National Aeronautics and
Space Administration

PDS 2010 Fundamental Questions

PDS 4 Data Architecture WG

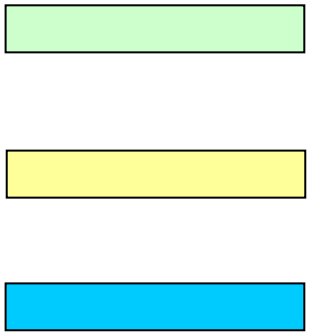
(right)

November 20, 2008

National Aeronautics and Space Administration

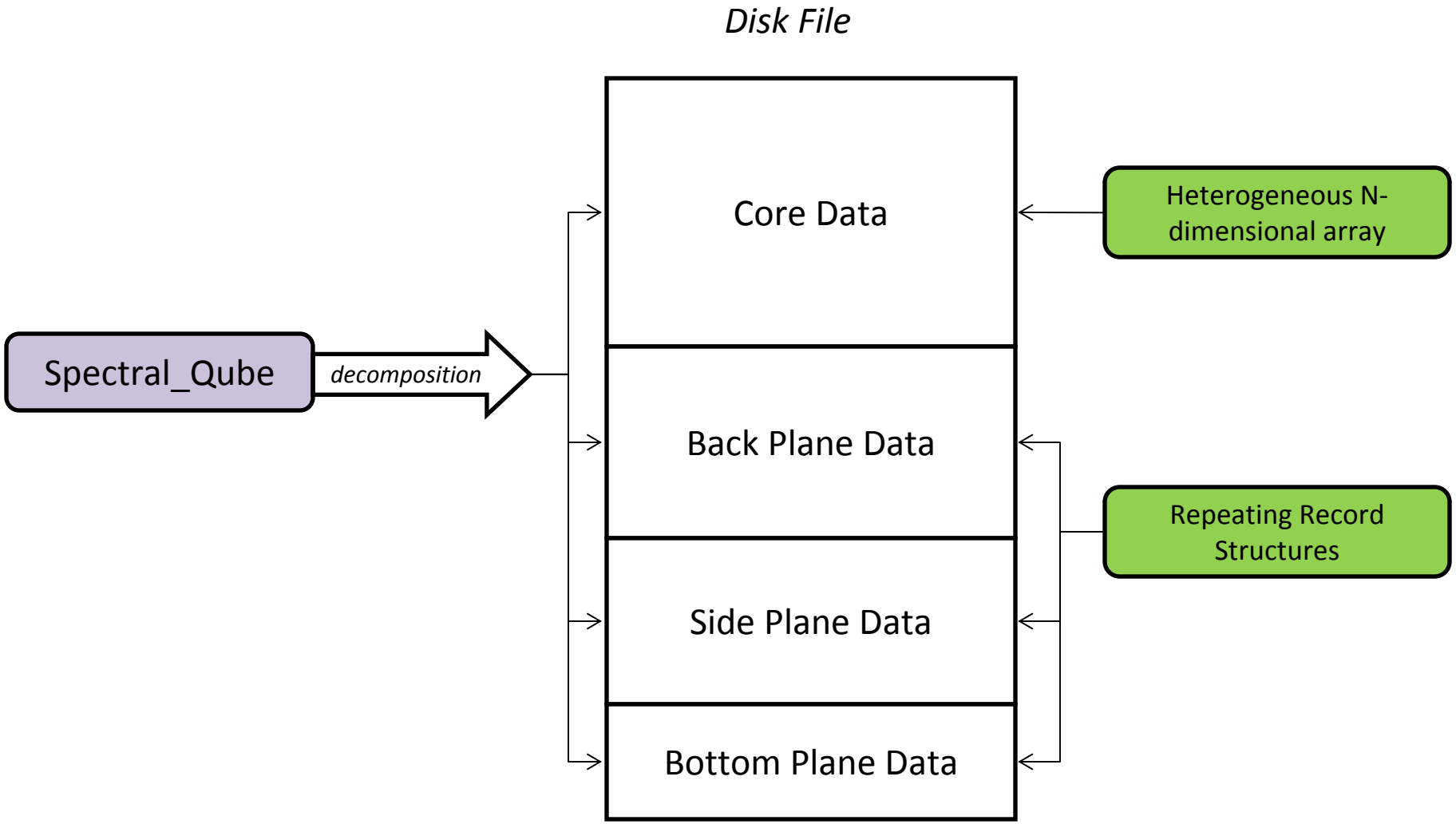
Straw man PDS4

- Four simple "**Base Structures**"
- PDS supported "**Abstract Classes**"
- PDS supported "**User Classes**"
- PDS Designed Utilities.



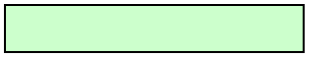
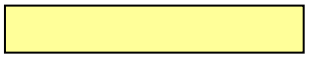
National Aeronautics and Space Administration

Decomposing Interleaved Structures



National Aeronautics and Space Administration

Straw man PDS4

- Four simple "**Base Structures**" 
  - Designed independent of interpretation.
- PDS supported "**Abstract Classes**" 
  - Anything beyond being able to read the bytes from the file and storing in the computer.
- PDS supported "**User Classes**" 
  - What users use.
  - Scientist perspective – (false color image)
- PDS Designed Utilities.
  - Conversions involve byte ordering, not alteration of the actual data.
  - Convert between Base Structures and Abstract Interpretations.
  - Convert to PDS supported set of User Interpretations.

National Aeronautics and Space Administration

Example - Images

PDS 4 will replace the IMAGE object with at least three Classes

- **All will have at least 6 required keywords**
  - **STORAGE_ORDER - of the pixels**
