

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”

Funded by:


arts and humanities research board



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

Funded by:


arts and humanities research board



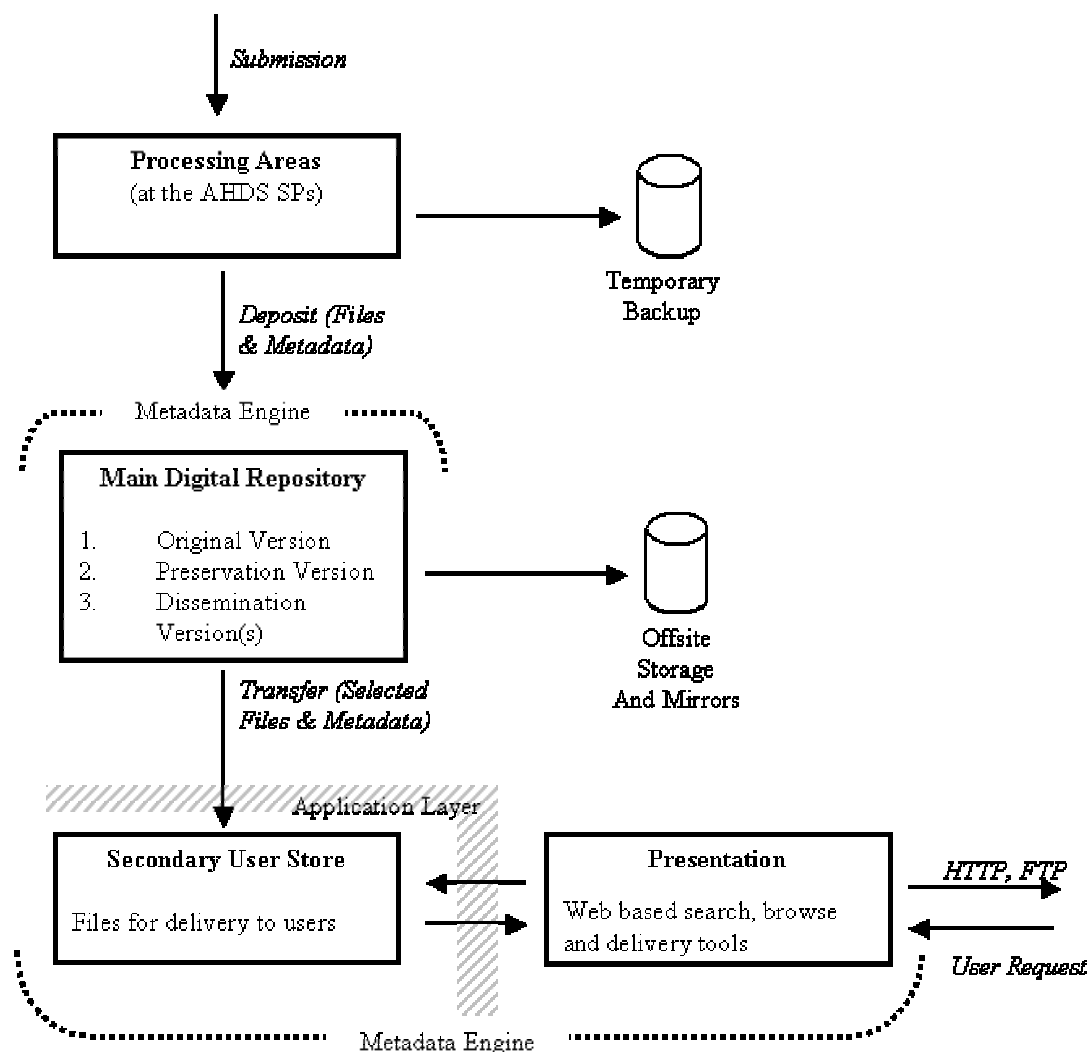
Strategic Context

- 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



Repository Requirements

- Security
- Reliability
- Long Term Integrity
- Large Capacity
- Manageability
- Affordability
- Master Copy
- Internal Validity

