National Aeronautics and
Space Administration

PDS 4 Data Architecture
Part II (1)

PDS 4 Data Architecture Team

September 2008

National Aeronautics and Space Administration

Leftover Fundamental Issue

- **Need clarification of MC requirement for "fewer, simpler formats"**
  - By the end of meeting (***not the end of this slide***), we need to be able to tell the MC what we hope they meant.
  - Referring to data file?
    - Fewer & simpler data storage formats
      - (e.g., number of different types of integer)
  - Referring data file label?
    - Fewer & simpler types of PDS data objects
    - Fewer options within PDS data objects

National Aeronautics and Space Administration

Data Structure Approach

- **Four options (extremes) – but one goal: an unambiguous, internally consistent data model.**

1. Only use structures defined by recognized external standards (e.g., ISO, FITS, etc).
   - Not necessarily all externally defined structures, but no PDS definitions.

2. Develop a syntax for PDS and data providers to use to define data structure.
   - Accept data in any structure that is adequately defined.

3. Clean-up and prune PDS3.

4. Start fresh

National Aeronautics and
Space Administration

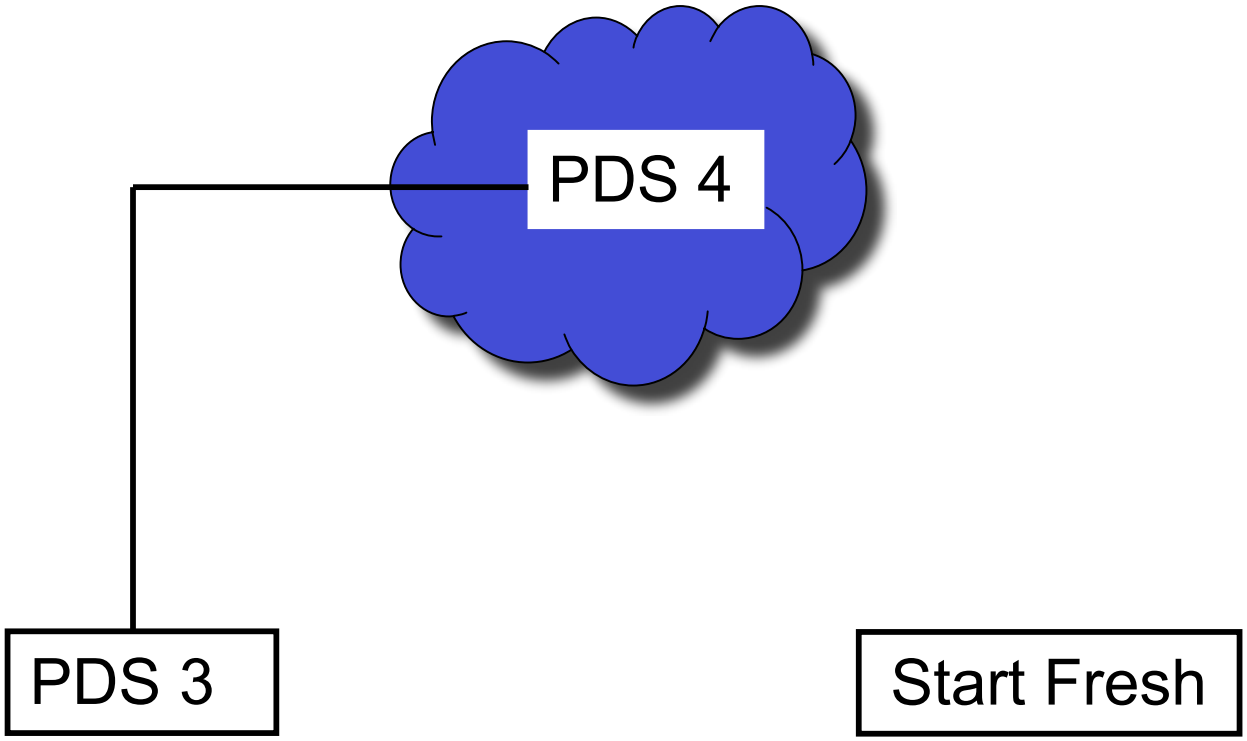
Boldly Going Somewhere @

PDS 3

Start Fresh

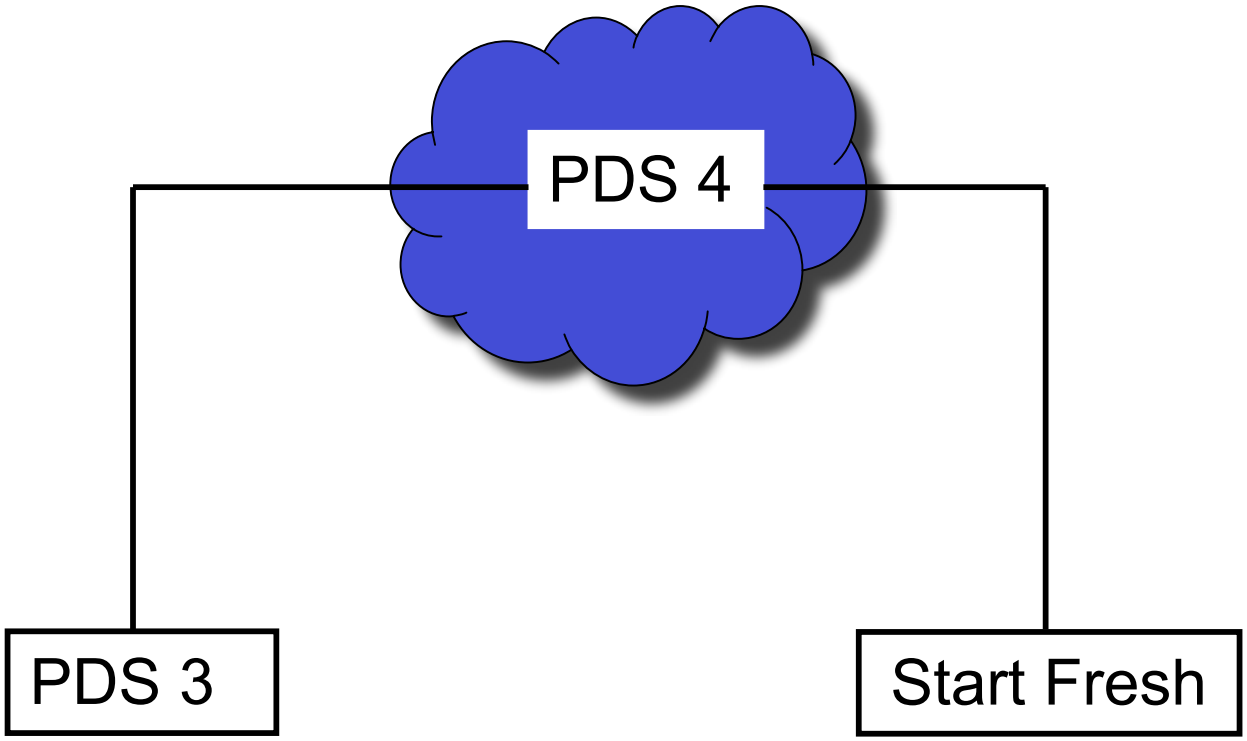
National Aeronautics and
Space Administration

