Data Integrity

Management Council F2F

Washington, D.C.

November 29-30, 2006

http://pds.nasa.gov

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
  - PDS Data Integrity Level 4 Requirements
  - Both were reviewed at the PDS Technical F2F on October 24-25

- 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.

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

Data Integrity L4 Requirements (review at Tech Session)

L4.DI.1 - PDS will only accept data files that are not corrupted. (2.5.1 UC-1)

L4.DI.2 - PDS will provide procedures for verifying whether a data submission has not been corrupted. (2.5.1, UC-1)

L4.DI.3 - 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.4 - PDS will ensure that archival data to be transferred between PDS Discipline Nodes has not been corrupted prior to transfer. (2.5.2, UC-8)

L4.DI.5 - PDS will verify that data submitted from a data provider have not been corrupted. (2.5.2, UC-1)

L4.DI.6 - PDS will verify that data submitted from a PDS Data Node have not been corrupted. (2.5.2, UC-4)

L4.DI.7 - PDS will verify that archival data transferred between PDS Discipline Nodes have not been corrupted. (2.5.2, UC-8)

L4.DI.8 - PDS will notify 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)

L4.DI.9 - PDS will ensure that a user receiving data from the PDS can verify that the data have been successfully transferred. (3.2.3, UC-2)

L4.DI.10 - PDS will periodically verify that data holdings have not been corrupted based on a schedule determined by the PDS Management Council. (4.1.2, UC-5)

L4.DI.11 - PDS will verify that data holdings are not corrupted prior to migrating to another medium. (4.1.3, UC-6)

L4.DI.12 - PDS will ensure that data migrated from one medium to another have been successfully migrated and not corrupted during the transfer. (4.1.3, UC-6)

L4.DI.13 - PDS will ensure that the archival data holdings, to be submitted to the deep archive (NSSDC) for preservation, are not corrupted. (4.1.5, UC-3)

L4.DI.14 - PDS will request the deep archive (NSSDC) to verify that the archival data holdings submitted to the deep archive have been received intact. (4.1.5, UC-3)

L4.DI.15 - PDS will ensure that the archival data holdings recovered from the deep archive (NSSDC) are not corrupted. (4.1.5, UC-7)

Related 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. (4.1.4, UC-5) [Allocated to disaster recovery]
  - PDS will ensure that all data holdings are accounted for. (2.2.2) [Allocated to tracking]

MC Request

- Policies/directives related to the following:
  - Archival Integrity for PDS
    - Develop a specific policy which covers frequency of verification, data integrity of the bits, tracking of the files, and that the files can still be accessed. For example,

      "Each node is responsible for verifying the integrity of their archival holdings on an annual basis including verification that all files are accounted for, are not corrupted, and can be accessed irregardless of the medium in which they are stored"
  - Disaster recovery planning for PDS
    - Develop a general set of requirements to guide PDS
    - Develop a specific policy. For example,

      "Each node is responsible for defining and implementing a disaster recovery plan for their node which covers catastrophic loss of data and/or system functionality"
- NSSDC/Preservation Planning
  - Need to align NSSDC/PDS MOU with archival integrity policy
  - Need to ensure NSSDC is meeting federal requirements for preservation of data (PDS-4.1.5)
- The WG continue to do the following:
  - Recommend an implementation for data integrity to the tech staff and then the MC
  - Develop requirements for Disaster Recovery and Archive Tracking (incl the tracking system)

Timeline

- Present requirements to MC (Nov 2006)

- Working Group to recommend to MC (Nov 2006)
  - MC to develop policies for archival integrity
  - Revisit MOU w/ NSSDC (WG to provide requirements), preservation policy/planning
  - MC to develop Disaster Recovery Policy, Preservation Policy

- Data Integrity Engineering Process
  - Working Group to identify and evaluate options for meeting the requirements (Jan 2007)
  - Review by Tech Group (Feb 2007)
  - Recommend solution to MC (Mar 2007)
  - Develop Implementation Plan upon MC approval of solution (Apr 2007)