Storage Requirements

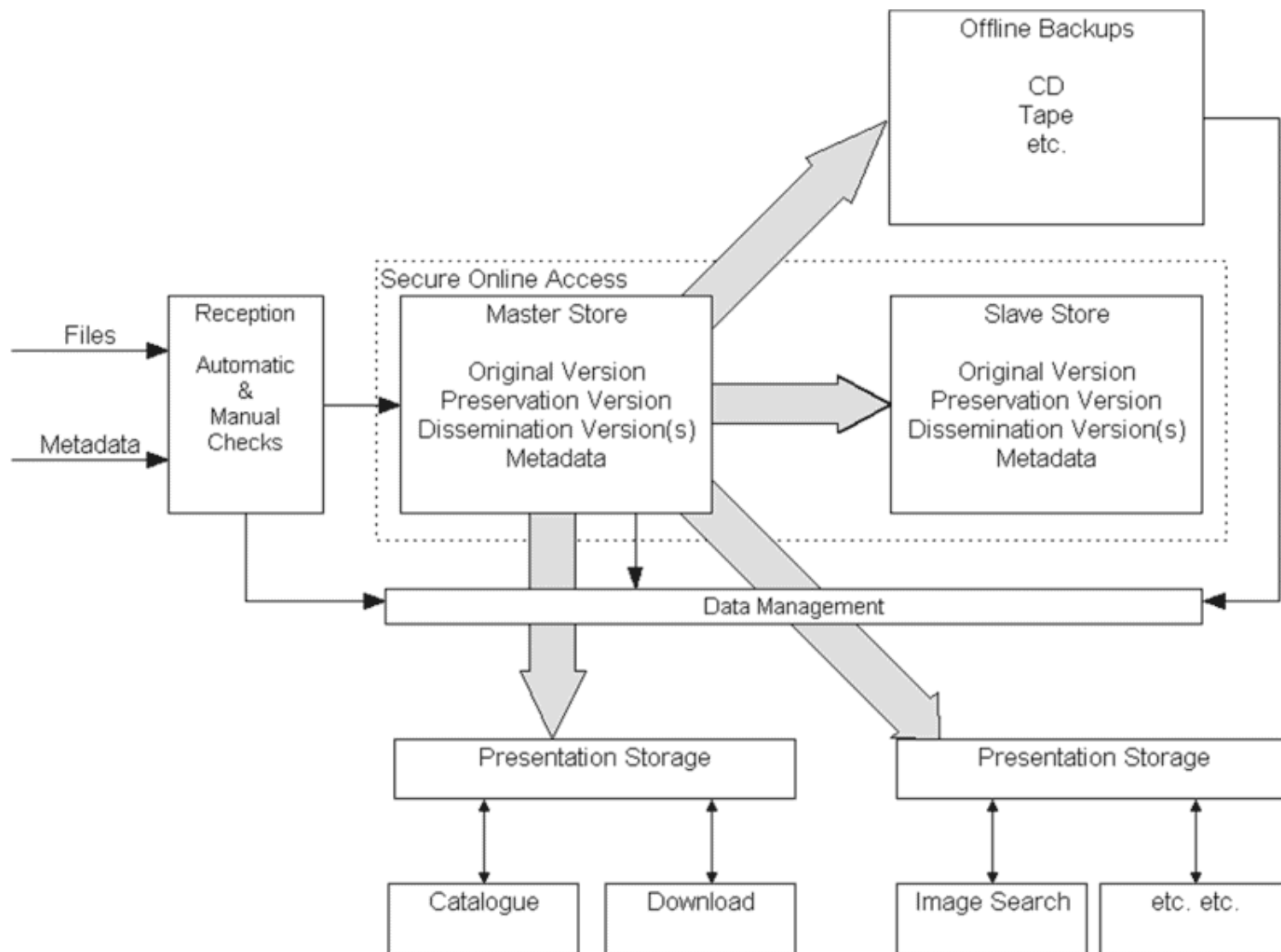
- Management and access control
- Distribution
- Replication
- Offline and online segments
- Integrity management
- Synchronisation and ingest mechanisms

