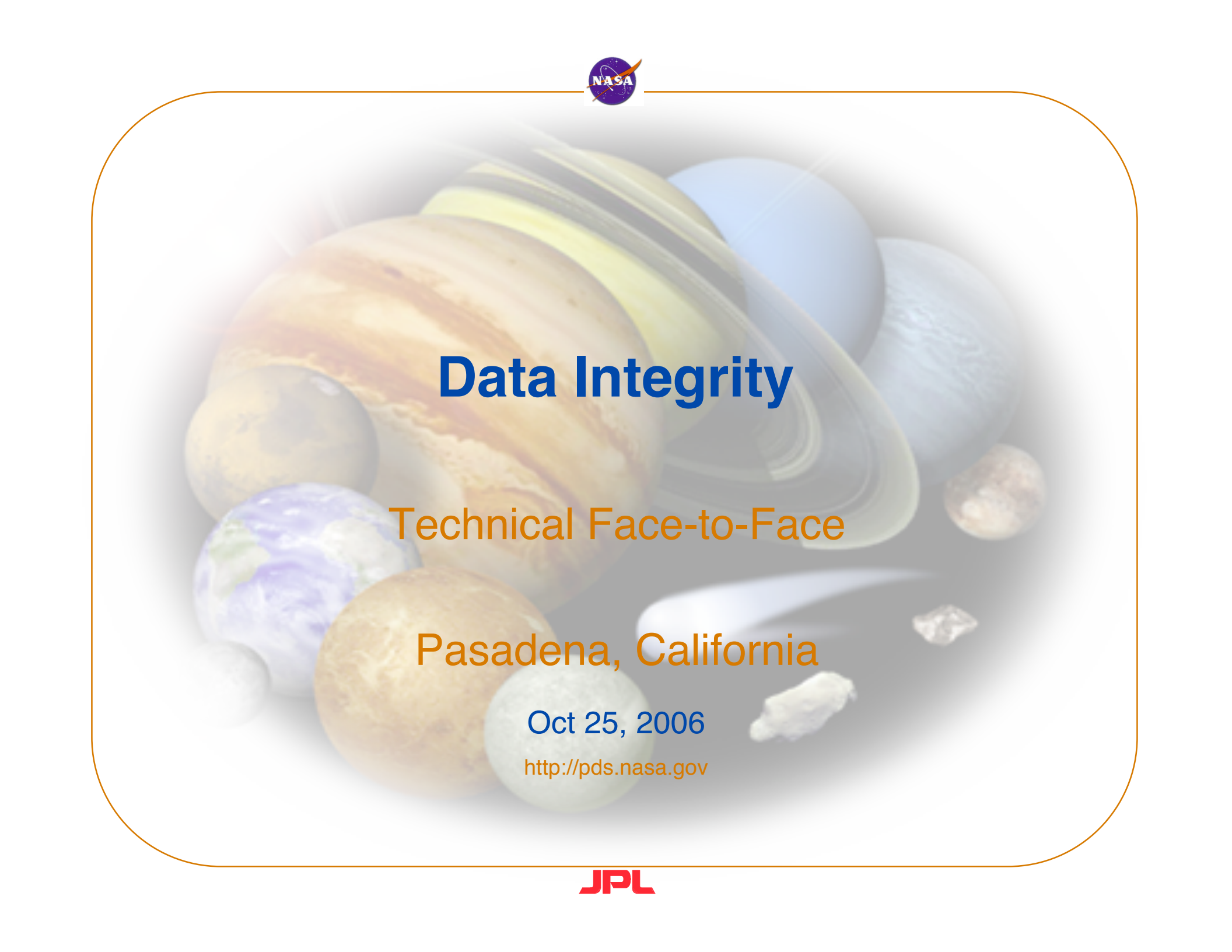

Data Integrity

Technical Face-to-Face

Pasadena, California

Oct 25, 2006

http://pds.nasa.gov

Agenda - Wednesday October 25

Check-in Visitor Control	8:15 AM
Agenda review	8:30 AM – 8:45 AM
Label design tool – continue use case/capabilities discussion	8:45 AM – 9:30 AM
Label design tool – requirements identification and discussion	9:30 AM – 10:30 AM
BREAK	10:30 AM – 10:45 AM
Label design tool – Summary actions, decisions, timeline and release scoping	10:45 AM – 12:00 PM
LUNCH	12:00 PM – 1:30 PM
Data Integrity WG Report (Use Cases, Requirements, PDS Needs)	**1:30 PM – 3:30 PM**
BREAK	3:30 PM – 3:45 PM
Data Integrity – Implementation planning Policy recommendation Standards/implementation plan	3:45 PM – 4:30 PM
Data Integrity – Summary actions, decisions, timeline	4:30 PM – 5:00 PM
Adjourn	5:00 PM

Topics

- Scope
- PDS Needs
- Data Integrity Working Group Report
  - Use Cases
  - Requirements

Scope

- Management Council Action:
  - Crichton agreed to chair an MC working group on data integrity. New suggested starting with Level 4 requirements before dealing with any implementation issues. The DIWG should have recommendations before the Tech Session tackles the technical issues in data integrity at its proposed face-to-face meeting. The Tech Session can then decide where to go with the existing SCR on checksums, including sub-issues of specificity and process.

