



SOFTWARE ARCHIVE WORKING GROUP (SAWG) REPORT

TODD KING

PDS MANAGEMENT COUNCIL MEETING

FEB. 4-5, 2016

A LITTLE HISTORY

Management did consider archiving software in 2010. The results were:

MC voted 8–0–1 that PDS has no requirement to archive software.

Minutes from the meeting indicate that the intent was to postpone those branches of the PDS 2010 development related to software archiving.

FORMATION OF SAWG

The SAWG was formed to take a new look at software archiving and what it means to PDS.

- Working group kick-off: Jan. 20, 2016 (telecon)

“Lots of enthusiastic ideas flying around.”
(T. Stein)

AND DIFFERENT PERSPECTIVES

Some concerns

- About potential cost.
- How to create an effective software archive.
- Provenance
- Reproducibility
- Impact on other PDS tasks

PRELIMINARY QUESTIONS

1. What is the driver for archiving software in PDS?
2. What does “archiving software” mean?
3. What support commitment do we expect PDS to give?

WHAT IS THE DRIVER?

- Publically funded software exists and should be preserved.

“there are some wonderful software tools out there that many people have dedicated a huge portion of their lives to. There is a strong desire to have that work preserved. The work cost the public lots of money and no one wants to see their work lost.” (Eric Palmer)

- NASA and OMB consider some software to be research data.

OMB AND NASA VIEW SOME SOFTWARE AS DATA

- According to OMB Circular a-100 (revise 11/19/93)

https://www.whitehouse.gov/omb/circulars_a110/

Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples).

- According to NASA Plan: Increasing Access to the Results of Scientific Research (November 21, 2014)

http://science.nasa.gov/media/medialibrary/2014/12/05/NASA_Plan_for_increasing_access_to_results_of_federally_funded_research.pdf

Data are understood to include not only the recorded technical information, but also metadata (describing the data), descriptions of the software required to read and use the data, associated software documentation, and associated data (e.g. calibrations)

- NASA PDART 2015 Announcement – Requires preserving software in NASA's Github.

DOES THIS CHANGE THE REQUIREMENTS ON PDS?

- NASA and PDART have set new expectations.
- Depends on how we define “data”

Does "data" in "Planetary Data System" refer only to "observational data" or to all "research data" (per the OMB and NASA definition)?

- Or is software a form of documentation?

Let's look at existing requirements

PDS4 LEVEL 1–3 REQUIREMENTS

1.4 Archiving Standards: PDS will have archiving standards for planetary science data.

There are also support requirements that are software related:

3.3.2 PDS will provide a capability for opening and inspecting the contents (e.g. label, objects, groups) of any PDS compliant archival product

3.3.3 PDS will provide tools for translating archival products between selected formats

3.3.4 PDS will provide tools for translating archival products between selected coordinate systems

3.3.5 PDS will provide tools for visualizing selected archival products

4.2.1 PDS will define and maintain a set of usability requirements to ensure on-going utility of the data in the archive.

