Physical Media WG Report

Physical Media Working Group

August 13, 2007

PDS MC Policies on Media, Integrity and Backup

- Data Delivery Policy
  - Data producers shall deliver one copy of each archival volume to the appropriate Discipline Node using means/media that are mutually acceptable to the two parties. The Discipline Node shall declare the volume delivery complete when the contents have been validated against PDS Standards and the transfer has been certified error free.
  - The receiving Discipline Node is then responsible for ensuring that three copies of the volume are preserved within PDS. Several options for "local back-up" are allowed including use of RAID or other fault tolerant storage, a copy on separate backup media at the Discipline Node, or a separate copy elsewhere within PDS. The third copy is delivered to the deep archive at NSSDC by means/media that are mutually acceptable to the two parties.

    (Adopted by PDS MC October 2005)

- Archive Integrity Policy -
  - Each node is responsible for periodically verifying the integrity of its archival holdings based on a schedule approved by the Management Council. Verification includes confirming that all files are accounted for, are not corrupted, and can be accessed regardless of the medium on which they are stored. Each node will report on its verification to the PDS Program Manager, who will report the results to the Management Council. (Adopted by MC November 2006)

Question presented to MC in March

- Given on-going DVD reliability problems observed in PDS and observed by the industry, should WG evaluate moving to mass storage for archiving of volumes vs DVD?
  - A question was then raised by the MC as to what the node capabilities were in terms of online RAID distribution and whether movement away from DVD would be feasible for PDS.

- Actions
  - Presented to MC. Recommended that data integrity checking be done on regular schedule for both optical and RAID.
  - Presented results at Archives in 21st Century Meeting
  - Performed a survey to understand node capabilities
  - Developed use cases, requirements for archive availability (primary, secondary/backup, deep archive)
  - Defined options for the secondary/backup for consideration by MC and nodes

Process

- Define requirements for data availability
  - Media/Technology Independent
- Conduct a survey of the PDS nodes to determine current capability for online repository management (capacity, operating system, etc)
  - Determine data holdings across PDS that are online vs offline
  - Determine data holdings using media/storage technologies
- Develop options for the “3 copies” of data
  - Compare media management costs (optical, RAID, tape)
  - Understand migration challenges and scope
  - Deliver electronic volumes to NSSDC

Survey Respondents

- The following PDS Nodes, SubNodes, and Data Nodes responded to the Repository Survey:
  - (1) RS -- [RS] Richard Simpson
  - (2) SBN -- [AR] Anne Raugh
  - (3) NAIF -- [BS] Chuck Acton / Boris Semenov
  - (4) ATMOS -- [LH] Lyle Huber
  - (5) GEO -- [TS] Thomas Stein
  - (6) IMG-Flagstaff -- [CI] Chris Isbell
  - (7) IMG-HiRise -- [EE] Eric Eliason
  - (8) IMG-KPL -- [MM] Myche McAuley
  - (9) IMG-LROC -- [EB] Ernest Bowman
  - (10) IMG-Themis -- [KM] Kimberly Murray
  - (11) SBN-PSI -- [CN] Carol Neese
  - (12) PPI-UCLA -- [TK] Todd King
  - (13) RINGS -- [MG] Mitch Gordon

Survey Purpose and Findings

- The PMWG distributed a survey to:
  - Characterize storage infrastructure of PDS nodes, subnodes and data nodes
  - Identify the amount of PDS data held in various media technologies (and online vs offline)
  - Understand current backup capabilities at nodes

- High Level Findings
  - > 95% of PDS data appears to be held in RAID
    - Much has already been migrated
  - 7 out of 13 respondents identified a backup source
  - 12 out of 13 serve data electronically via a primary online RAID repository
    - Many of these are CD, DVD sources with two sites using tape
  - Three copies of data may not exist (or may not exist in any real organized way)
    - Data sets not delivered to NSSDC
    - Data managed online may/may not be backed up at a suitable site
  - Cost to manage storage varies across nodes
    - It's clear the cost is non-linear (I.e. costs drop as volumes increase) for RAID management and it appears to be more linear for CD, DVD and Tape
    - The WG also noted that optical management costs are more linear
  - Large data increases seen in out years (particularly in imaging)

