DSRB and DSpace 2

Richard Rodgers
MIT Libraries
DLF Spring Forum – San Diego
April 13-15, 2005

DSpace 2 - Objectives

- **Improved modularity via re-factored APIs and plug-in mechanisms**
- **Better preservation support by clearer AIP formulation and embodiment**
- **Scalable storage and service components**
- **SRB good platform for exploring and prototyping answers**

What do scalable storage architectures look like?

- They can be complex: distributed, remote, heterogeneous hardware and software
- They must support data mobility: ability to re-host, adopt, register, transfer, mirror content
- May be composed of distinct administrative domains – local, consortial/community, or 3rd party commercial storage providers
- Therefore, they must also be auditable, and optimally even self-diagnosing and healing

How do preservation services and activities operate at scale?

- Automated: must support high-level directives and policies that are translated to low-level and bulk data operations
- Yet flexible: should accommodate trade-offs among time, risk mitigation, cost, performance, administrative autonomy, etc
- Replication as an example: replicate-on-ingest v.s. on-demand
- Interoperable: many of these activities could involve inter-repository relationships and agreements

How do data and collection management operate at scale, and how do they relate?

- How much and what kind of low-level data management information is relevant to content life-cycle management?
- The SRB maintains a rich set of technical, administrative, and even descriptive metadata.
- Drinking from the metadata fire-hose: managing large scale event streams
- Can semantic web techniques and technologies be brought to bear to capture and expose this data?

How can scalable storage enable or complicate repository federation?

- Are any storage services required beyond those underpinning repository services?
- Some examples: replication networks, shared storage facilities, others
- How to exploit SRB optimizations such as parallel I/O?
- Relations to common grid infrastructure: authentication, computational resources

DSpace http://dspace.org
http://wiki.dspace.org