  - **STORAGE_TYPE - correlation between number of bits and data type**
- **None will have PREFIX or SUFFIX BYTES**

The three classes will be:

- **2D_Image_Data (no BAND keywords)**
- **A_Banded_Image**
  - **1 option for storage (what ever we decide, perhaps BAND_STORAGE_TYPE=sequential, BAND_SEQUENCE=rgb).**
- **A_Movie**
  - **A stack of images, (e.g., a time series of 2D images)**

National Aeronautics and Space Administration

Example - Labels

Required Identification Elements
Required Catalog Reference Elements
Node-Specific Elements
Mission-Specific Elements
Dataset-Specific Elements
Object Definition

National Aeronautics and Space Administration

Fundamental Questions – Definitions (1)

The Archive

- The collection of data sets that have passed review, completed editing and been accepted by a PDS node for permanent preservation.

node data base

- The collection of files and information used by a node to serve requests from contemporary users.

viable

- An Archive data set is viable if the information in its files can be recovered with reasonable effort and used in contemporary applications (analysis, display, etc.). "Reasonable effort" may include programming to reorder bytes, but would not include cryptography or detailed research into legacy formats or storage structures.

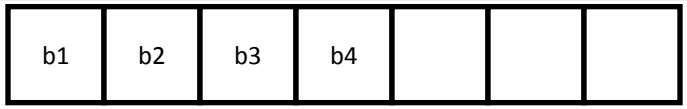
National Aeronautics and
Space Administration

