National Aeronautics and
Space Administration

PDS 2010 Fundamental Questions

PDS 4 Data Architecture WG

(left)

November 20, 2008

National Aeronautics and
Space Administration

Overview

- **Where we're headed**
  - **Straw man for PDS 4**
  - **Example replacement for IMAGE object**
  - **Labels**

- **Fundamental questions**

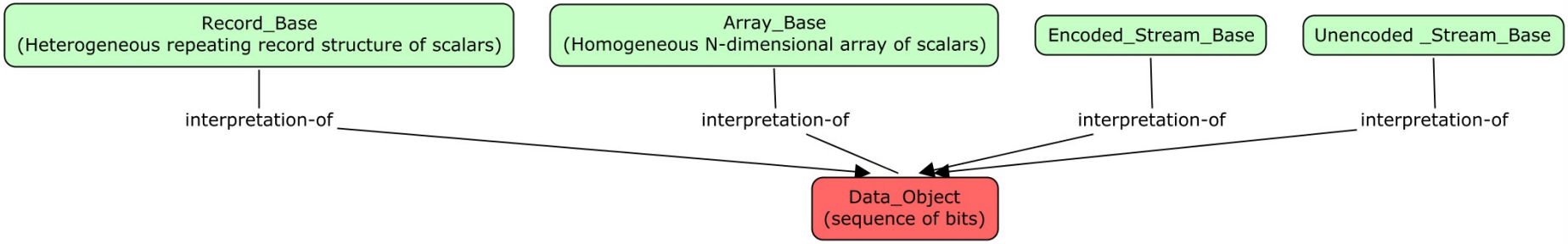
National Aeronautics and Space Administration

Straw man PDS4

- Four simple "**Base Structures**"
  - Used for storage and possibly transfer.
  - Designed independent of interpretation.
  - I/O level – (e.g., 3D Array)
- PDS supported "**Abstract Classes**"
  - What programmers use.
  - Anything beyond being able to read the bytes from the file and storing in the computer.
  - Programmer perspective – (e.g., a banded image)
- PDS supported "**User Classes**"
  - What users use.
  - Scientist perspective – (e.g., false color image)
- PDS Designed Bidirectional Utilities.
  - Conversions involve byte ordering, not alteration of the actual data.
    - Convert between Base Structures, Abstract Classes and PDS supported set of User Classes.

