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 descriptions of the reconstruction process in AD.3 and AD.4 are good. It would be of great help to users if the reconstruction code were archived; this is still TBD. The input data actually used for the reconstruction should be specified as well as possible (eg, HASI ACC XSERVO file xxx.TAB) and should all be available to users either in other instrument archives (HASI, GCMS, SSP, etc) or in the DTWG archive (entry state, aerodynamic database used to get AOA, etc). The proposed labels and tables do not have consistent structures. The data products are high level and will support the scientific goals of Huygens. Scientifically-useful data products from the DTWG can be reproduced by users based on the archived data only if all the input data is clearly specified and archived somewhere.

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?

Possession of AD.3 and AD.4 are essential for this review. They were not included on the CD. I used **huygens.oeaw.ac.at/Papers/HUY-DTWG-IF-0001_v78.pdf** as AD.3. I used DTWG_PhaseA_Report_R4.pdf from the CD as AD.4.

The preamble pages preceding Section 1 contain a couple of typos, but are understandable and complete. Section 1 notes that AD.3 is very important and that "AD.3 is therefore planned to be provided as part of the DTWG archive volume". This is non-binding. It should state that AD.3 WILL be provided.

Since the FDF files of Section 2 will not go to the PDS, they should not be discussed here. Nor will users care that this is the fourth DTWG delivery, since they will have no access to the previous three. The events that define the start and end of the entry and descent phases should be defined - eg, entry starts at 1270 km interface and ends at initiation of chute deploy sequence and descent starts when entry stops and stops at first impact. The probe entry state is very important and deserves to be archived separately, not just as the first line in the reconstruction results. It should be given in EME2000 and also in something like "alt, lat, lon, speed, flight path angle below horizontal, flight path azimuth clockwise from north, time". It should be stated whether the velocity components are Titan-relative or inertial. The sampling rate of the data to be archived should be stated here. I think it was stated as 1 second somewhere deep in AD.3 or AD.4.

Entry Phase Reconstruction in Section 2.2. A reader should be told that a time must accompany the initial conditions. “1270 km above the planet's surface” - Titan's not a planet, say above Titan's surface instead. The AOA discussion that follows is a useful introduction to AD.4. It should be noted that accelerations in the lift direction are not used in the reconstruction, since they (a) should average to zero and (b) don't have a well-determined direction. The iterative process is not quite correct. The first reconstruction of density assuming AOA=0 needs values of CD, and getting values of CD needs an assumed Ma, Re, density, temperature. It is more likely that a first trajectory will be calculated using AOA=0 and a first atmosphere calculated using CD=2. These can then be iterated. Both Ma and Re are needed in this process, the present discussion doesn't mention Re as much as Ma.

The paragraphs on Entry/Descent Phase Reconstruction should be rewritten to improve their clarity. Some phrases are redundant or confusing. The basic content is fine, but these critical paragraphs need to be very clear.

Section 3. “supposed to be the most accurate one”. Please say instead that it “will be the most accurate one.” Or omit any mention of preceding reconstructions, as I mentioned above. Table 2 refers to “vertical (descent) velocity”, Table 3 refers to “descent speed”. Please be consistent and accurate. Is this velocity with respect to the reference 2575 km sphere or to the radar-detected surface?

“The Browse Directory is likely to contain”. Make this statement binding. More files are listed with bullets than are mentioned in the sentence.

Geometry Directory. State that HUY_EVENT... is defined in AD.3. The contents of the other three files are not defined anywhere.

Software Directory. The reconstruction codes would be of great use to all users. I hope they will be archived. Once the DTWG/PSA/PDS has decided whether they will be archived or not, say so. Please avoid non-binding phrases such as “is likely to”. If archived, they will need documentation.

Section 4. Only HUY_DTWG_ENTRY_POS.TAB has a /* DATA QUALITY discussion, the other files do not. Presumably all files should have this. This file has three column 5s. Almost all files declare the wrong number of columns in the OBJECT=TABLE section. Some lines appeared bold in my PDF reader. They should not be bold. Quotation marks are used inconsistently many times. I presume the preferred usage is an opening (angled as \) double quotation mark followed by a closing double (angled as /) quotation mark. I see many lines where opening double quotation marks appear at the start and end (eg ET_EPOCH) and where pairs of mismatched single quotation marks are used (eg ELAPSED_TIME). Seconds and milliseconds are missing from the UTC XXX in ELAPSED_TIME. I do not know what is meant by “The relative time = 00.00 corresponds to”. Does this mean “An elapsed time of 00.00 seconds corresponds to”?

Future users will not be able to validate or test the results of the DTWG for themselves without the following information. (A) Knowledge of which datasets were used in the reconstruction. Plans are discussed in AD.4, but they will not be followed exactly. The archive should contain a file that says “We used files XXX.TAB from HASI, etc, in the reconstruction” where the files referred to are present in other parts of the PDS archive. (B) The aerodynamic database used to obtain the AOA. (C) The Mach number used to get AOA, together with the ratio of specific heats, temperature, and mean molecular mass used to get the Mach number. (D) The Reynolds number used to get AOA, together with the parameters used to get it. (E) The mass, mass loss, and area used to get the temperature. This information is more important than having a copy of the Fortran code.

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)?

It is understandable when read in conjunction with AD.3 and AD.4

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.

Time references used are consistent with Data Archive Plan. EAICD consistent with Data Archive Plan Appendix J.