- Members
  - Dan Crichton, EN, chair
  - Mitch Gordon, Rings
  - Ed Guinness, Geosciences
  - Bill Harris, PPI
  - Steve Hughes, EN
  - Al Schultz, GSFC
  - Mark Showalter, Rings
  - Tom Stein, Geosciences

Scope Cont…

- A set of use cases and level 4 requirements were developed based on the Level 3 requirements and the end-to-end flow of data across PDS
  - PDS Data Integrity Use Cases (October 11, 2006)
  - PDS Data Integrity Level 4 Requirements (October 19, 2006)

- Scope of use cases/requirements is on file corruption

- NOTE: Data accountability for files and collections is also a critical function necessary to verify the integrity of collections such as volumes. PDS currently plans to address integrity of collections as part of the end-to-end tracking subsystem engineering.

PDS Needs

- Discussion questions to think about:
  - Do we have the right set of requirements captured?
  - Have we addressed PDS' needs for file corruption?
  - What does PDS need in terms of policies, tools, etc?
- Can we develop an implementation plan?
  - What's short term?
  - What's long term?

Summary of Use Cases

- 1. Data Delivery
- 2. Data Distribution
- 3. Transfer to Deep Archive
- 4. Data Node Termination
- 5. Archive Integrity
- 6. Media Migration
- 7. Recover Data from Deep Archive
- 8. Data Transfer between Nodes

Use Case
Use Case #1 Data Delivery

- Description
  - Data files delivered to PDS, the receiver verifies the integrity of the files received.
- Actors/Users
  - Data Provider, PDS Node Staff
- Scenarios
  - Data Provider ensures that data files have not been corrupted
  - Data Provider transfer the data files to a PDS Node
  - PDS Node Staff verifies that received data files have not been corrupted
  - PDS Node Staff finds a corrupted data file and rejects the data delivery

Use Case
Use Case #2 Data Distribution

- Description
  - The PDS makes the data products / volumes available (i.e. distributed) electronically to Data Consumers. (2.7.3, 3.2.3)

- Actors
  - PDS Node, Data Consumer

- Scenarios:
  - The PDS Node ensures that the data file(s) to be distributed has not been corrupted.
  - The PDS Node readies the data file(s) for transfer.
  - A Data Consumer checks that the data file(s) received has not been corrupted during transfer.
  - A Data Consumer requests replacements for those files corrupted during the transfer.

Use Case
Use Case #3 Transfer to Deep Archive

- Description
  - The PDS preserves data products / volumes in long-term storage at the NSSDC deep archive. (2.8.3, 3.2.3, 4.1.5)

- Actors
  - PDS Node, NSSDC

- Scenarios:
  - The PDS Node ensures that the data files, to be transferred as a data transfer package, have not been corrupted.
  - The PDS Node ensures that the created data transfer package file is not corrupt.
  - PDS transfers data products / volumes captured in the data transfer package to long-term storage at the NSSDC.
  - The NSSDC rejects and requests retransmission of a corrupted data transfer package.

Use Case
Use Case #4 Data Node Termination

- Description
  - A Data Node is dissolved and the responsibility for archiving and distributing the data holdings passes to the permanent PDS Nodes. (2.5.1, 2.5.2, 2.5.3, 2.8.3, 3.2.3)

- Actors
  - PDS Data Node, PDS Discipline Node

- Scenarios:
  - The PDS Data Node ensures that the files in its data holdings to be transferred have not been corrupted.
  - The PDS Data Node transfers its data files to a PDS Discipline Node.
  - The PDS Discipline Node rejects and requests retransmission of corrupted data files.

Use Case #5: Archive Integrity

- Description
  - PDS Nodes maintain the integrity of their archives by periodically accounting for submitted data and verifying that no files have been corrupted (2.6.1, 2.6.2, 2.6.3, 2.8.3, 4.1.1, 4.1.2, 4.1.4)
- Actors
  - PDS Nodes
- Scenarios
  - The PDS Nodes ensure that all data files received from the Data Providers are accounted for.
  - The PDS Nodes ensure that the data files on existing media have not been corrupted.
  - The PDS Nodes ensure that they have an accessible backup copy of their archive holdings.
  - The PDS Nodes develop a plan for recovery of missing or corrupted files on existing media.

