



# Library of Congress Designing Storage Architectures Workshop – Sept. 2007

- Rick Matthews
- Sr Staff Engineer – Archiving BU, Storage S/W
- Sun Microsystems, Inc.

# Retention and Change

- Long term archives are all about (to me):
  - > Reliable and economical retention of data
  - > Changing media and media formats
  - > Changing metadata and discovery formats
  - > Changing data formats
- Picking questions 2 & 3
  - > What is a least worrisome portion of data reliability
  - > What standards in the near term will affect archive systems

# Retention

## Reliability of Stored Data is Paramount

- A retention solution needs to be affordable
  - > Acquisition of equipment, power and manpower
- Multiple copies of stored data increases reliability
  - > Media failure usually recovered from an additional copy
  - > Detected “checksum” failure may be recoverable from an additional copy
- Separate reliability metadata
  - > Checksums, hashes and signatures (and more)
  - > Reliability techniques may differ between applications
    - And the archive system is also an application

# Change

## Change Happens

- Archive media, metadata and content all change
  - > Tape, drives and media types change for age, density and economics
  - > Applications change data formats (checksum, security, new formats, etc.)
  - > Search/discovery metadata is emerging and will change.

# Workshop questions

## 2. What am I least concerned about for data reliability

- Techniques for reliability are well defined
  - > Level of reliability detection is tunable (none, checksums, hash, etc.)
  - > Recovery from a detected failure is as automatic as possible
  - > Device reliability is increased via drive statistics when available
  - > By configuration, reliability of data can be a no-brainer

# Workshop questions

## 3. What are my concerns regarding standards

- Manufacturers want exclusivity by proprietary formats (DVD, DVD+, etc.)
- Applications change formats (MP2, MP4, etc.), when do they de-commit older formats?
- How do we detect content format change errors?

# My questions to workshop

- Tape continues for lots of of good economic reasons.
  - > Capacity per inch<sup>3</sup>; power at rest; degradation at rest
  - > Are there any emerging storage techniques/media that can approach tape economics?
- Where do you think some functions belong?
  - > Search? Deduplication? Content verification?
  - > Are these archive system or application ISV problems?



# Library of Congress Designing Storage Architectures Workshop

**Rich Matthews**

[rick.matthews@sun.com](mailto:rick.matthews@sun.com)