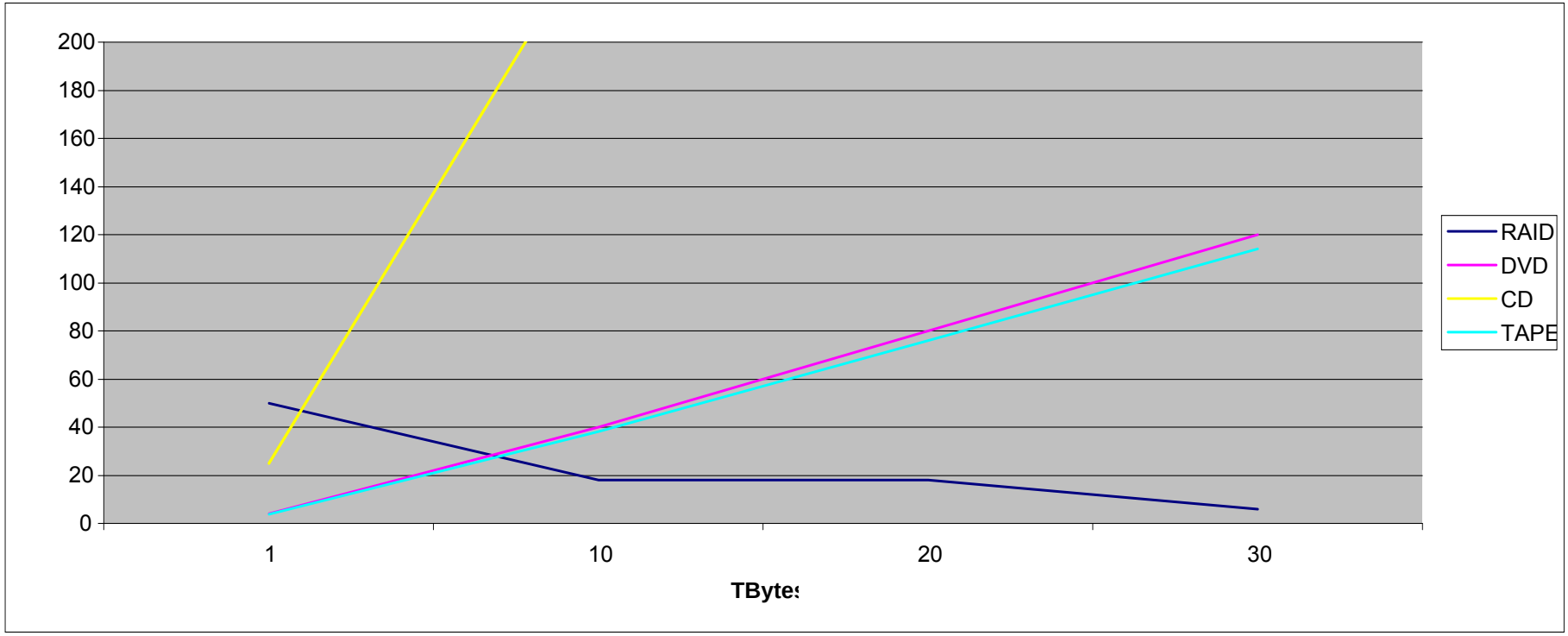
Data Volumes Matrix from Survey*

NODE	CURRENT DATA VOLUME (TB)	CURRENT ONLINE CAPACITY	% ARCHIVE RAID (FY07)	% ARCHIVE OPTICAL (FY07)
RS	1	-	0%	100% (CD/DVD)
SBN (Maryland)	1	3.5	100% (RAID)	
SBN (PSI)	0.087	.4	100% (RAID)	
NAIF	0.05	.1	100% (RAID)	
ATMOS	0.5	1.5	85% (RAID)	15% (TAPE)
GEO	14	18	100% (RAID)	
IMG (USGS)	7	9.5	100% (RAID)	
IMG (HiRISE)	8.6	6	100% (NAS)	
IMG (JPL)	9	33	95% (RAID)	5% (OTHER)
IMG (THEMIS)	9.9	25.5	100% (RAID)	
PPI	2	4	100% (RAID)	
RINGS	1	-	100% (RAID)	

Totals: ~54+ TB ~100+ TB

* Primary Repository

Preliminary* Survey Costing Results for Archival Storage



* Costs varying widely from node-to-node. This is an initial average cost based on four years with a few samples of data. Later MC request is do we want to compare to industry numbers.

NOTE (1): If you factor in data integrity, all media seems to have on-going costs to verify its usability.
NOTE (2): Cost-wise, optical media makes more sense for smaller volumes of data.
NOTE (3): PMWG estimates (based on PDS input) that costs start to level out at roughly $4.5K/TB/Year as the data increases to larger volumes.