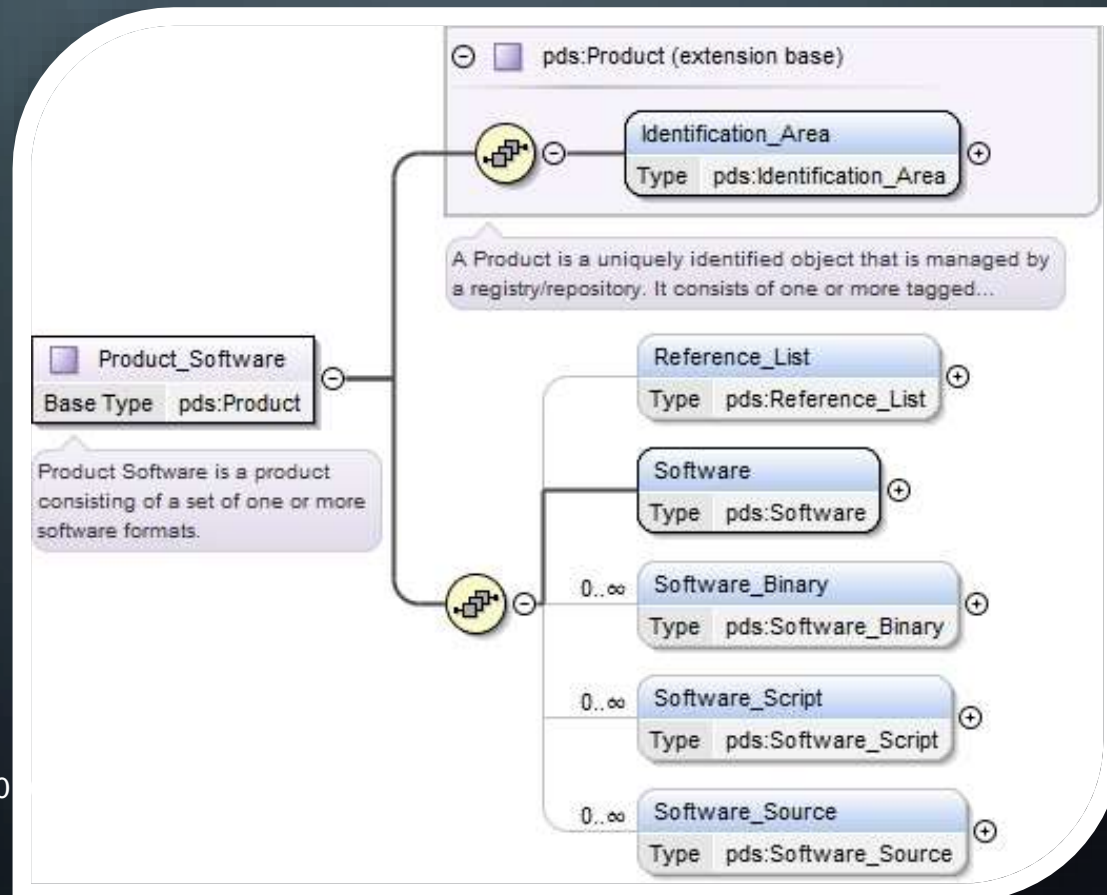
HOWEVER....

- There is a large variety of possible software i.e., processing, visualization, conversion and analysis
- Early PDS volumes included software:
 - imaging 15 out of 90 datasets.
 - Half is decompression source and/or executables. A few have display software and a few have processing software. It is all pretty ad-hoc.
 - PPI 269 out of 779 volumes
 - Mixture of source code, references to “software” volumes, references to web sites, applications.

How might this be handled in the PDS4 world?

