

September 20, 2012
Harvard Library
Digital Repository Service

Digital Repository Service (DRS)

- A preservation and access repository for digital assets
- For content that has long-term value for scholarly research



Content

- Paintings, drawings, sculpture and architecture.
- Photographs (e.g., prints, lantern slides, and daguerreotypes)
- Archaeological objects (e.g., baskets, pottery and ironwork)
- Page-turned object (books, manuscripts, etc.)
- Musical scores and performances
- Correspondence (written, typescript, and printed); diaries; and journals

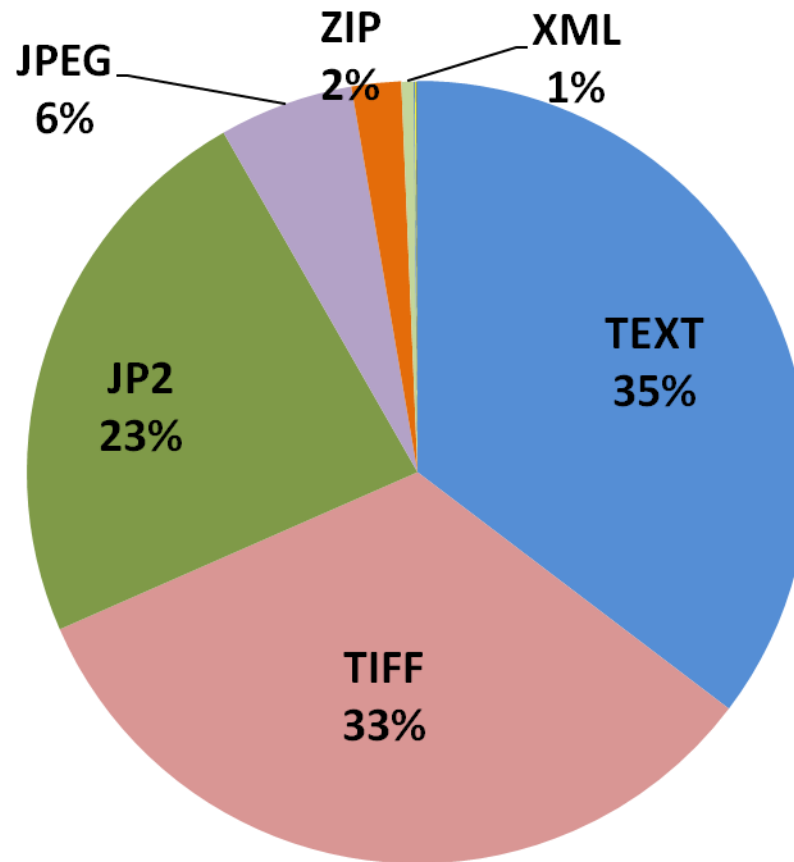


Planned Content

- Email
- Video
- Office formats
- Databases
- Datasets



Formats (by file count)

