Review report of the Experiment to Archive Interface Control Document (EAICD)

EAICD: XXXXX (e.g. ACP)

1.2.3 Preamble

As an independent reviewer of the Huygens archive, you are asked to read first the Data Archive Plan. This document provides you with an overview of the archiving activities within the Huygens project. It also summarizes the products that will be archived (see the appendixes C to K). Those products have been negotiated with the teams and are listed in the table:

1.2 Experiment	**Raw data**	**Calibration information**	**Reduction algorithm**	**Calibrated data**	**High level data**
HASI	X	X		X	X (TBD)
SSP	X	X	X	TBD	
ACP	X	X		X	
GCMS	X	X		X	X (TBD)
DISR	X	X	X		
DWE	X	X	X	X	
DTWG			X		X
HouseKeeping	X	X		X	

TBD: To be defined

Then, you are asked to read the individual Experiment to Archive Interface Control Documents. You are only assigned to review some of these documents, depending on your field of expertise.

The EAICD provides users of each experiment with a detailed description of the product and a description of how it was generated, including data sources and destinations. Also, it is the official interface control document between each team and the archiving authority.

As the EAICD itself will be part of the documentation of each data set, it is one of the entry points for scientists interested in the Huygens data. It is therefore very important that such a document must be clearly written.

The Huygens Data Archive Team (Olivier Witasse, Joe Zender for ESA; Lyle Huber for PDS) has been working with the teams and helped them generating this document, which contains three important parts:

Section 2: Overview of process and product generation

Section 3: Archive format and content

Section 4: Detailed interface specifications

We ask you to answer to the following questions:

Main Comments

The acronyms list is incomplete and not alphabetical. Modes should be defined in more detail in the EAICD, not just in the references. There are lots of non-binding terms used to describe what will be archived. If only DNs and voltages are archived, which seems to be all that is presently commited, then this archive will be useless to the scientific community. Archived sampling rates of several instruments are not stated.

First topic: Structure and completeness of the EAICD.

Question #1: In your point of view, is each section understandable? Do you miss some important information?

The acronyms list is (very) incomplete and not in alphabetical order. A few examples: DPL, TBD, TBW, DTWG, PSSRI, NASA, ESA...
Section 1. SP 1177 should be in the reference documents.
The contact names and address need countries and street addresses.

Section 2. Many references to AD.x, which will not be available to any users of the data. This is useless for potential users of the final SSP archive.
Fortan used for Fortran here and elsewhere.
The sequence of sensors in Table 1 and elsewhere in the text should always be the same. API-V and API-S switch back and forth.
2.3 RD-01 should be RD-1, references to, eg, Lorenz et al that are taken directly from RD-1 should be listed as RDs in Section 1.5
2.4 When does M1 start? And end? And so on. A qualitative discussion here would be useful to the user. Are the definitions in SP 1177 still valid?
Section 2 is very understandable, except for the multiple AD.x references at the start.

Section 3.
SSP uses a consistent, logical structure with good definitions. This detailed description will be appreciated by users,
I presume the “3” and “3/4” in the Data Set IDs refer to data levels. If so, then the data processing level is clear to users.
Does “data processing level” of raw/calibrated mean the same as the PDS/PSA data processing levels of 1-8 in the Data Archive Plan? If so, then the same scale should be used in each case, if not, then it needs a different name to prevent confusion.
3.1.4 Defining mode and id in AD.3 is not useful to a user who does not have any AD documents.

How does a spec of ATMOS correspond to SSP's operating mode? And so on.
3.2.2 Zero time is defined as when Huygens crosses the entry interface of 1270 km. Entering the “atmosphere” is not a well-defined event.
3.1.2 and 3.4.1 “provisionally foreseen” and “aim to deliver” are non-binding. Use binding terms to describe what will be delivered.
Table 3 is potentially very useful as it describes exactly what's in the archive for each sensor and mode. However, I found it difficult to understand.
My copy of Word shows that ACC-E has a horizontal line separating modes 5 and 123, but API-S does not. This should be consistent.
What is an array? I don't see any difference between ACC-I-1236 in 4.2.17.6, which doesn't have an array, and API-S-23 in 4.2.18.1, which does.
What is an hktable?
Some tables have (n x 3), some don't.
A format like:
DL + (n x 4) table + (n x 4) hktable + (512) array (mode 2,3)
could be used consistently for all entries.
Lab data still TBD.