Boldly Going Somewhere @



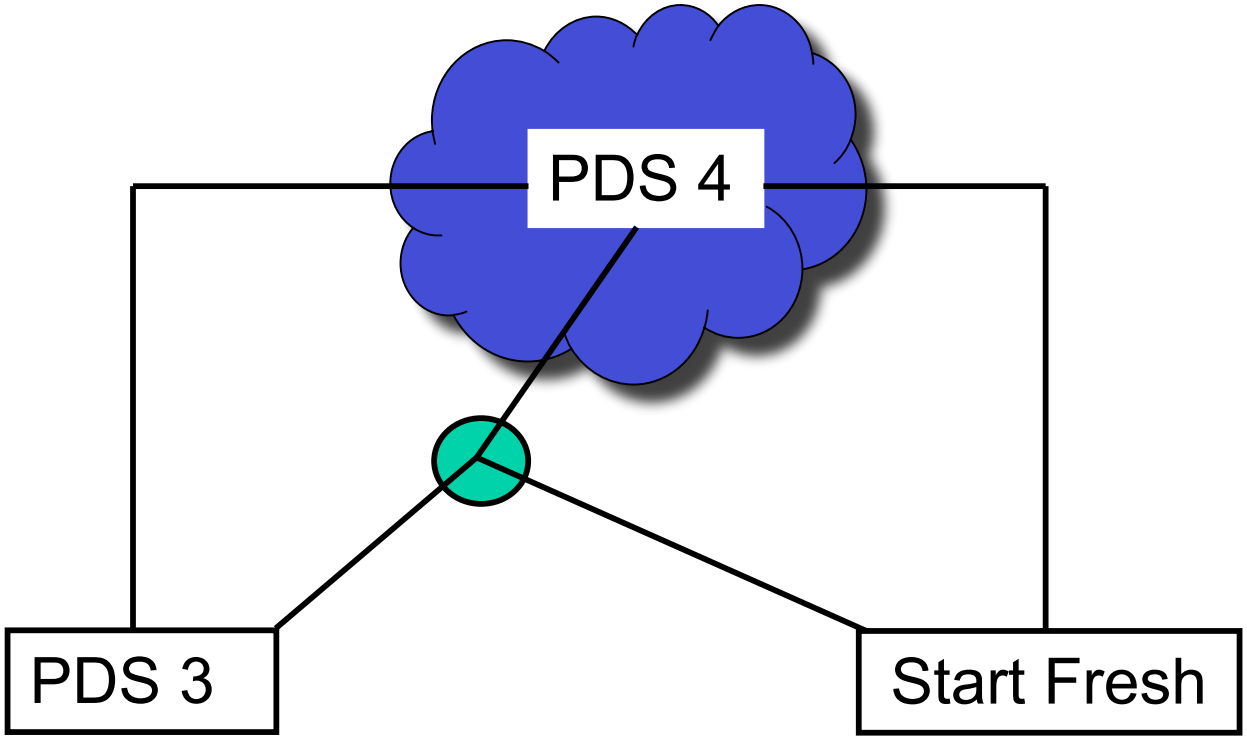
National Aeronautics and
Space Administration

Boldly Going Somewhere @



National Aeronautics and
Space Administration

Boldly Going Somewhere @



National Aeronautics and Space Administration

PDS3 Modification vs. Start Fresh Implications and Consequences

PDS3 Modification

- Do a clean-up and prune PDS 3 (based on the PDS3 Specification)
  - Explicitly define underlying structures & possibly restrict them.
  - Improve consistency, remove ambiguity.
  - Identify and state explicitly any underlying assumptions
- Familiarity - for us and the user communities
  - The user communities are familiar with PDS3 (although many may be annoyed by that familiarity)

Start Fresh

- Develop new structures (skeleton), and graft on pieces from PDS 3
  - In many ways parallel existing structures, some of which are currently implicit rather than explicit; design out the issues that have presented hurdles in the past.
  - 'Cherry-pick' items & aspects of PDS3, avoiding ambiguities and inconsistencies.
  - Avoid unstated assumptions
- Data preparers and data users will have to learn a new model for data structures.