Second topic: Scientific and technical content.

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

First para of Section 2 says that the DTWG will be given all the information necessary to do X, but does not say that the DTWG will do X. I would like to see something like: The DTWG will reconstruct the position and velocity of the Huygens probe from the entry interface to the surface and the angle of attack during the entry phase only. The sampling rate will be one second. Then some text can follow about how this is done.

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)?

Not in Section 2. AD.3 gives a clear discussion of what measurements (eg HASI x-servo accelerations) the instrument teams plan to provide to the DTWG, but there is no way to determine whether the DTWG results are based on (eg) the HASI results archived with the PDS, subsampled data from the archive, preliminary data from HASI that are inconsistent with what was archived, etc.

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

No, they have said that it is still being discussed (Software Directory). The discussions in AD.3 and AD.4 are very clear and useful, but they refer to present plans. What actually occurs will inevitably differ from the plans. 6.4.7 of AD.3 looks like a very useful auxiliary product file. These results should also be archived.

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?

Are the latitudes geodetic or geocentric or something else? Is the probe (vertical) descent velocity relative to the 2575 km sphere or the radar-detected surface? Why does 6.4.5 in AD.4 suggest that it is relative to the atmosphere? I didn't see that fact mentioned anywhere else.

Will these products support the scientific goals?

Yes.

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

Not applicable.

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

A reference to the plans of Section 16 of AD.3 would be helpful here. The discussion in the EAICD is very brief. Section 6 of kazeminejad&atkinson.pdf addresses validation, but I am concerned by Figure 9 of that paper.

Question #6: Is the geometrical information addressed?

Yes, with the exception of geocentric/detic latitude. A statement such as "the entry interface is 1270 km above the reference surface, which is a sphere of 2575 km radius" would be helpful.

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?

Unlike the other instruments, the DTWG is still developing its "instrument" [reconstruction code] and this will probably continue beyond Jan 2005. This makes it difficult to write documentation for the process. What I have used as AD.3 and AD.4 are very good. kazeminejad&atkinson.pdf complements what I have used as AD.4, being less detailed but easier to read.

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)?

Yes.

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

The choices seem reasonable.

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?

I presume "keywords" means the names and types in the header. The descriptions of each of the five files in their headers are not consistent. The key points are whether the file is (a) position or velocity, (b) entry or descent phase, and (c) Titan-fixed spherical coordinates or inertial EME2000 cartesian coordinates. All three of these need to be specified in the description in the header. I suggest "Reconstructed probe position during entry in Titan-fixed spherical coordinates" as the description of the first file, with position/velocity, entry/descent, Titan-fixed spherical coordinates/inertial EME-2000 cartesian coordinates being altered as appropriate for the others. The first description is "... entry phase parameters", which is much less informative. The PRODUCT_NAME and DATA_QUALITY_ID of the first file are different from the other four.

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

There's no need to say "profile" in the description of each column in the first file. Why do half the columns in the first file say "Huygens probe xxx" and the rest just say "xxx"? There is not much consistency across the five files.

Consistency could be achieved by using:
(1sigma error in...) Altitude above reference surface/west longitude/latitude during entry/descent phase in Titan-fixed spherical coordinates

(1sigma error in...) x/y/z-axis position/velocity during entry/descent phase in inertial EME2000 cartesian coordinates (1 sigma error in...)

and three properties that aren't as repetitive: Measured HASI pressure, Altitude above radar-detected surface, Vertical descent speed

Col 5 in Descent_Vel has the wrong description

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

XXX EAICD	[Minor] List of editorial comment
	What is "archive responsible"? Is it a person or an activity? It is used many

	times. The Issue/Rev No at the top of each page has an unnecessary space before “Issue 2” Distribution List: Huber is at the PDS, not PDA. Tomasko is at University of Arizona, not Arizona. Is Univesite [Fulchignoni] spelled correctly? Many acronyms in this list are not defined on the following page. Some acronyms (eg ADRS, TL) in the A and A list are not used in this document. EAICD/ESA/ESOC entries are in a different font to all others in my PDF reader. HSWT not in list. 1.1 HSWT is a Working Team, not Working Group. 1.2 Data is presented to PDS and PSA, not just PDS. 1.4 AD.4 is missing a closing double quotation mark AD.4 provides, not provide, a lot of... 1.5 Does Atkinson need a departmental affiliation? Should contact personnel provide phone/fax number? 2 Furhtermore should be furthermore 2.2 argon mole fractions as functions, not function, of altitude Under parachute the molecular weight - this doesn't make much sense 3.3 RD.7 should be AD.7 Fig 1. PERS.CAT should be PERSON.CAT? Compare Fig 1 with Fig 19.3 of PDS standards, I don't see an INST or INSTHOST file in the DTWG plan. 3.4 Descriptions of files in parentheses have inconsistent capitalization. Description of DTWG.CAT misses “Working” Browse Directory. Is AOA plot vs time? Is desc_vel plot vs time? Descriptions of these two plots are not consistent with the rest. Document Directory. Change “documents to understand” to “documents needed to understand” AD.3 reference is both before and after the title of this document. PhD thesis is referred to as AP4, not AD4, twice. START_TIME and STOP_TIME in the PDS labels are sometimes labelled as UTC, sometimes as Z. If Z is used, is there any way to make clear that this is a UTC time, since so much else is in ET? I presume that “6” in the DATA_SET_ID refers to the data level, if so then it is consistent with the Data Archive Plan

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