Table 4. Since there is no commitment to archive any level 4 data, the data set names should contain “3”, not “3/4”. I see from my instructions that ESA and SSP have not yet negotiated the delivery of any calibrated data, which is consistent with Table 4. DNs and voltages will be archived. The CALIB directory is still TBD. If this archive does not contain sufficient calibration information for DNs to be processed into physical properties, then it will be useless to all users. If it does contain sufficient calibration information, then why not archive the calibrated data? That saves the time of users and prevents different groups arguing over whose derived version of the calibrated data is correct. Presenting it in a publication is not sufficient.

Table 5 is very useful. The information from SSPTEAM.PDF should go in PERSON.CAT.
Contents of CALIB directory are TBD.
What is {file} and where is {dpl} used?
The text listing of the CATALOG directory doesn't match Table 5.
Geometry. Huygens position and velocity will be archived by the DTWG. Any information on the attitude of Huygens would be useful here, but I do not know if that will be available.
Table 5 says “software ... to access/process the data products”. Later, the software directory is described as containing software to extract SSP data from the telemetry packets. The telemetry packets will not be in this (or any?) archive, so I do not know which users will ever find the software useful. Software that uses the archived data files as input in some way would be most useful to users.
Documents directory. Do any of these articles have AD/RD numbers? The present names are too terse for most users to understand.

Section 4
4.1 Data products will be tables. Any arrays from Table 3?

4.2 “Sample labels are subject to revision” There is a lot of TBD material in this EAICD.
I do not understand what sets the sequence of sensors in 4.2. They have been alphabetically ordered everywhere else.
Should the time in product creation time have “Z” at the end to signify UTC?
Mission phase name of “All modes-1(descent)...” doesn't match with description in Section 2.4 where Mode 1 = upper atm.
Data_quality_id refers to quality.cat, instrument_mode_id refers to ssp_modes.cat. Neither of these cat files are in Table 5.
The various data labels all seem to be consistent with each other, which is good.
Mission_phase_names should be consistent with mode names in section 2.4. There are inconsistencies within the mission_phase_names as well, eg MODES-1,2,3(Atoms) and MODES-123(ATMOS). Atoms is clearly wrong.
ACC-I data labels. Data Label 2 has M1236 for the data and M123456 for the HK data. Why the inconsistency? Where is M4 for ACC-I? 4.2.17.6 says that it operates continuously.
API-S. The M45 file promised in Table 5 is now just M4. The Product ID is wrong as well.
There are two HK Data Label 1's, the only difference being in whether “TEMPS” is in the entry for ^TABLE. Using HKTEMPS, rather than HK, throughout this label is inconsistent with how the other sensor names have been treated. HK Data Label 2 looks like HK Data Label 1 with “TEMPS” removed. I haven't seen anything else in this document that distinguishes between HK and HKTEMPS.
There is no data label for STATUS.
Data Quality flags are not defined.

Data Object Definitions
The units of each measurements (s for time, V for voltages, etc) should be defined for each object in each table. Will these times be UTC, seconds since T0, or some other time?
ACCE_MODE5_TABLE. Names including (Raw A/D Values) and (V) are not formatted correctly, see Section 5.4 of the PDS Standards. Try, say, “_V” instead. This also occurs in other tables.
ACCE_MODE123_DN_SERIES. Why is this a SERIES, not a TABLE? Hyphens and underscores are mixed in object names. Please use only underscores. This also occurs in other tables.
There is no table for ACC-I M5.
ACCI_MODE1236_TABLE. Are “raw ADC counts” the same as “Raw A/D values”? Column 7 is labelled as Column 3.
APIS-MODE23_TABLE. The text here doesn't make much sense. The last sentence is incomplete and refers to ACCI.
APIS-MODE4_TABLE. This data reduction description doesn't make much sense to me either. Does “averaged by 4” mean that 4 samples are averaged together? Are the “further 140 samples” taken before the 60s surrounding the peak, immediately after the 60s surrounding the peak, or some other time? 60 uncompressed + 140/4 + 800/20

= 60 + 35 + 40 = 135, not 136. I'm not sure what PEAK_POS refers to. Is it the number (0-999) of the sample that had the largest DN, the time of that sample, the number from 0-136 of that sample, or something else entirely?
APIS-MODE6_TABLE. Again, unclear text describing the reduction.