backups

National Aeronautics and Space Administration

Data – A Simple Example

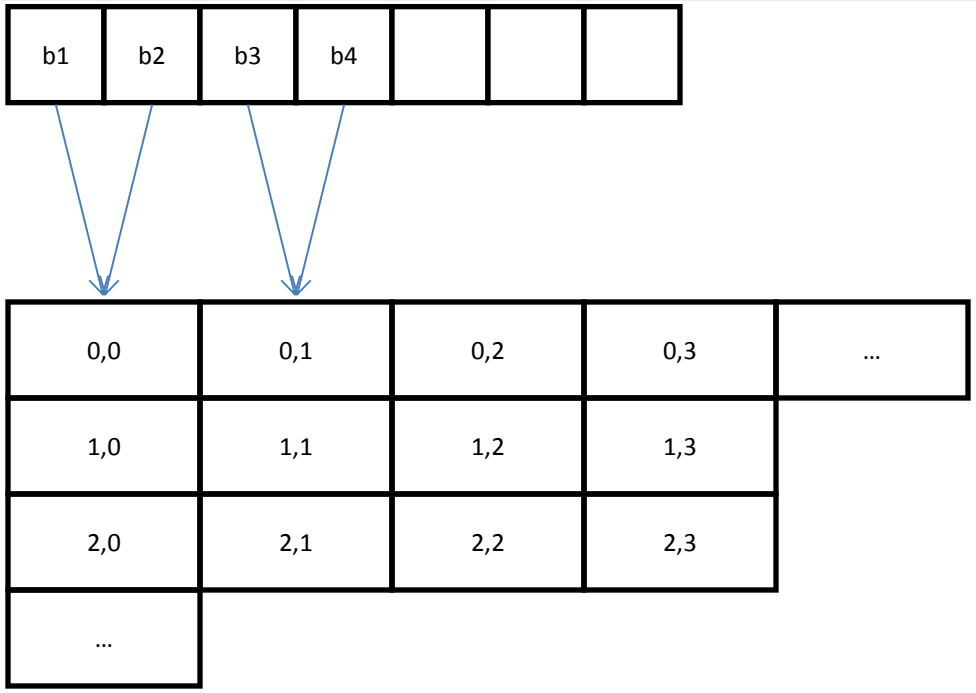
- In the disk file, the data exists as a simple sequence of bytes.
- The lowest-level PDS utilities read the sequence of bytes into one of the basic storage structures.



National Aeronautics and Space Administration

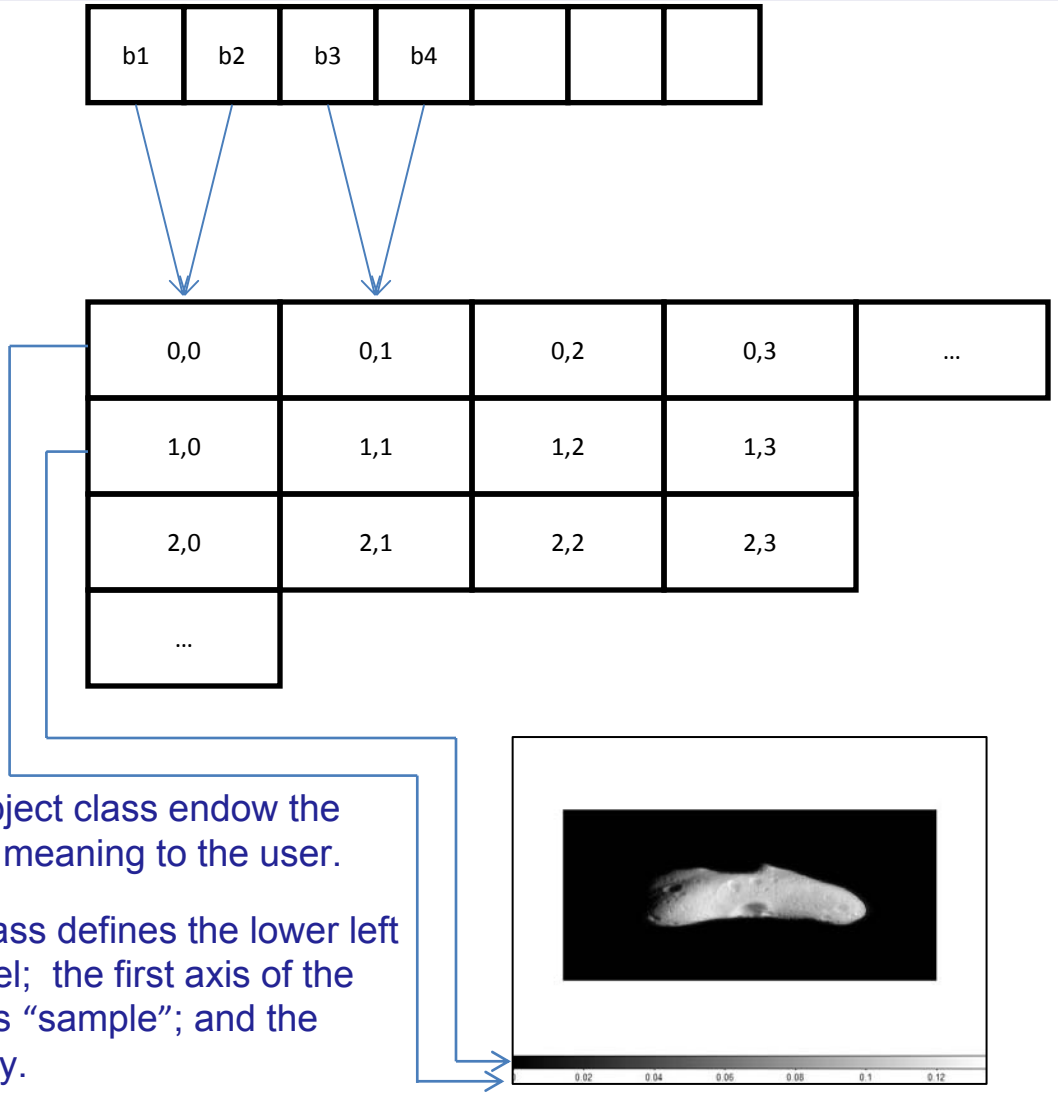
Data – A Simple Example (2)