National Aeronautics and Space Administration

PDS4 (base structures) @

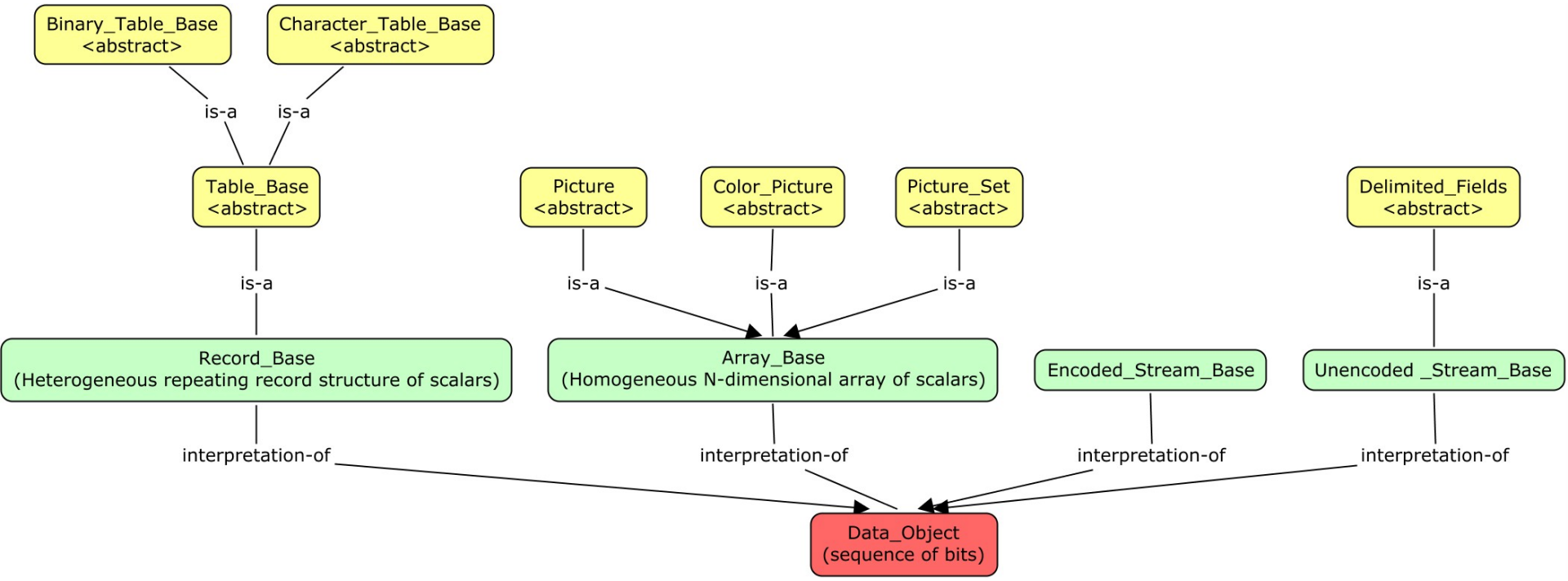


Legend

User Classes
Abstract Classes
Base Structures
Data Object

National Aeronautics and Space Administration

PDS4 (building up) @

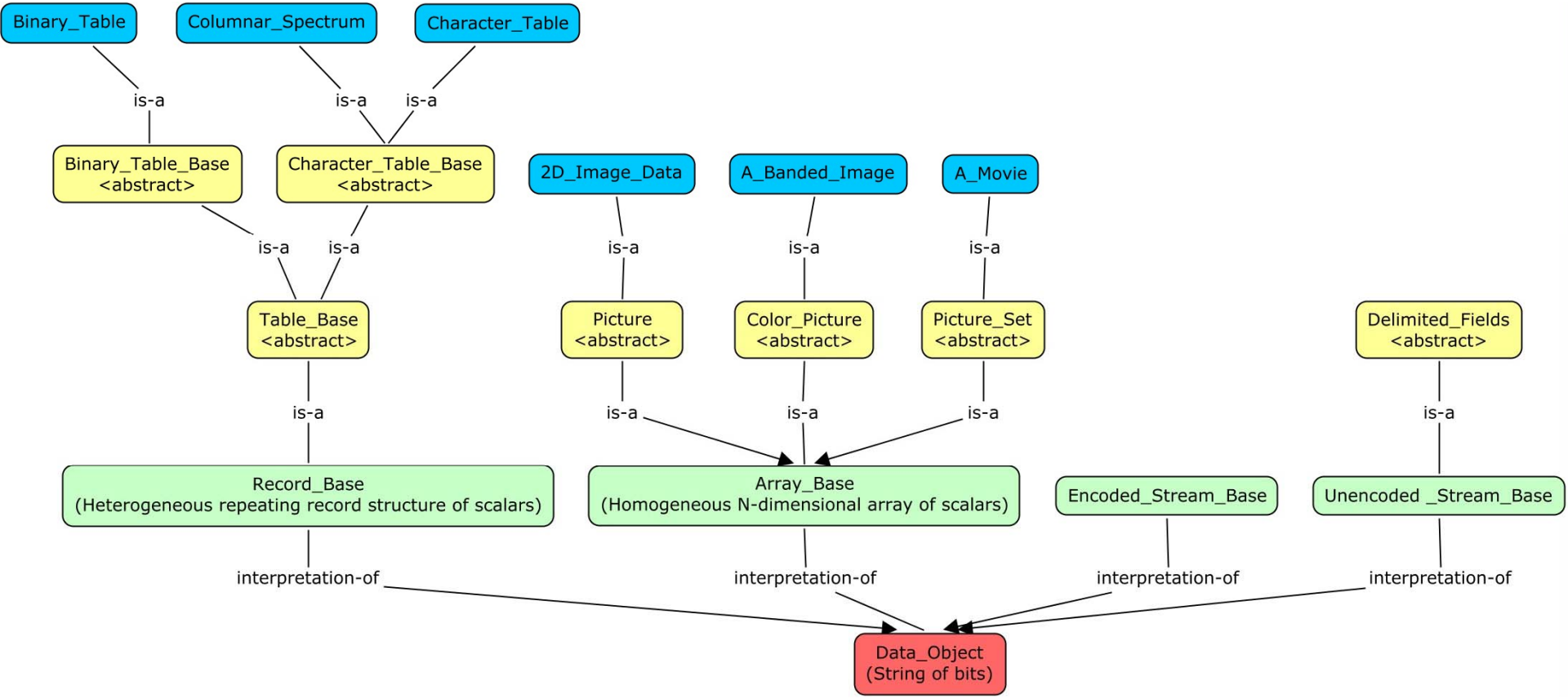


Legend

User Classes
Abstract Classes
Base Structures
Data Object

National Aeronautics and Space Administration

