New York Public Library
Digital Storage Requirements

Barbara Taranto
Managing Director, NYPL Labs
Office of Strategic Planning
Library of Congress Meeting Storage Architectures for Digital Preservation
September 27-28 2010
Goals for NYPL Preservation Environment

- Create a robust, sustainable and extensible architecture
- Consolidate storage platforms
- Consolidate repository services
- Tie production workflow to preservation repository
- Create clear governance of preservation repository
- Herd sheep – gather all the silo’d content into a format agnostic preservation space
Stuff
Size and Type

- 60 TBs legacy content on SAN – simple objects – database layer monitoring submissions and recording minimal technical metadata
- 12 TBs legacy content on external drives – including datasets and complex objects – spread sheet inventories
- 10 TBs vector datasets (GIS) in open and proprietary formats
- 40 TBs media – first of several streams from vendors
- JP2000 output from Google Books and OCA initiatives
- New initiatives – Digital Humanities
Considerations (challenges)

- The submission process requires almost the same amount of storage space (and same requirements) as destination requirements.

- Content from vendors and from born digital sources require staging areas for quality control and processing. Space requirement may be higher than destination requirements.

- Destination storage is part of lifecycle management. Migrations requires more than destination storage – format migrations are space hogs.
Investigating local and cloud services
Use case for running processes (pre-ingest) on newly acquired assets