- In the process of reading the file, the sequential bytes are interpreted into a storage structure in memory.
- In this case our memory storage structure is a 2-dimensional array of 2-byte integers, stored in row-major order.
- This structure and its attributes are defined as the basic n-dimensional homogeneous array (with an element type of MSB-I2).



National Aeronautics and Space Administration

Data – A Simple Example (3)



- The attributes and methods of the object class endow the storage structure and its values with meaning to the user.
- In this case, the 2-D Image object class defines the lower left corner as the location of the (0,0) pixel; the first axis of the structure as “line”; the second axis as “sample”; and the element value as a greyscale intensity.

National Aeronautics and Space Administration

Examples of Differences Exercise 1

- While most nodes listed Archiving and current users support as the top priorities, two nodes considered helping data submitters to be a top priority, and two (different) nodes considered facilitating PDS operations to be a top priority.
- Half the nodes responding considered archiving data in simple formats to be the top priority - the other half rated this at zero.
- One node considers "one-stop-shopping" to be a high priority. Only one other node even placed this in the top 10.
- On the subject of archiving in "easy-to-use" formats, three nodes rated this as very high priority; the rest rated it at the lowest priority.
- One node rated "easy data submission" at a very high priority; most rated it around the middle, but three nodes rated it 2 or lower.
- One node rated archiving in contemporary formats as its highest priority. Only one other node included this in its top ten, and rated it only 1.

National Aeronautics and Space Administration

Examples of Differences Exercise 2

- While most nodes indicated a preference for strictly centralized data structures, two nodes indicated a preference for the possibility of node-specific structures.
- While most nodes indicated a desire for the possibility of node-specific metadata, three nodes leaned more heavily towards most metadata being centrally defined.
- Five nodes indicated strong preference for the standards to be globally defined, but three nodes indicated a desire to accommodate node-specific supplements (or extensions, if you prefer) to the common standards.

National Aeronautics and Space Administration

The Process Option B – From First Principles

- Flesh out the Product Model
  - data set, volume, document, software, label structure and organization, etc.
- Develop the Data Dictionary Model
- Identify the PDS supported “interpretation objects”
- Write a Specification
- ‘Push a button’ and let software build a data model from the specification.