Tables are needed for DEN, PRE, REF, and THP. I do not yet know if the sampling rates of those sensors are 1 per hour, 1 per second, or something else. Will DNs be archived? Will voltages be archived? These four sensors have different entries in Table 5 (table/array/hktable), so there will be some differences.

Where are tables for TIL, HK, and STATUS?

The file in the Appendix has an inconsistent mission_phase_name. Also, Object ACCE_MODE1,2,3_DN_TABLE (Appendix) is not the same as ACCE_MODE123_DN_SERIES (4.2.17.2)

Question #2: Is the EAICD itself understandable with respect to potential future users (taking into account the long-term preservation of the data - overall coherence of the document)?

The sensors are described well in Section 2 and the SP 1177 article. The format is described fairly well in Section 3. The instrument modes, discussed in SP 1177, should be discussed in the EAICD more than they are, the data object definitions are incomplete, and there are some inconsistencies between Table 5, the data labels, and the data object definitions.

Question #3: Is the EAICD coherent with the Data Archive Plan? Check in particular the conformance to the standards (section 6.6) and to the appendixes.

Except for the few incomplete parts noted above, yes. It is unfortunate that data will only be archived as DNs and volts. That will make the archive useless for most practical purposes.

Second topic: Scientific and technical content.

Question #1: Are the scientific objectives clearly and concisely described (in section 2)?

Yes.

Question #2: The processes involved in the data flow from the Huygens probe to the ESA Planetary Science Archive are very important, in order to understand how the data are processed and transformed.

Is it clearly described in the document (section 2)?

Once the data object definitions are improved, it should be clear to a user what sampling and averaging has been performed onboard the probe. Software will be provided that extracts raw data from the telemetry packets. The path from the probe to the telemetry packet sitting in Milton Keynes is not described in Section 2.

Has the team committed to provide to providing algorithms that will allow long-term use of the data and comparison to future datasets?

No. DNs and voltages are not useful to scientists.

Question #3: Taking into account the reference papers (e.g. the space science review paper), do you think that the data products are clearly identified?

Are they clearly described?

Will these products support the scientific goals?

The measurements are well-described in the reference papers. The data products are also described well in terms of DNs and voltages. These will not be scientifically useful.

Question #4: Do you think the calibration information is carefully addressed (sections 2 and 3)?

Calibrated data (see Table 4) will not be archived and calibration information is TBD.

Question #5: Do you think that the validation of the data is carefully addressed in this document (section 3)?

It is briefly addressed in three lines of Section 3.3

Question #6: Is the geometrical information addressed?

Yes. DTWG results will be used.

Question #7: Is the set of documentation (intended to be delivered with the dataset) is complete and sufficient for data calibration and processing, data visualization and analysis?

The contents of the CALIB directory are still TBD, so I do not know whether it will be sufficient or not.

Third topic: Long-term access to the data.

The data will be archived under the directory /DATA. Each team is free to organize the content of this directory. For each data product (e.g. a table, an image, etc..), a label file is provided.

Question #1: Are the selected data structure clear and useful (section 3)?

Once Table 3 is made clearer, they will be fine.

Question #2: The filenaming convention is explained in section 3. Please comment on the specific choices that have been made.

Generally, it is consistent and logical.

Fourth topic: Data Product Labels

PDS data products labels are required for describing the content and format of each individual data products within a data set. Examples of label are given in section 4 of the EAICD.

Question #1: From the proposed labels (see section 4), is the list of keywords clear and understandable?

Good.

Question #2: In the proposed table objects, is the description of the columns clear enough? (Column name, text description, unit…)

Good.

Please list here the additional comments you may have on this document, if any.

XXX EAICD	[Minor] List of editorial comment

XXX EAICD	[Major] Comment in section xx

Please list here the additional comments you may have on the Data Archive Plan (D.A.P.), if any.

D.A.P.	Comment in section xx

D.A.P.	Comment in section yy