National Aeronautics and Space Administration

PDS3 Modification vs. Start Fresh Implications and Consequences

PDS3 Modification

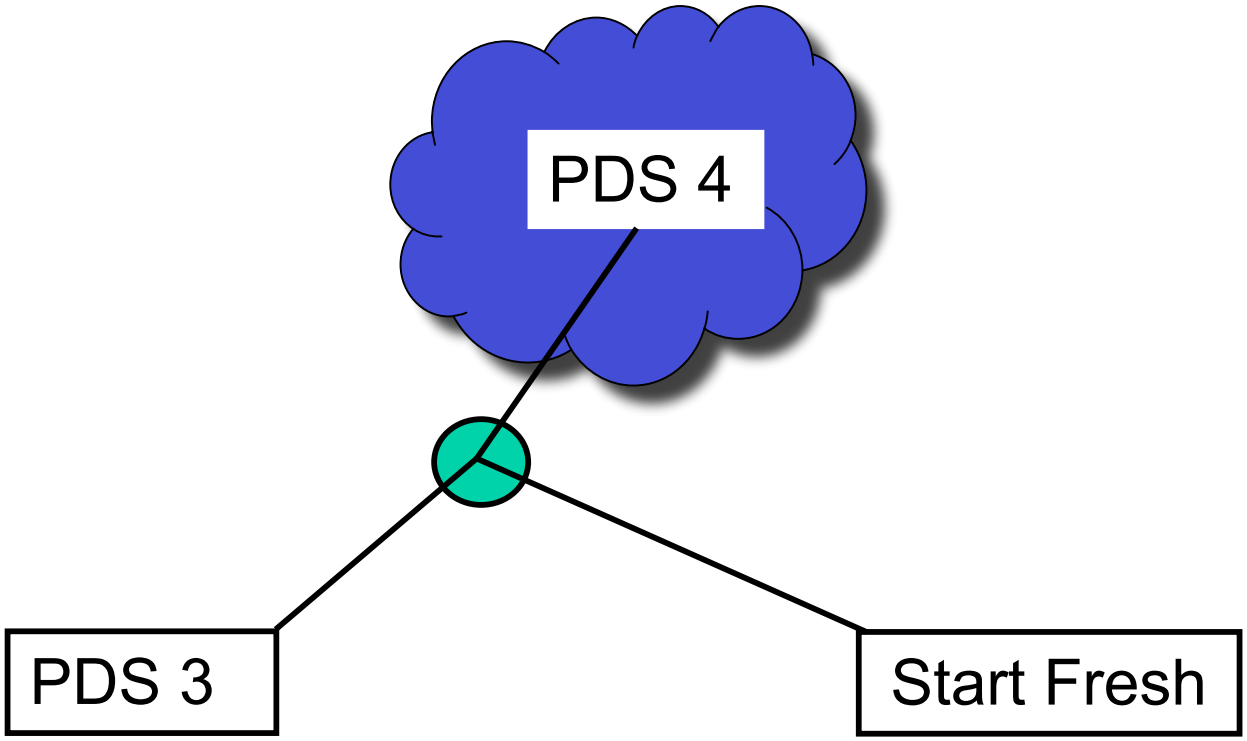
- Could result in fewer data objects and fewer options within them.
- Porting legacy data to the new system would probably be substantially easier.
- Many of the 'broken' items involve mutually exclusive options – they can't be fixed within the boundaries of PDS3.

Start Fresh

- May result in more, but simpler, data objects.
- Can develop a set of structures optimized for whatever the design priorities are (code re-use, processing, display, portability, etc.)
- High probability that at least some of the object structures currently supported cannot easily be supported in the new system.

National Aeronautics and Space Administration

Boldly Going Somewhere @@



- First we need answers and priorities for the big questions
- Need complete assessment of costs & benefits for a couple of options (How far left or right does the green circle shift?)

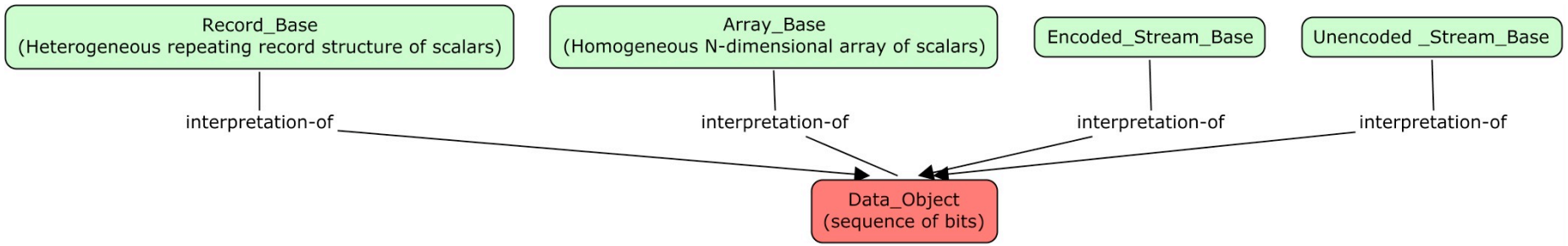
National Aeronautics and Space Administration

Fresh Start – Strawman: A 'Minimalist Approach' @

- Four simple "**Base Structures**"
  - Used for storage and possibly transfer.
  - Designed independent of interpretation.
  - I/O level – (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 – (a banded image)
- 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 Classes and to PDS supported set of User Classes.