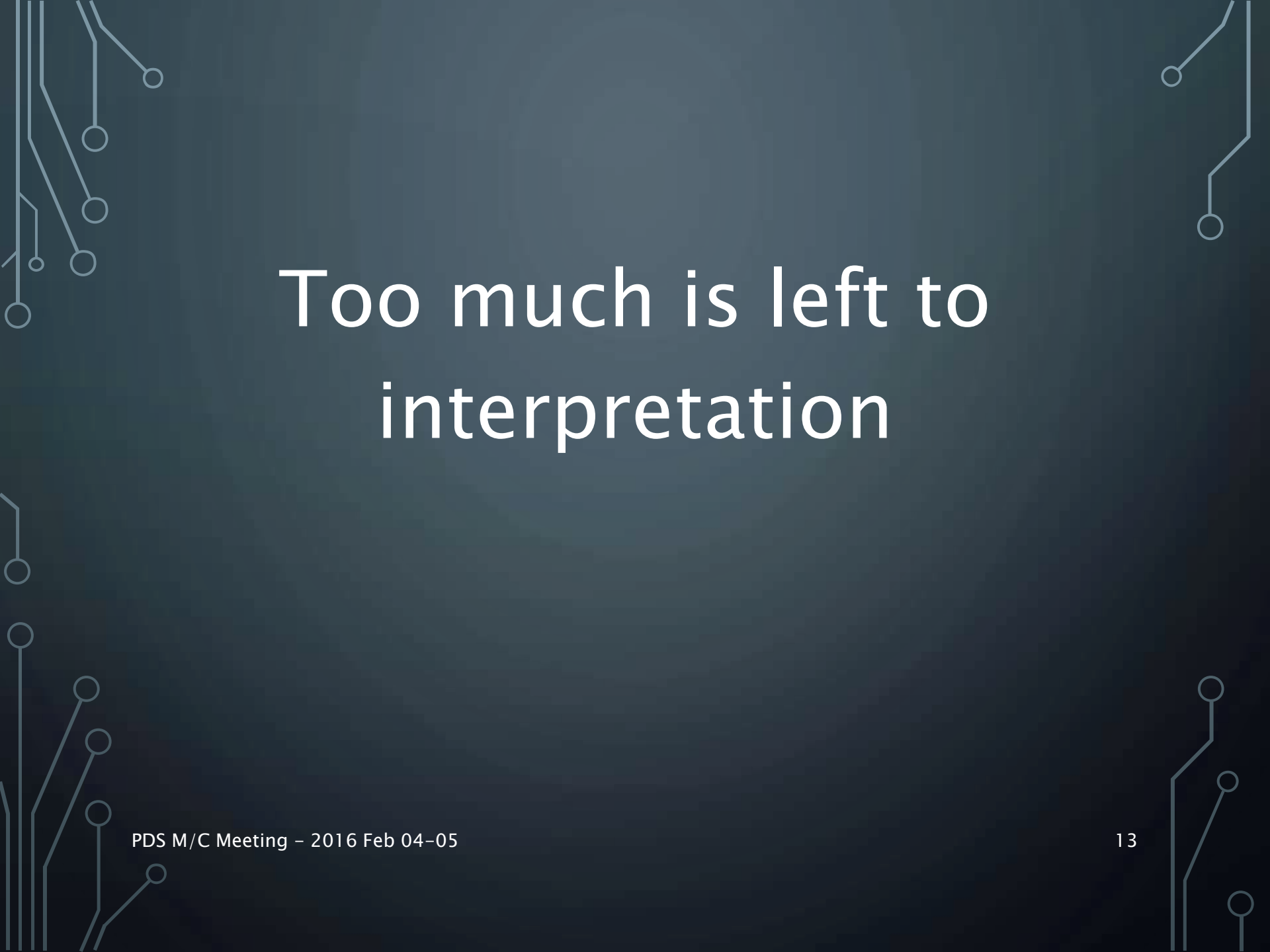
THE ABILITY TO DESCRIBE A SOFTWARE PRODUCT DOES EXIST IN PDS4

PDS4 has Product_Software



... BUT ITS USE IS UNCLEAR

- There is no Software collection type.
- What types of software products do policies allow?
 - The “PDS4 Data Formats” policy could be interpreted to mean only source code (flat* UTF-8 text) is allowed.
 - Software_Binary doesn't fit without changes in the allowed supplemental formats.

The slide features a dark blue background with white decorative circuit-like lines in the corners. These lines consist of straight segments and small circles, resembling a stylized PCB or network diagram. The main text is centered in a large, white, sans-serif font.

Too much is left to interpretation

DISCUSSION

- What is the PDS position on software today? and how do we align with NASA's 2014 policy
- We need to state that position as a formal policy.

If software is allowed we need to:

- Add collection type of “software”
- Define what must be in a software collection
- Define allowable binary software “formats”
- Address reproducibility requirements

BACKUP MATERIAL

PDS M/C Meeting – 2016 Feb 04-05

15

WHAT DOES “ARCHIVING SOFTWARE” MEAN?

- “Assuming that the context is the PDS, then “archiving software” could mean that software is classified and archived as provenance information for science digital objects.” (Steve Hughes)

Should the provenance scope be

1. Describe how the digital object (software) was produced.
(We have Product_Software in PDS4 that can do this)
2. Software as documentation for science digital objects (data)
3. Sufficient information for the digital object (software) to be reproduced (run and used).

Reproducibility is a big question.