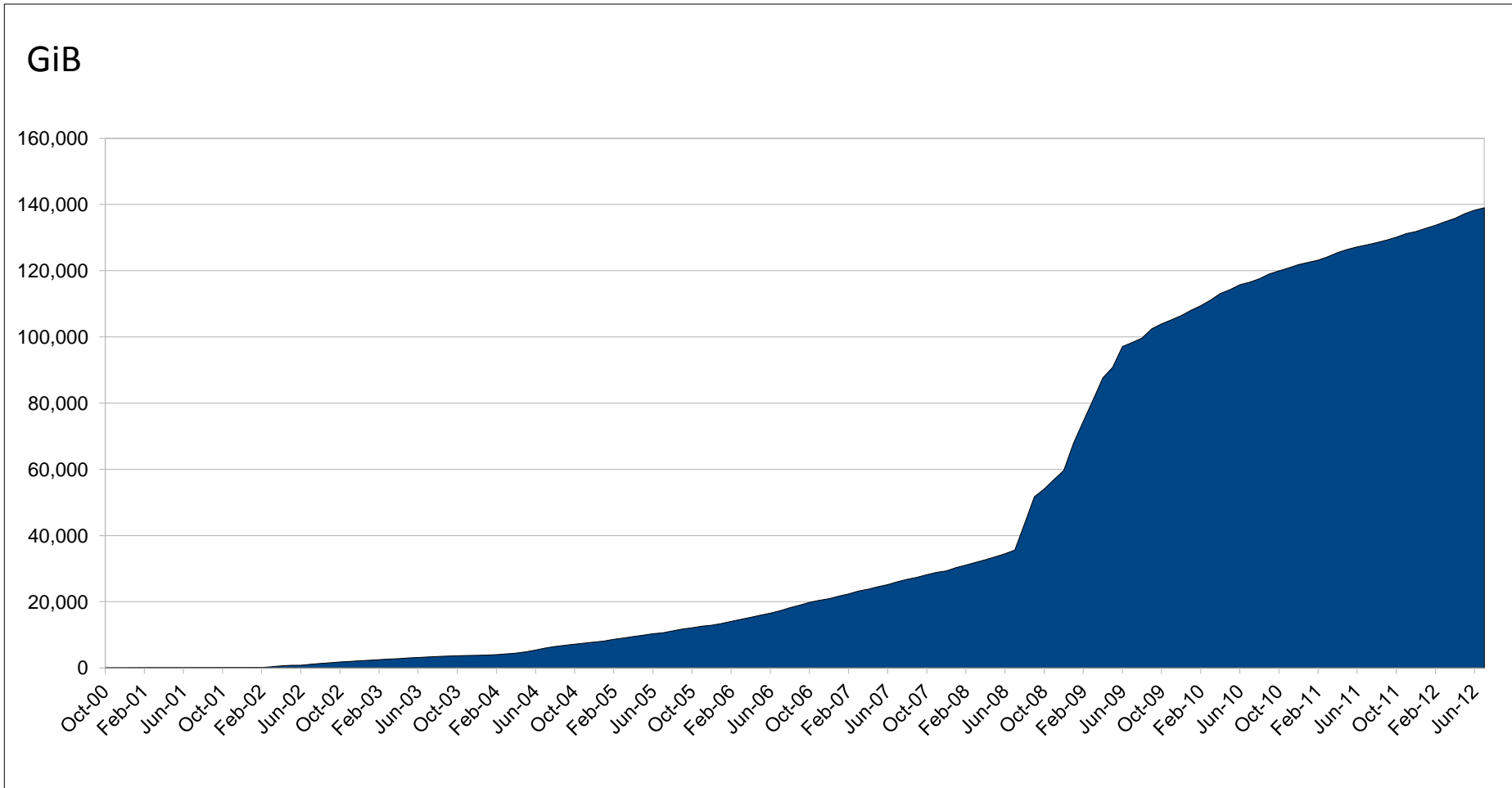


Content Growth

8/1/2012 :

1 copy: 149 TB, 42 mill. files

all copies: 499 TB, 168 mill. files



HARVARD
LIBRARY

University-wide Use

- Cultural Resource Repositories
 - Museums
 - Archives
 - Libraries
 - Special Collections
 - Institutes, Centers
- Cross-department collections and initiatives



Present Systems Architecture

- Two sites using Oracle SAM/QFS 4.x.
- Replicated disk copies between the two sites.
- One onsite tape copy.
- One offsite tape copy.
- Application servers only at one site serving content from DRS.
- DR capability to second site.
- In-house data integrity checking on 2nd disk copy.