Use Case #6: Media Migration

- Description
  - The PDS Nodes and the NSSDC periodically verify the integrity of data files stored on a physical media in the repository. Transfer data files to new media to ensure data integrity because a) existing media is failing, b) reduce physical space needs by moving to denser media, or c) reduce cost by moving to denser media. (2.8.3, 3.2.3, 4.1.2, 4.1.3, 4.1.4)
- Actors
  - PDS Node, NSSDC,
- Scenarios
  - The PDS Nodes and NSSDC ensure that the data files on the existing media have not been corrupted.
  - The PDS Nodes and NSSDC transfer data files from one physical media to another.
  - The PDS Nodes and NSSDC ensure that the data files transferred have not been corrupted.

Use Case #7: Recover Data from Deep Archive

- Description
  - A PDS Node determines that integrity of an existing set of archival data may have been compromised. The PDS Node recovers data from the NSSDC (4.1.5).
- Actors
  - PDS Node, NSSDC
- Scenarios
  - The PDS Node requests a copy of the archival data from the NSSDC
  - The NSSDC ensures that the data files to be transferred have not been corrupted.
  - The NSSDC prepares and transfers data files to the PDS Node
  - The PDS Node rejects and requests retransmission of data files corrupted during transfer.

Use Case #8: Data Transfer between PDS Nodes

- Description:
  - A PDS Node transfers data holdings to another PDS Node. (2.7.3, 3.2.3)
- Actors
  - PDS Node
- Scenarios
  - A PDS Node ensures that the data files to be transferred have not been corrupted.
  - A PDS Node readies the data files for transfer.
  - The data files are transferred from the source Node to the destination Node.
  - The destination PDS Node rejects and requests retransmission of data files corrupted during transfer.

Proposed Data Integrity L4 Requirements

- L4.DI.1 - PDS will provide procedures for verifying whether a data submission have not been corrupted. (2.5.1, UC-1)
- L4.DI.2 - PDS will ensure that a PDS Data Node verifies their data submissions are not corrupted prior to submission to a PDS Node. (2.5.2, UC-4)
- L4.DI.3 - PDS will ensure that archival data to be delivered between PDS Discipline Nodes has not been corrupted prior to submission (2.5.2, UC-8)
- L4.DI.4 - PDS will ensure that data submitted from a data provider have not been corrupted (2.5.2, UC-1)
- L4.DI.5 - PDS will ensure that data submitted from a PDS Data Node have not been corrupted (2.5.2, UC-4)
- L4.DI.6 - PDS will ensure that archival data transferred between PDS Discipline Nodes have not been corrupted (2.5.2, UC-8)
- L4.DI.7 - PDS will require a data provider, PDS Data Node, or PDS Discipline Node to resubmit data that have been corrupted during transfer (2.5.3, UC-1, UC-4, UC-8)

Proposed DI L4 Requirements Cont…

- L4.DI.8 - PDS will provide a mechanism to allow a user receiving archival data from the PDS to verify that the data have been successfully transferred (3.2.3, UC-2)
- L4.DI.9 - PDS will periodically verify that data holdings on existing media have not been corrupted based on a schedule required by the PDS Management Council (4.1.2, UC-5)
- L4.DI.10 - PDS will verify that data holdings are not corrupted prior to migrating to another media (4.1.3, UC-6)
- L4.DI.11 - PDS will ensure that data migrated from one media to another have been successfully migrated and not corrupted during the transfer (4.1.3, UC-6)
- L4.DI.12 - PDS will ensure that the archival data holdings, to be submitted to the NSSDC for preservation, are not corrupted (4.1.5, UC-3)
- L4.DI.13 - PDS will request the NSSDC to verify that the archival data holdings submitted to the NSSDC have been received intact (4.1.5, UC-3)
- L4.DI.14 - PDS will ensure that the archival data holdings recovered from the NSSDC are not corrupted (4.1.5, UC-7)

Additional Requirements

- The following requirements were captured during the Data Integrity functions, but were allocated to other requirements areas.
  - PDS will ensure that they have an accessible backup copy of their archive holdings (4.1.4, 2.6.1, 2.6.2, UC-5) [Allocated to disaster recovery]
  - PDS will develop a plan for recovery of missing or corrupted files on existing media (4.1.4, UC-5) [Allocated to disaster recovery]
  - PDS will ensure that all data holdings received from the Data Providers are accounted for (2.2.2) [Allocated to tracking]