Data – Simple Examples @



Legend

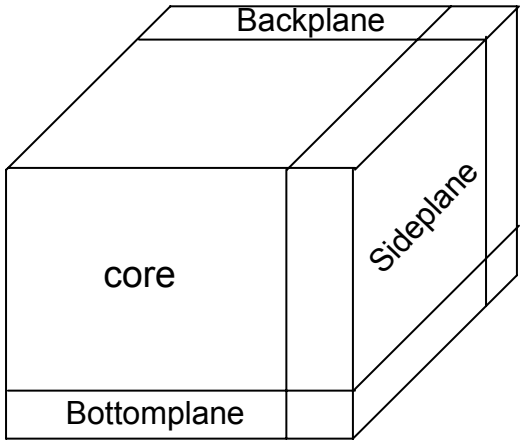
User Classes
Abstract Classes
Base Structures
Data Object

National Aeronautics and Space Administration

Decomposing Interleaved Structures @

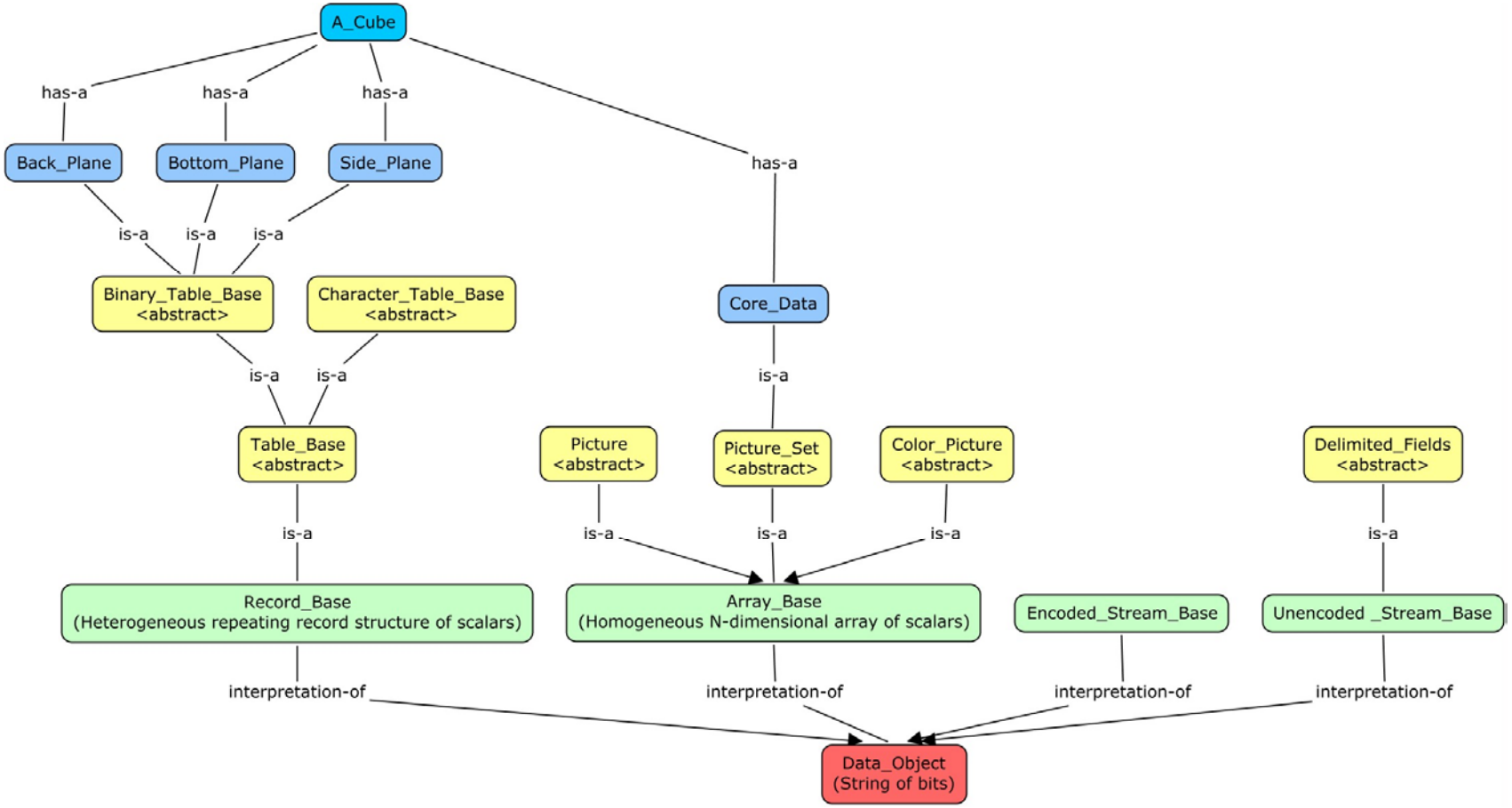
- In current PDS design, the logical data areas that constitute the parts of a SPECTRAL_QUBE object are interleaved with the core data and each other.