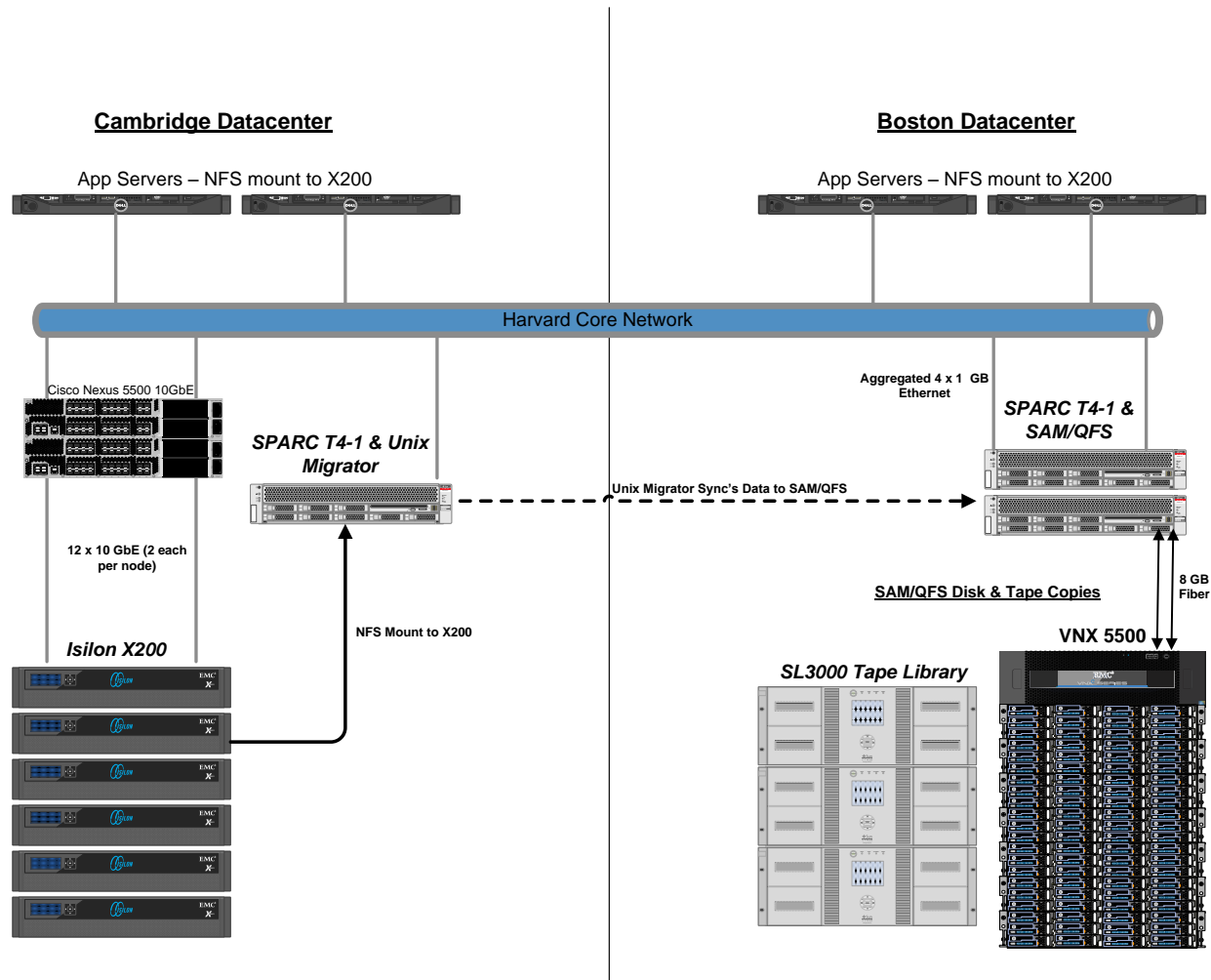


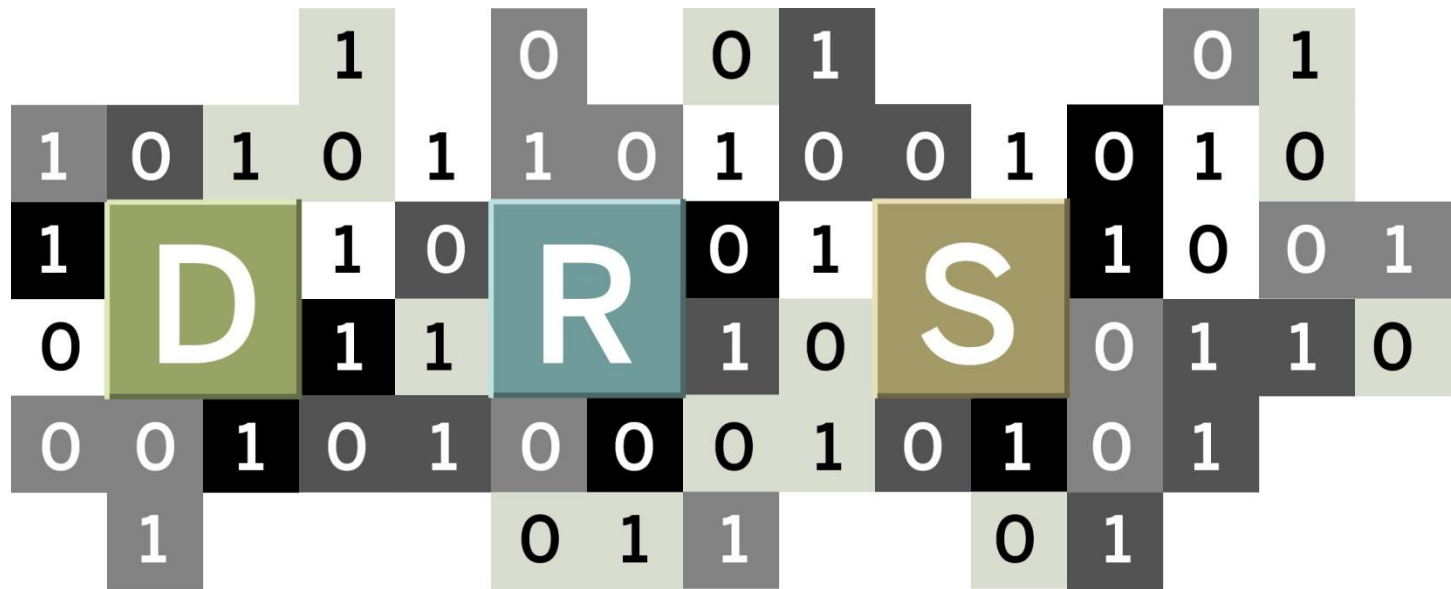
New Systems Architecture

- Two complete disk copies located at two sites using two different storage types.
- Offsite/Onsite complete tape copies.
- Improved tape drive technology: Oracle T10000c (built in CRC checks, faster)
- Improved content delivery to applications using scalable NAS solution.
- HA application server configuration, load balanced between two sites.
- Failover capability for storage to second site.
- Data integrity checking using Cambridge Computer's Starfish application.



New Systems Architecture





Harvard Library

Thomas M. Scorpa

thomas_scorpa@harvard.edu