National Aeronautics and
Space Administration

Minimalist (base structures) @

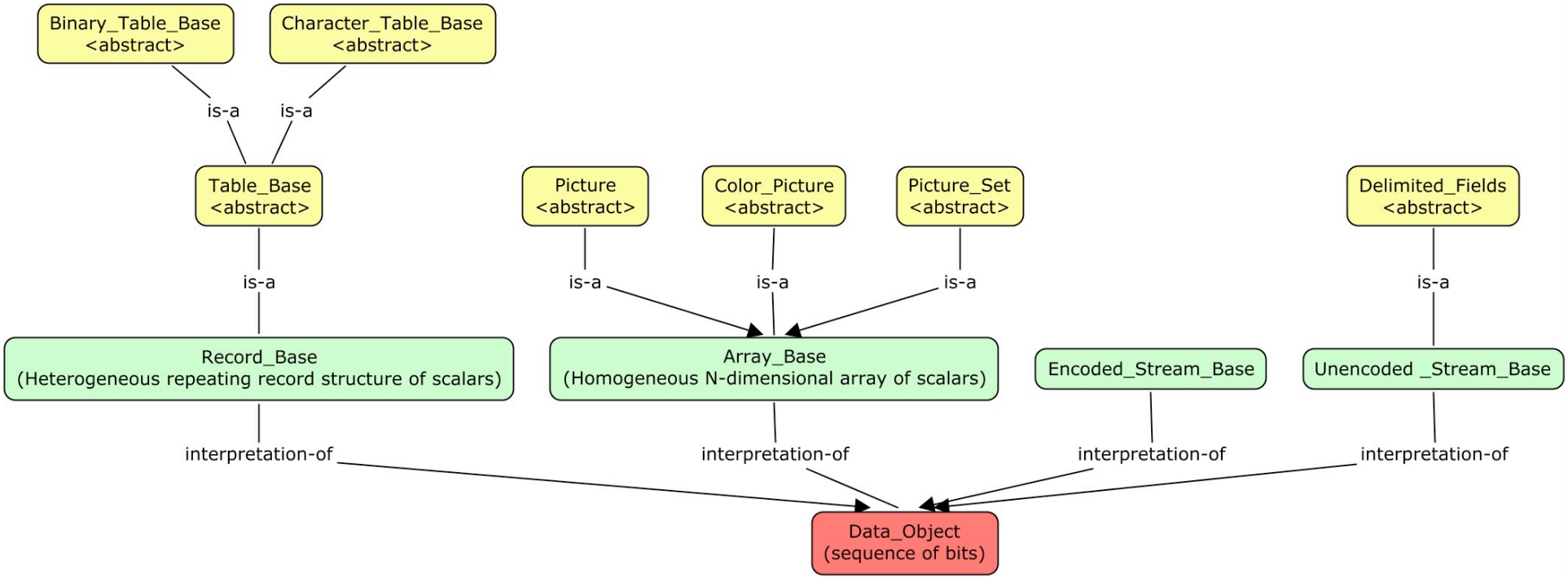


Legend

User Classes
Abstract Classes
Base Structures
Data Object

National Aeronautics and
Space Administration

Minimalist (building up) @