Backup Options*

Solution	Type	Media Type	Cost
Commercial	External	Varies	$12000/TB/Year
SDSC	External	Online RAID	$1000/TB/Year
SDSC	External	Near-line Tape (served by robotic tape device)	$400/TB/Year
NSSDC	Node Independent	Replicate Node Systems	Cost of hardware + $20K to $50K labor
Node to Node	Internal	LTO Tape	Cost of media + labor

* The PMWG noted that one of the key areas that needs to be addressed are plans for backing up the archive (e.g., operational copy of data for recovery of a node)

San Diego Super Computing Center (SDSC)

- Presented by Reagan Moore at Archives in 21st Century Meeting
  - SRB has been “extremely successful”
  - They have expressed interest in being a backup site for PDS
- SDSC provides storage services and software for managing data collections online, near-line, etc
  - Flagship product is the Storage Resource Broker (SRB) which provides
    - Persistent naming of distributed data
    - Management of data stored in multiple types of storage systems
    - Organization of data as a shared collection with descriptive metadata, access controls, audit trails
- Storage Resource Broker (SRB) manages 2 PBs of data in internationally shared collections
  - SDSC offers a storage service storing data at San Diego based on a Service Level Agreement
- Data collections for NSF, NARA, NASA, DOE, DOD, NIH, LC, NHPRC, IMLS
- Goal has been generic infrastructure for distributed data

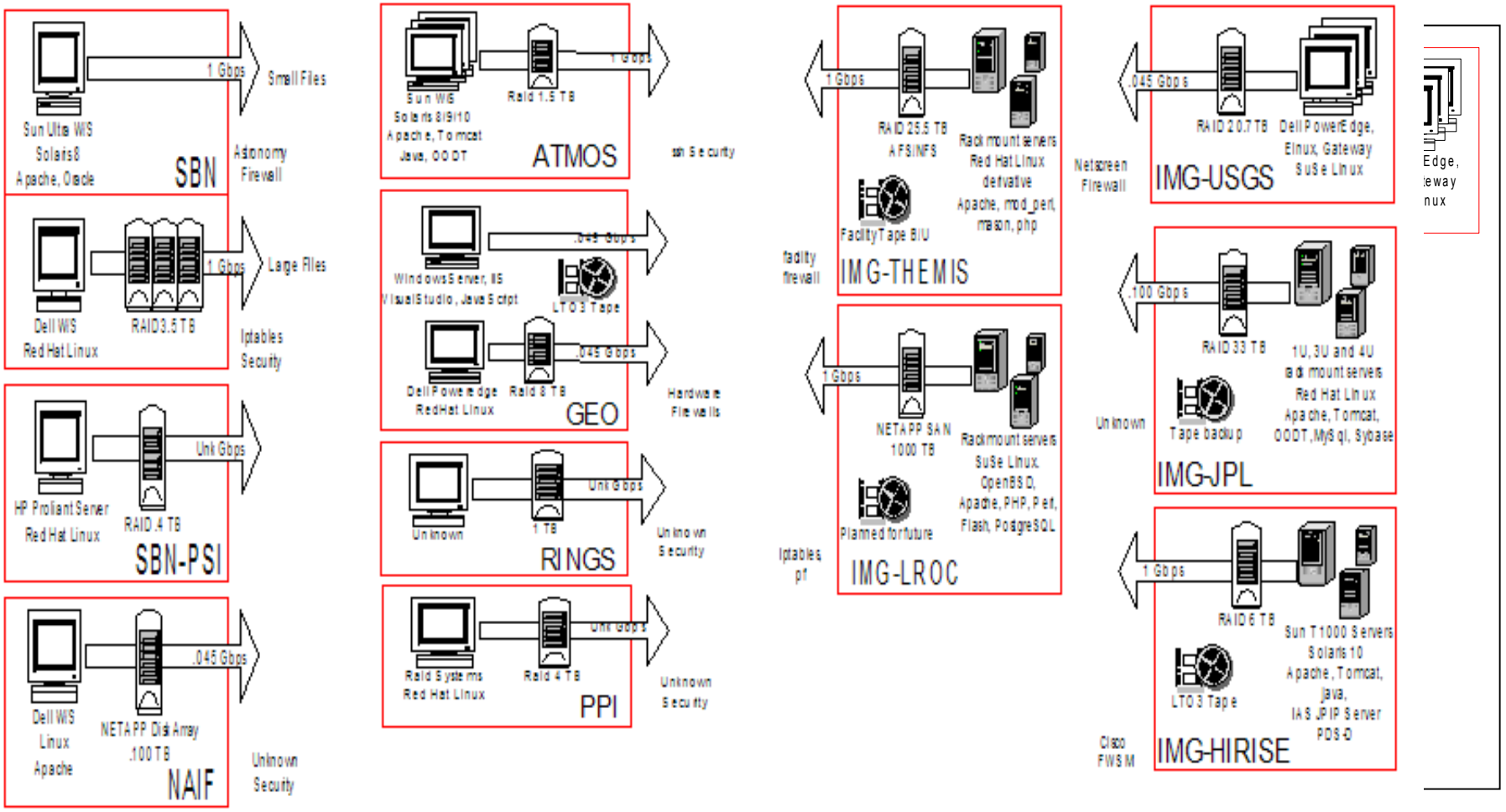
Date	5/17/02		6/30/04			4/23/07		
Project	GBs of data stored	1000Õs of files	GBs of data stored	1000Õs of files	Users with ACLs	GBs of data stored	1000Õs of files	Users with ACLs
Data Grid								
NSF / NVO	17,800	5,139	51,380	8,690	80	119,278	17,828	100
NSF / NPACI	1,972	1,083	17,578	4,694	380	36,514	7,483	380
Hayden	6,800	41	7,201	113	178	8,013	161	227
Pzone	438	31	812	47	49	25,681	14,793	68
NSF / LDAS-SALK	239	1	4,562	16	66	193,959	196	67
NSF / SLAC-JCSG	514	77	4,317	563	47	20,620	2,152	55
NSF / TeraGrid			80,354	685	2,962	293,539	8,038	3,267
NIH / BIRN			5,416	3,366	148	20,800	33,748	424
NCAR						1,567	8	2
LCA						1,834	39	2
Digital Library								
NSF / LTER	158	3	233	6	35	260	41	36
NSF / Portal	33	5	1,745	48	384	2,620	53	460
NIH / AfCS	27	4	462	49	21	733	94	21
NSF / SIO Explorer	19	1	1,734	601	27	2,750	1,202	27
NSF / SCEC			15,246	1,737	52	168,931	3,545	73
LLNL						13,784	1,374	5
CHRON						6,398	2,064	5
Persistent Archive								
NARA	7	2	63	81	58	3,793	4,983	58
NSF / NSDL			2,785	20,054	119	5,699	50,600	136
UCSD Libraries			127	202	29	190	208	29
NHPRC / PAT						1,888	521	28
RoadNet						2,608	975	30
UCTV						7,359	2	5
LOC						9,693	256	8
Earth Sci						3,794	511	5
TOTAL	28 TB	6 mil	194 TB	40 mil	4,635	961 TB	153 mil	5,516

