



National Aeronautics and  
Space Administration

# ***PDS 4 Data Architecture Part I (1)***

**PDS 4 Data Architecture Team**

**September 2008**



National Aeronautics and  
Space Administration

## *Discussion Topics*

- **Part I**
  - **Unresolved Fundamental Issues**
  
- **Part II**
  - **More Unresolved Fundamental Issues**
  - **Straw man Information Model**



National Aeronautics and  
Space Administration

## *Fundamental Issues @@*

- **Big Goals.**
    - Who are our customers?
    - What are the drivers we are trying to meet to support them?
    - What are the priorities?
  - **How should the full PDS Data System be organized?**
    - Where should it fall in the range from a ‘Consensus Extreme’ to a ‘Distributed Extreme’?
      - DNs do what all agree to do vs. DNs do what ever they want
- 
- 
- **Clarify MC requirement for “fewer, simpler formats”**
  - **Data Structure**
    - Decide how to get to wherever it is we want to be



National Aeronautics and  
Space Administration

## *Big Goals – an exercise for small groups @*

- **Customers/Drivers.**
  - a) Make science data **available** to the current non-mission science community.
  - b) Make science data **easy to use** (contemporary formats).
  - c) Make science data **submission easy** for Missions and other providers.
  - d) **Preserve** science data in **contemporary formats** (future migrations).
  - e) **Preserve** science data formatted for **long term preservation** (suitable for unambiguous interpretation – simple formats).
- **Small group exercise – phase 1 (3 Minutes)**
  - What is missing from the list?



National Aeronautics and  
Space Administration

## *Big Goals – an exercise for the group @*

- **Customers/Drivers.**
  - a) Make science data available to the current non-mission science community.
  - b) Make science data easy to use.
  - c) Make science data submission easy for Missions and other providers.
  - d) Make science data usable to science community in the future.
  - e) Preserve science data for future scientists.
- **Small group exercise – phase 2 (10 Minutes)**
  - Prioritize (wait for the instructions)



National Aeronautics and  
Space Administration

## *Data System Organization* @

- Consider a spectrum of options from ‘Consensus Extreme’ at one end to a ‘Distributed Extreme’ at the other.
  - Neither end point is likely to be a satisfactory organizational arrangement.
  - Need to identify which aspects of each the full PDS data system should and should not incorporate.



## *Consensus vs. Distributed Extreme – Aspects to use or avoid.*

### Consensus Extreme

- All objects & elements defined globally.
  - may not be altered, amended or augmented in any non-global way.
- One interface into the catalog of PDS holdings.
- Curating Node & physical location transparent to users.
- All local repositories must conform to a single layout.

### Distributed Extreme

- Only highest-level requirements defined globally.
  - Each discipline node develops its own object classes, attributes, data repository structures, and interface procedures with data providers.
- Each discipline runs a catalog for its own archives.
  - Central catalog exists with minimal information to direct users to the appropriate discipline-specific catalog.
- User sees the discipline node as the source, not the PDS.
- Local repositories independent.



## *Consensus vs. Distributed Extreme – Aspects to use or avoid.*

### Consensus Extreme

- Structures & formats highly rigid.
  - Any globally controlled data system will be greatly limited in the number of attributes it can sensibly handle for searching, comparison, and such.
- Global configuration control.
- All general utilities support all PDS4 classes (e.g., PDS image tool must be able to support anything identified as an “image”).

### Distributed Extreme

- High flexibility.
  - Any change the node decided to make that did not violate the minimal Archive requirements would be within its purview.
- Each node provides its own configuration control procedures.
- Any utility to manipulate, analyze or reformat the data will almost certainly come from the node.
  - general utilities will involve things like syntax checking and global standards compliance verification.





## *Consensus vs. Distributed Extreme – Aspects to use or avoid.*

### Consensus Extreme

- No mission or discipline specific tailoring.
- Cheap, universal software solutions.

### Distributed Extreme

- User-Oriented Service.
  - Each node would be in a position to tailor its products, catalog and services to its specific user community.
- Limited interoperability.
  - Unless two or more nodes specifically set out to coordinate them, attribute definitions, catalog interfaces, and software tools from one node would not be applicable to/useful on data from a different node.



## *Consensus vs. Distributed Extreme – Aspects to use or avoid.*

### Consensus Extreme

- "1 PDS".
  - Fewer arguments with data providers.

### Distributed Extreme

- Requirement Proliferation.
  - Each node will have an additional set of archiving requirements over and above the PDS baseline requirements, developed to support their user community.
  - Fewer arguments internally.
  - Mission Annoyance. A large mission providing data to more than one discipline node will likely find itself in the position of having to follow what will seem like two different sets of standards for generating data products.



National Aeronautics and  
Space Administration

## *Archive Organization @*

- **Group exercise – 15 minutes – work as individuals & small groups using slides 7 – 10 (summarized on the other screen):**
  - Make additions to the slides.
  - From your slides prioritize what you want to see in PDS 4.
  - From your slides prioritize what you do not want to see in PDS 4.
- **Discussion**



National Aeronautics and  
Space Administration

*End of Part I*

***Well that was fun –***

***@***