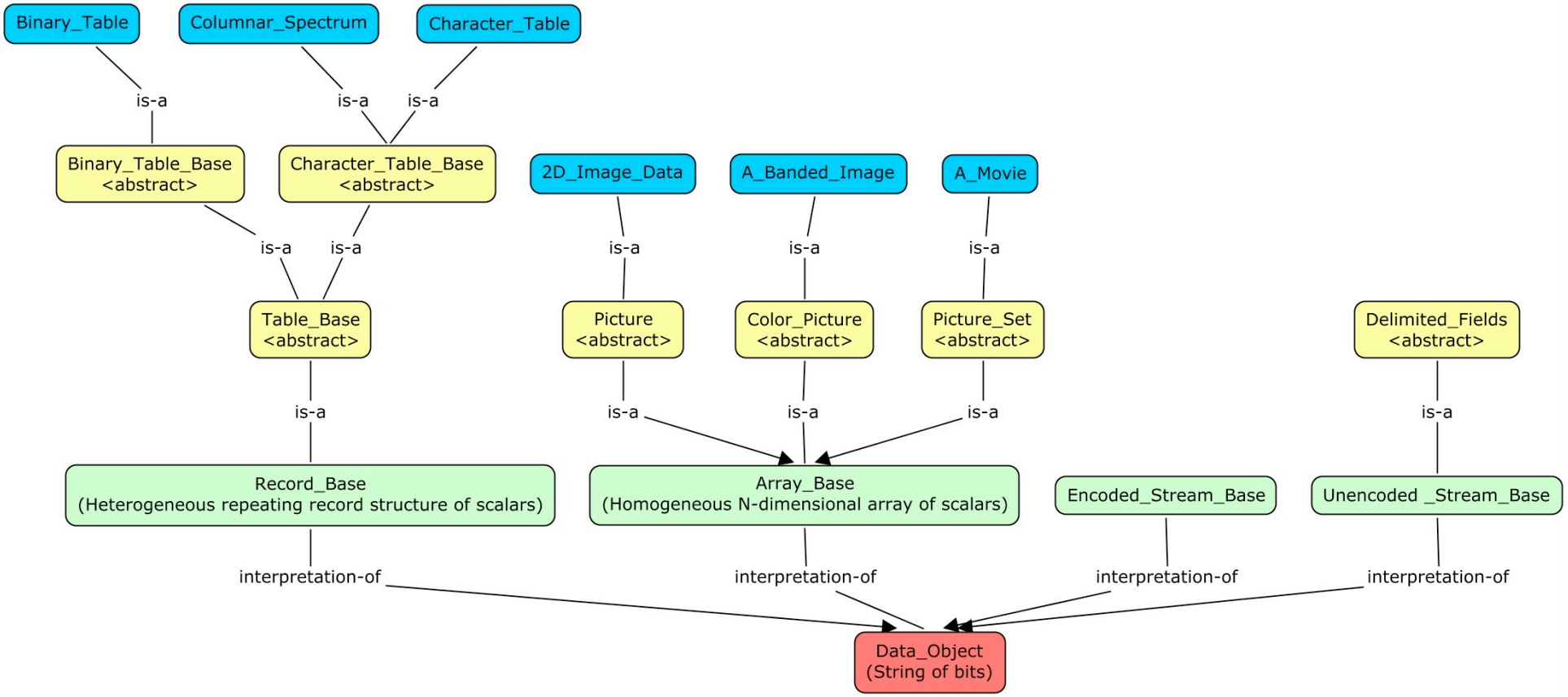


Legend

User Classes
Abstract Classes
Base Structures
Data Object

National Aeronautics and Space Administration

Data – Simple Examples @



Legend

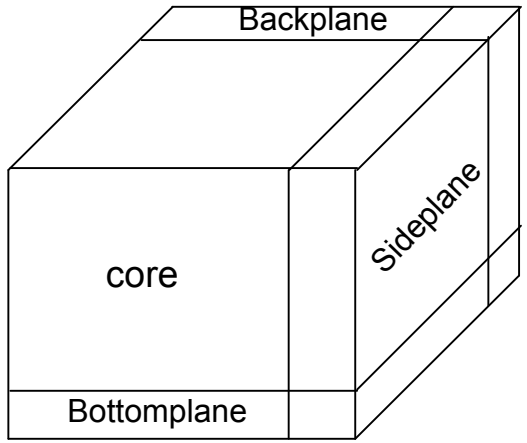
User Classes
Abstract Classes
Base Structures
Data Object