PMWG Recommendations to PDS MC

- Develop "framework" report for nodes to provide information to PDS Management regarding backup and integrity plan (and their local implementation). Roll into a PDS plan and post at MC website.
- Explore use of SDSC services at SDSC for PDS as a PDS near-line backup source
  - PMWG would explore use of SRB clients
  - PMWG would test putting data at SDSC
  - PMWG will report back to the MC on results for a full recommendation
- Given that NSSDC-PDS I/F is effectively, ready, start transferring the electronic volumes to NSSDC
  - Transfer those data sets where the NSSDC MPGA client will run (Unix-based)
  - Work with NSSDC on beta testing Windows version of software
- Continue to develop PDS storage projections and defining how they map to media solutions (e.g., produce a high level PDS storage architecture diagram from the survey, etc)
- Determine whether to migrate volumes not online?
  - Survey showed that a small percentage is not online. Some are old volumes (e.g., Magellan DODR and FBIDR, several hundred SDDPT volumes at EN)

Backup

Survey Cost Results

- Data System Capacity and Costs (excluding LROC)
  - Current capacity is 100 TB, storage used is 50 TB.
  - Total system cost (including FY07 labor) is on the order of $940K or $9.5K per TB of storage capacity but varies widely between nodes.
  - Labor cost is in the ballpark of $200K per year.
  - Data acquisitions are on the order of 200 TB for the next 3 years and replenishment costs are on the order of $1000K per year ($5K per TB). Most acquisitions are at IMG node.

Preliminary Hardware Architecture Diagram*