- The result is a physically complex file structure that requires dedicated software to read the data and properly extract the interleaved logical sections.



National Aeronautics and Space Administration

Complex Example

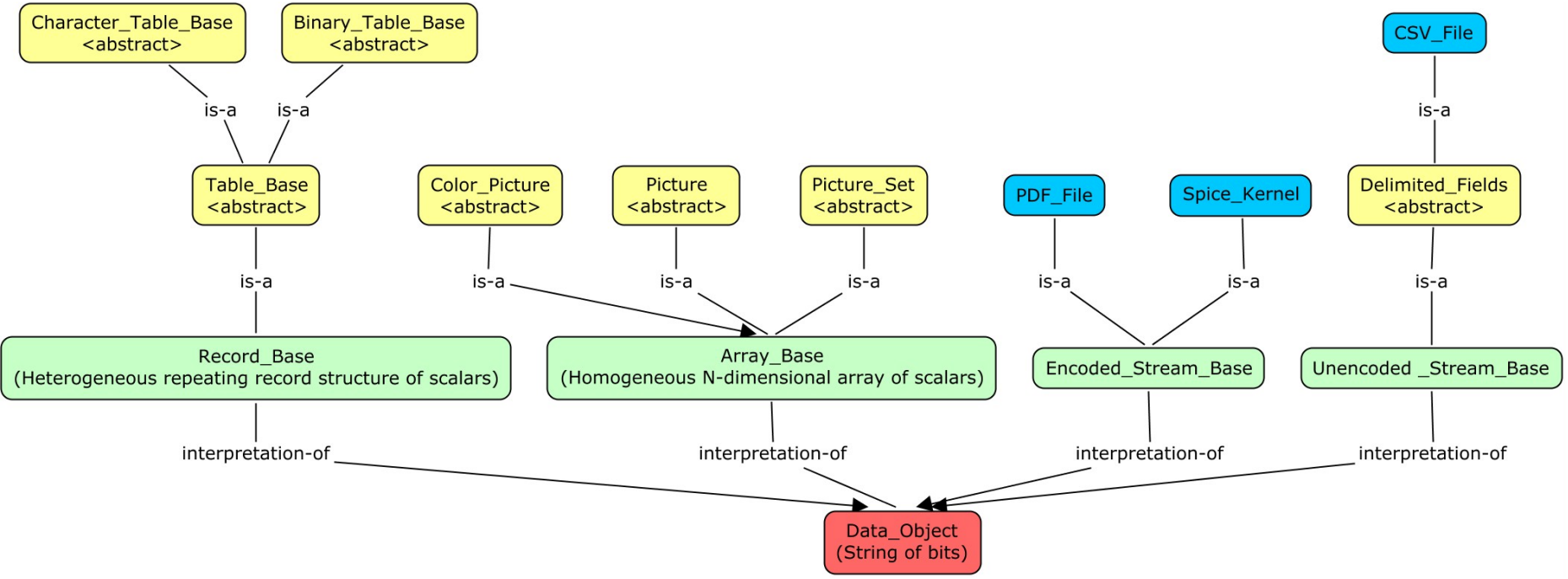


Legend

User Classes
Abstract Classes
Base Structures
Data Object

National Aeronautics and Space Administration

Data – Other Examples @



Legend

User Classes
Abstract Classes
Base Structures
Data Object

National Aeronautics and
Space Administration

Example - Images

"PDS is too complex"