National Aeronautics and
Space Administration

Decomposing Interleaved Structures @

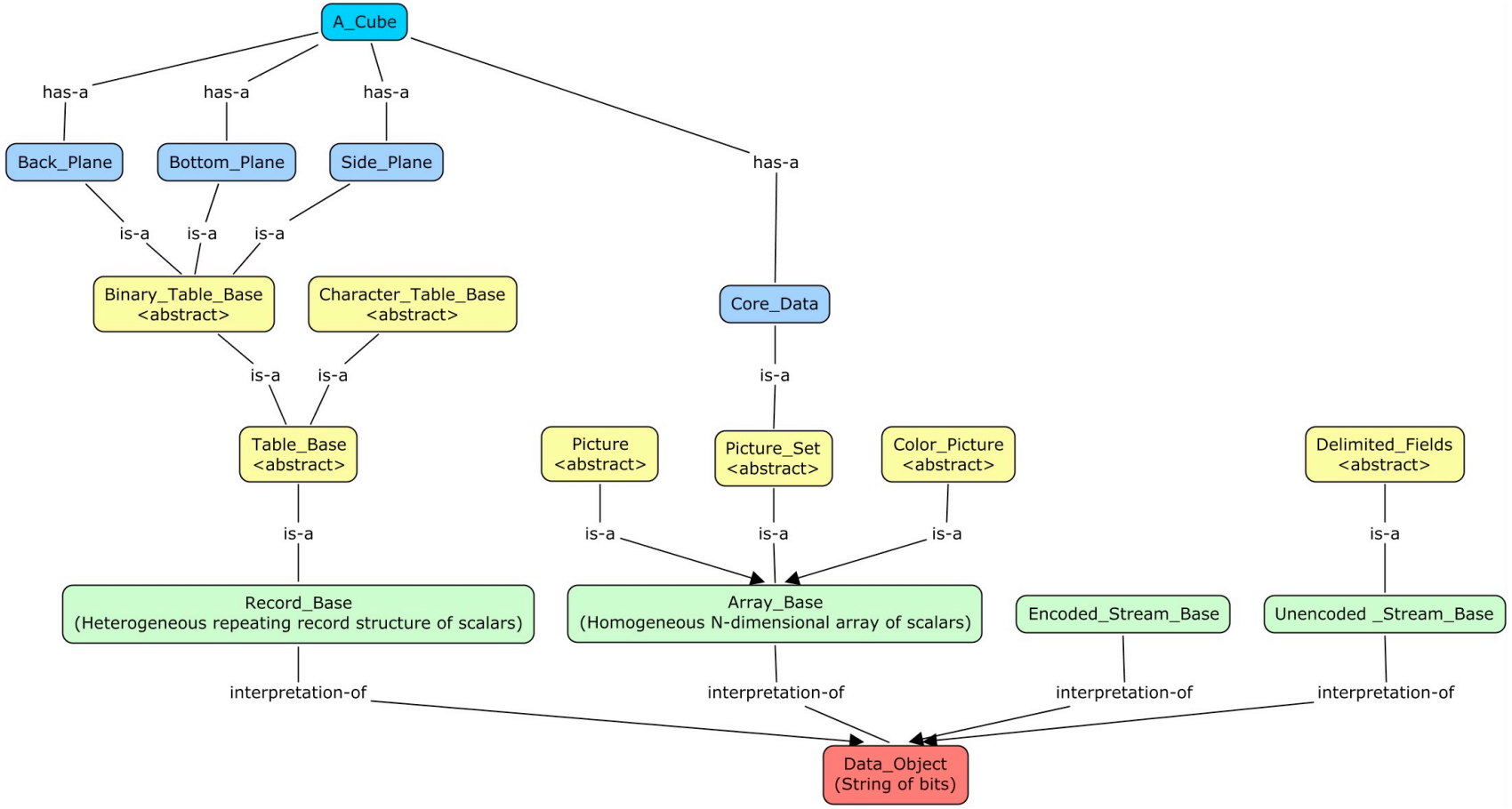
- In current PDS design, the logical data areas that constitute the parts of a SPECTRAL_QUBE object are interleaved with the core data and each other.

