

DP comment: Each bundle_readme file already referenced the SIS in the last paragraph. I added more explicit language.

RMS: Done – the following sentence was added to the bundle_readme.txt in the final paragraph: "The SIS (sis_vgj_rs) describes in more detail how the products in this archive were constructed from the original raw data files."

ykaspi #1: [it] would be useful to have a contact person to inquire about original data [in bundle_readme.txt].

The function of bundle_readme.txt is to provide a high level overview of the bundle. Details — such as contacts, file naming, and data location — are in the SIS.

DP action: None

RMS: As for the point above, it would be helpful to direct users to the SIS from the bundle_readme.txts.

DP comment: As ptortora #2 above.

ykaspi #2: bundle_readme.txt doesn't provide a clear reference to which folder each data product could be found.

See two previous responses.

RMS: As for the two points above, it would be helpful to direct users to the SIS from the bundle_readme.txt's.

DP comment: As ptortora #2 and ykaspi #1 above.

ykaspi #3: the documentation concerning the trajectories is not very complete/clear but this is signified in the description file accompanying the data and can at this point be done easily with SPICE kernels.

The trajectory data are provided primarily as historical documentation — a future analyst might use these to reconstruct the occultation geometry as seen by the Voyager Radio Science Team members. For a second *science* analysis, the SPICE kernels would likely be preferred.

DP action: None.

RMS: The data provider's response is valid.

ykaspi #4: bundle_readme.txt explains the transformation of the .ODR files to the .TAB files containing the X-band and S-band data, but could state the naming of the file again (what is X band and what is S band data).

DP action: See response to ptortora #2 above.

RMS: See response for ptortora #2 (it seems as though users are having difficulty finding the relevant information in the SIS, so more signposting to the relevant section would be helpful).

ykaspi #5: The ASCII file with record header/trailer data is actually the data that will be used for further radio science (frequencies for calculations), which is not clear from the data. The explanation of how to use the data and what the actual frequencies are (since there are three columns) is also unclear. I suggest more clarity concerning the useful part of the data, while referring clearly with examples to the nomenclature of each product.

DP action: Section 5.4.1 has been added to the SIS; it includes an example calculation based on [4] and notes the relationship among the frequency and phase quantities in the header/trailer file that was documented in [2]. Refining the example calculation to include USO drift, Doppler effects due to spacecraft and receiving antenna motion and general relativity, corrections for propagation through Earth's atmosphere and ionosphere, and various timing and frequency errors in data acquisition and storage is well beyond the scope of this restoration project.

But the reviewer may be misunderstanding what is contained in the header/trailer file. These are *not* the measurements that would be used for further radio science analysis; they are settings for the equipment used during data acquisition. See response to ykaspi #9 and ykaspi #10 below.

RMS: Section 5.4.1 of the SIS is a useful addition.

ykaspi #6: The information about the Voyager missions and the radio science instrumentation itself is good but the scientific papers referring to the radio science published by use of the data is insufficient. I have noticed that it is necessary to archive the relevant articles and even notes that authors had, together with the rest of the documentation in a separate folder, so that future generations of scientists can easily find each phase of the science.

Archiving science papers that report science results is difficult given copyright protections. PDS usually considers that a complete citation is sufficient; users can then negotiate terms under which they can access the papers. Known notes from the Voyager Radio Science Team have been included in the archive or referenced documents; there are very few.

DP action: None.

RMS: Yes, it's a copyright issue.

ykaspi #7: If any data processing tools/codes exist and are public, they should be made available for future users.

Almost nothing remains. This is a data recovery/restoration project carried out long after the Voyager Radio Science Team disbanded. No one who participated in the original data collection and analysis has been involved in the restoration; most of the original analysts are no longer alive. The few who remain alive are either retired or no longer active in the field.

DP action: None.

RMS: This addresses the reviewer's request.

ykaspi #8: Also the small ASCII file containing the header has an additional line of 34 fields that could confuse users. The descriptions of the useful frequencies are difficult to understand (definitely for new users), so here additional clarity could be added (what are these frequencies, what is the final product, how was it achieved?).

DP action: A footnote has been added to SIS section 3.1.2.1 noting that the text file has two records — one with field names, the second with field formats. The various frequencies are explained in [2], which is now referenced by new SIS section 5.4.1.

RMS: This clarifies.

ykaspi #9: Much of the processed Radio Science data products are in the form of a .TAB file and an accompanying .LBL file, which is in my opinion easier to work with and much clearer in

ways of description of each field and use. It would be beneficial to split the instrumentational specifics (receiver information) and the useful data (frequencies, time) for future science.

DP comment: I'm not aware of any processed radio science data from the Voyager Jupiter occultations. If the reviewer is referring to processed data from experiments conducted with other spacecraft, then there are two answers to the comment:

(1) The .TAB and .LBL pairing was standard for products archived under PDS3 standards; the bundles being archived here meet PDS4 standards which, among other things, use .XML labels rather than .LBL labels.

(2) The .TAB files mentioned by the reviewer (from other missions and experiments) may contain time stamps, frequency residuals, and (possibly) estimates of signal strength — they are *measurements* of signal parameters provided by the data acquisition equipment. In these Voyager 1/2 Jupiter bundles, the frequencies in the .TAB files give the output frequencies of the ground receivers' programmable local oscillators (and their drift rates). They are settings for the data acquisition equipment; they are not measurements of the received signals.

DP action: None.

RMS: The .tab labels in these Voyager bundles make it clear that the frequencies refer to the receiver settings, rather than received signals: searching for "(F/f)req" in the TAB labels finds Commanded frequency (clearly refers to receiver, and which is described in Section 5.4.1 of the SIS, under 'Receiver Operation'), and Ramp Start Frequency (its <description> references the Dana Synthesiser).
No action required of data provider.

ykaspi #10: An explanation on how the data was created could be helpful, starting from information about the X-band and S-band raw data files from the Voyager Jupiter era DSN radio science open loop receivers to the technique used to achieve the frequencies of the signal (which is the data product that was created and will be used in the future for science).

The data were not "created". I believe the reviewer misunderstands the contents of the .TAB files — they document settings of the receivers; they are not measurements of signal frequency (see response to ykaspi #9 above).

DP action: None.

RMS: As above, no further action required of data provider. If the data provider thinks that a sentence or two to describe his response would be helpful in the documentation, this would be supported.

Additional RMS comment: *In the bundle XMLs, under the <Context_Area> in <Target_Identification>, it states collection_to_target instead of bundle_to_target.*

DP action: Corrected.
RMS: Done.

Additional RMS (minor) comment: *We noticed there were a couple unclosed parentheses in data XMLs, which you might consider fixing – e.g. data/vg1j_63mr_1979064t123001.xml: Line 460 "(Receiver 1 operated at S-band" and line 475: "(Receiver 2 operated at X-band." Please only fix these if it requires minimal effort while you are updating the bundle.*

DP action: Corrected.
RMS: Done.