Designing Storage Architectures for Digital Preservation: Panel 1- How Would You Store Data

September 20, 2012
Library of Congress
Henry Newman
Our Distinguished Panelists

- Dave Anderson Seagate
- Shawn Brume IBM
- Robert Fontana IBM
- Ed Childers IBM
- Alan Poston Xyratex
- Bob Raymond Oracle/Sun
- Erik Riedel EMC
How you would store the following non-compressible files in a long-term storage environment requiring very high reliability, which would have a request rate of 10% of the data per year:

<table>
<thead>
<tr>
<th>2012</th>
<th>2015</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 PB</td>
<td>5 PB</td>
<td>5 PB</td>
</tr>
<tr>
<td>20 PB</td>
<td>20 PB</td>
<td>20 PB</td>
</tr>
<tr>
<td>50 PB</td>
<td>50 PB</td>
<td>50 PB</td>
</tr>
</tbody>
</table>
Ground Rules for all of us

• 4 minutes per panelist timed
• No discussion about specific hardware or software products is allowed
• We expect 10 minutes for discussion among panel members and moderator after the individual presentations, and then questions from audience
Thoughts and considerations

• Long term storage requires an understanding of technologies and costs:
  – What about power and cooling for some technologies
    • Power costs are not coming down
  – What about the impact of migration
    • All media must be migrated
      – Given failure rate increases with age
      – The lack of interface support for storage devices
  – What about hard error rates and silent data corruption rates
    • None of these rates have changed in many years nor are they going to change
  – What about standards
    • What archival standards are here and what are missing
LOC Challenges

• Long term media costs include:
  – Migration costs
  – Floor space, power and cooling
  – Cost for interface (storage controller, tape drive etc)

• Data integrity is a critical concern
  – What is the impact of each media type in its interface and packaging (e.g. storage controller, tape drive configuration?)
  – End to end silent corruption rate

• What media type best optimizes all these issues and why?
Panelists: Take it away

Audience: Please hold questions until after the panelists have spoken