- 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

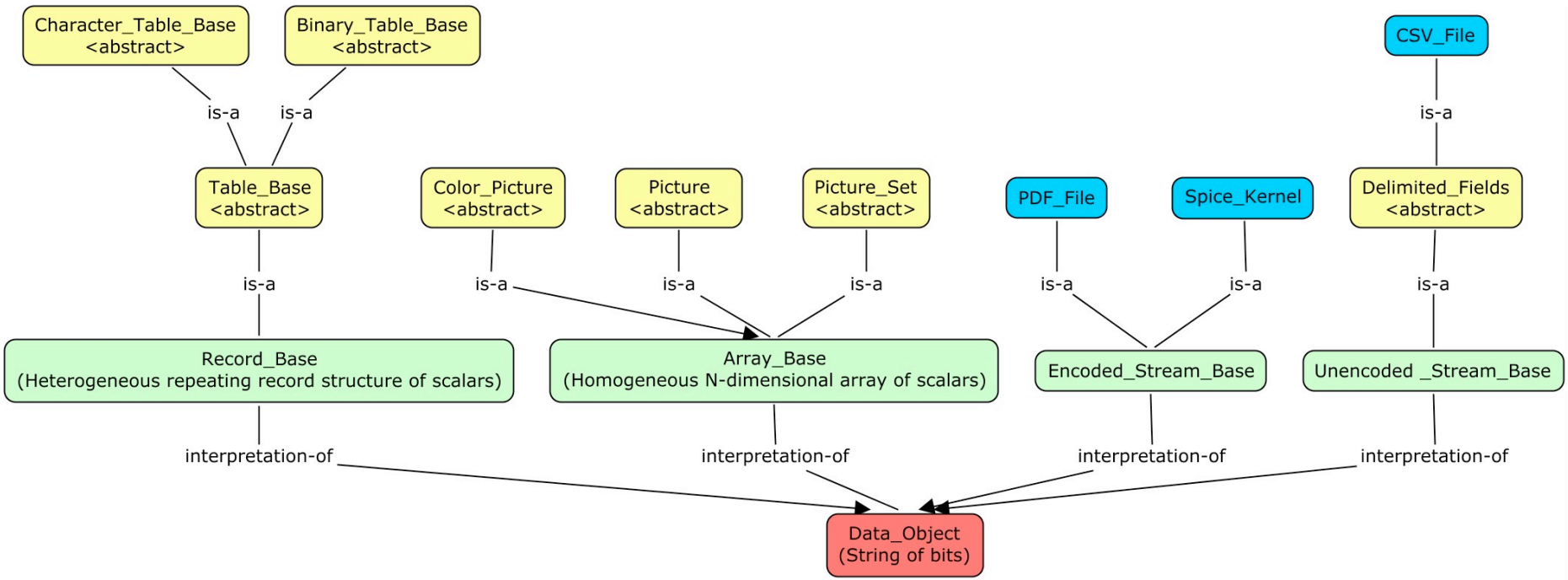


Legend

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

Minimalist Approach

- This approach probably leads to more ‘data objects’ (e.g., Binary_Table and Character_Table), but the objects will be simpler.
- No undeclared underlying assumptions.
- Questions about the ‘Minimalist’ straw man?

National Aeronautics and
Space Administration

Conclusions @

- Whatever approach we take, the process will probably take us to the same place.
  - Which approach (prune or graft) is easier, cheaper?
  - What are the additional benefits of the other approach? Do they make the extra cost and effort worthwhile?

National Aeronautics and
Space Administration

Backup Slides

National Aeronautics and Space Administration

1. Accept structures defined by recognized external standards @

- **Use External Standards.**
  - There are a number of national and international standards bodies (e.g., ISO, ANSI, IEEE, FITS, etc.) researching, writing, reviewing and promulgating standards for pretty much any sort of data we might want to take.
  - As long as a format meets an acceptable international standard, we can merely label it and refer users to the appropriate standard when they want to know intimate details of file structure.
- **External Standards – Pros**
  - Standard formats tend to work with publicly available tools - convenient for contemporary users.
  - We can concentrate our own standards work on documentation and data base integration.
- **External Standards – Cons**
  - Long-term maintenance of the archive becomes an issue if usability is to be maintained.
  - Where there are overlapping standards (for images, for example), we will have to either favor one standard over another, or deal with the issue of users who want all their data of a similar type in the same format.

National Aeronautics and Space Administration

2. Develop a Structure Definition syntax

- **Structure Definition Syntax.**
  - We could define a broad, flexible syntax to describe file structures.
  - Broadest application implies we do not define 'PDS compliant' structures.
  - Data preparers would use this syntax to describe the structure of their files and submit that description as part of the archive.
  - We would then have tools available to convert those descriptions into structure definitions for use in, say, C or Java programs.
- **Syntax – Pros**
  - Porting PDS3 data might be reduced to the problem of merely writing the description of the old data structure in the new grammar.
  - Fewer arguments with mission data providers. We could accept anything they could accurately describe with the grammar. Completely arbitrary data structures could be supported.
- **Syntax – Cons**
  - Completely arbitrary data structures might have to be supported.
  - Probably difficult to write a syntax that could accept a sufficiently large variation in file structures to keep missions happy.