PDS 3 IMAGE object
- 4 required keywords
- 25 optional keywords

BANDS optional keyword allows storage of color images
- There are 18 permutations allowed for RGB images

PREFIX & SUFFIX BYTES keywords allow storing non-image data interleaved with the image data.
- 4 possible ways (line or row)

National Aeronautics and
Space Administration

Example - Labels

PDS 3 labels have very little fixed structure

- **Start with PDS_VERSION_ID & finish with END**
- **Objects have defined beginning & end**
- **No other required ordering**
- **Required keywords mixed with optional keywords**
- **No fixed way to relate relational conditions (if have ‘A’, must have ‘B’)**

National Aeronautics and Space Administration

Example - Labels

- Actual design not done yet
- Have a sample of one possible set of rules & a label
  - Basic idea - labels contain information grouped in Classes which have rigid structural and content requirements.
  - If you use a Class, you must include all of the attributes belonging to that Class (no optional “keywords”).
  - Note (PDS 4 “attribute” ~ PDS 3 “keyword”)

National Aeronautics and Space Administration

Node Consensus or Lack There Of

- During and following the tech F2F, we explored the node technical representatives sense for the desired characteristics for PDS2010.
  - Note: there were some ambiguities and issue overlaps in the survey exercises that influenced some responses.

1. In the broadest terms, there was consensus (but not unanimity) that the two major tasks of the PDS are **archiving data** and **serving the current general user community** - not necessarily in that order.

2. Even among the nodes that agreed to the above, there was a fair amount of broad disagreement regarding the best ways to meet those goals.

3. Much of this disagreement arises from differences in the ways various nodes serve their user communities. In fact, there are a few areas where the priorities of individual nodes seem to be polar opposites of each other.

National Aeronautics and Space Administration

Fundamental Questions – Definitions (2)

archiving

- The process of adding to and maintaining the Archive.

user

- Someone who wants to make use of files in the node data base. This includes, but is not limited to, researchers who were not involved in the original mission or data collection effort.

user service

- Any and all node activities concerned with serving the needs of users. It includes interfaces, hands-on help with locating and interpreting data, and developing utilities to assist in transforming, displaying, or otherwise manipulating the data.

data preparer

- A person or team responsible for creating, submitting, and correcting any liens on data sets intended for ingest into the Archive.

National Aeronautics and Space Administration

Fundamental Questions – Definitions (3)

MC agrees

- There is genuine disagreement among the nodes on some aspects of the PDS mission. The MC is being asked to make a collective and binding decision in a number of these cases.

- Resolving these questions should help the MC decide on the direction forward for PDS2010, and will influence the design decision for both Data and System Architecture.

National Aeronautics and Space Administration

Fundamental Questions

1. Given the preceding definitions, does the MC agree that archiving and user service are the top two priorities of the PDS? (No ordering need be specified at this point.)

2. How long must the data in the Archive be viable?

3. Does the MC require that data be delivered to users in the same format in which it exists in the Archive?

4. Does the MC require that data files in a node data base be in the same format in which they exist in the Archive?

5. Does the MC agree that data preparers must submit data for archiving in the Archive format?

6. If our shoestring budget was cut to half a shoestring and no aglets, in what order would the MC direct the nodes to carry out these activities:

National Aeronautics and Space Administration

Final Observations

- We assume that the move to PDS4 will have impacts on providers, end users, and PDS operations. We need to be sure that it is so much better that everyone will want to learn it.

- The draft PDS4 structures are designed to minimize the long term cost of archiving. Once the system is in place with the translation software written, we can focus most of our time and budget on current users and providers.

- Data providers will have to adjust to the new, more rigid requirements, but will have real prospects for long term stability in PDS requirements which will be much simpler to implement.

- Questions?

National Aeronautics and
Space Administration

backups

National Aeronautics and Space Administration

The Process
Option A – Overhaul PDS 3

- Address all of the items and issues set aside by the specification
  - In order to make the specification unambiguous, anytime there was a conflict or ambiguity, the WG picked something and the rest got set aside.
- Consciously separate the Data Model from the Object Access Library .
- Develop the Data Dictionary Model
- Redefine problematic items
  - data set, volume, document, software, label structure and organization, etc.
- Update the PDS 3 Specification.
- 'Push a button' and let software build a data model from the specification.