SRB at the AHDS

Andrew Speakman
Hamish James
SRB

- Storage Resource Broker (SRB) is client-server middleware that provides a uniform interface for connecting to heterogeneous data resources over a network and accessing replicated data sets

- http://www.npaci.edu/DICE/SRB/
AHDS

- The AHDS “preserves arts and humanities digital resources created by Higher Education" and provides “rich, deep access to the intellectual content of arts and humanities digital resources”
Collection Nature & Growth

• Datasets, electronic texts, GIS, audio, video, images, CAD ...
  - no software

• Collection currently >1TB
  - imminent growth to >3TB
  - uneven growth, small number of deposits contribute bulk of collection
  - most data is static
Strategic Context

• The AHDS 2002-05 Strategic Plan envisions replacing a “substantially devolved structure with a much more centralised service”
  - Replace separate arrangements for preservation of digital material with shared facilities
  - Replace separate delivery systems with interoperable shared system
  - Maintain distributed organisational structure
Consultancy Recommendations

• “A preservation centre should be established to hold preservation copies of all materials held by [AHDS Centres] which require indefinite preservation”
  - incorporate preservation metadata
  - incorporating a collection management system
  - offer sufficient storage for all AHDS Centres
  - offer sufficient security of archived data while providing AHDS Centres with flexible and easy access to their digital collections
Early Design Plan

Submission

Processing Areas
(at the AHDS SPs)

Deposit (Files & Metadata)

Main Digital Repository
1. Original Version
2. Preservation Version
3. Dissemination Version(s)

Temporary Backup

Metadata Engine

Transfer (Selected Files & Metadata)

Secondary User Store
Files for delivery to users

Application Layer

Presentation
Web-based search, browse and delivery tools

HTTP, FTP

User Request

Funded by:

AHDS
Repository Requirements

• Security
• Reliability
• Long Term Integrity
• Large Capacity
• Manageability
• Affordability
• Master Copy
• Internal Validity
Storage Requirements

- Management and access control
- Distribution
- Replication
- Offline and online segments
- Integrity management
- Synchronisation and ingest mechanisms
Repository Plan
Implementation

- **Immediate:**
  - Hardware RAID (2TB) and tape library (10TB)
  - scripts and procedures
  - offsite duplicates

- **Medium term:**
  - evaluating Storage Resource Broker
  - partnership with CCLRC, Atlas Petabyte Storage Service (APS)
About SRB

• Project
  - http://www.npaci.edu/DICE/SRB/

• Data Grid
  - storage ‘virtualisation’
  - platform / technology neutral
  - data duplication and synchronisation

• Free to academic institutions
  - not Open Source
Key Features

• Logical namespace
• High performance parallel I/O
• Scalable capacity (terabytes)
• Many clients and APIs
• Many supported storage systems
• Integrated metadata
• Integrated security and access control
• Wide user base, mature system
Infrastructure

- OSs - Unix, Linux, Windows, Mac
- MCAT - Oracle, DB2, Sybase, Postgres
- Storage vaults:
  - file systems
  - archive systems (HPSS)
  - databases (Oracle, DB2)
  - tape library
- Resources
  - Physical / logical / compound
Clients and APIs

- Web based browser client
- Windows GUI client
- Java GUI client and admin interface
- Java API (also C, Python)
- Command line tools
- No direct file system access
Windows Client

- SRB Admin Interface
- Resource Management
- Attribute Table
- File List
- Dataset Information

Funded by: AHDS

© AHDS
Metadata

- **MCAT**
  - objects
    - metadata (system level, application level)
  - resources
  - collections
  - users
- For resource control and data discovery
- One per domain
- Simple structure
- Containers
SRB Evaluation

- Install on Postgres + SuSE Linux
- Access through firewall
- File system sync (command line)
- Database backup and recovery
- Client upload and download (Windows, Web)
- Web service integration via Java API
- Tested admin tool remotely
- Demonstrated replicated resources
Open Issues

- Large networked data volumes
- Windows client (user acceptance)
- Metadata issue – in band vs out of band
- Integration with tape library software (compound resources)
- GSI (X.509 and SSL)
- Federated zones
SRB Strengths

- Satisfies many requirements
- Good technical architecture
- There isn’t an equivalent
- Mature and popular
- Highly functional
- Federated
SRB Weaknesses

• Technical support
• GUI clients
• MCAT dependence
• Not for faint hearted
• Not open source
• Not standards based
• Monolithic (does too much)
Conclusions

• SRB has potential:
  – to simplify day-to-day operations
  – to simplify distributed management of data
  – OAIS

• AHDS is looking for partners
  – DSpace
  – SHERPA II