Funded by:


arts and humanities research board



Repository Plan

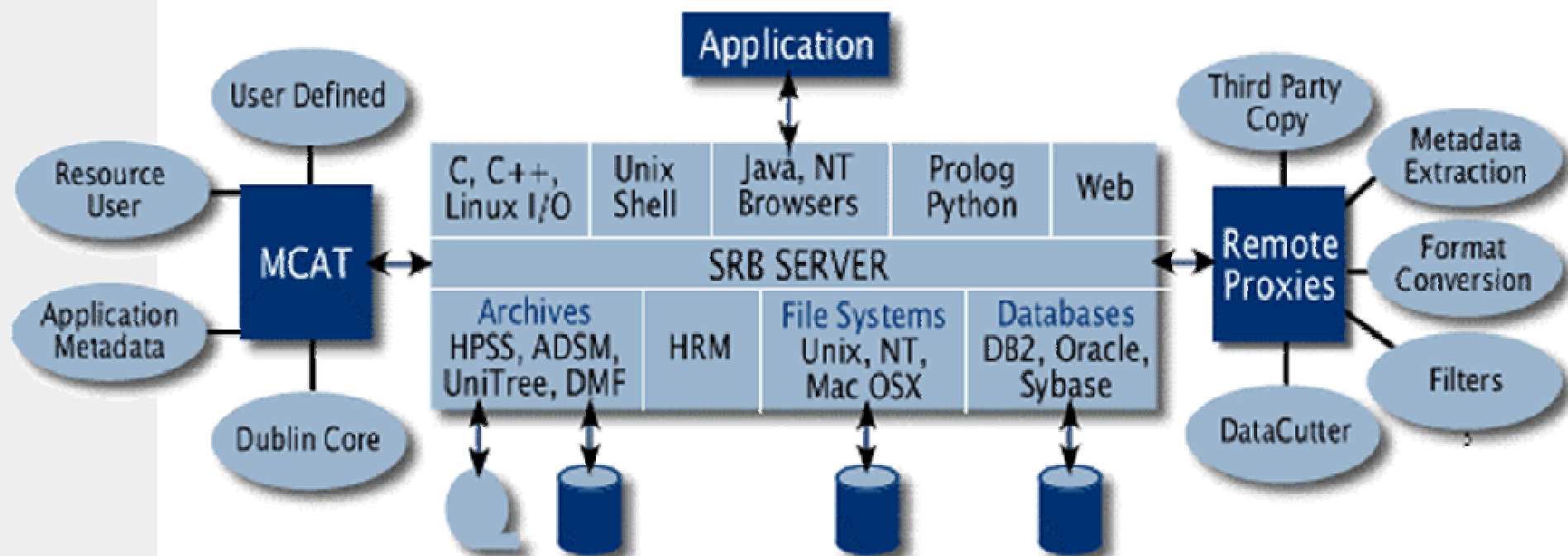


Implementation

- 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

SRB Admin

The Srb/Mcat Admin Main panel

File Zones Users Resources Tokens

Display Tokens panel

Select a Token type:

- Location
- ResourceType
- ResourceClass
- DataType
- UserType
- Domain
- Action
- AccessConstraint
- Zone

Name: compound	NetPrefix:
Name: demozone1oc	NetPrefix:
Name: home	NetPrefix:
Name: level1	NetPrefix:
Name: level2	NetPrefix:
Name: level3	NetPrefix:
Name: level4	NetPrefix:
Name: null	NetPrefix:
Name: jem	NetPrefix: 192.168.0.103
Name: 192.168.0.103	NetPrefix: 192.168.0.103
Name: ghidorah oracle Vault	NetPrefix: ghidorah
Name: ghidorah sdsc	NetPrefix: ghidorah.sdsc.edu:NULL:NULL
Name: db2 mda-18 sdsc	NetPrefix: mda-18.sdsc.edu:1gnadb2sd:db2v2
Name: mda-18 sdsc	NetPrefix: mda-18.sdsc.edu:NULL:NULL
Name: unix file mda-18 sdsc	NetPrefix: mda-18.sdsc.edu:NULL:NULL
Name: mda-18 db2 Vault	NetPrefix: mda-18.sdsc.edu:vaults:db2v2

Add New Location panel

Location Name:

Host Address:

Parent Location:

Domain of Server user:

Server user ID:

The following is needed only for DB Large Object L...

Database Name:

Schema Name (for DB2 only):

Status:

Execute Clear / Refresh CL...

Display Resources panel

Name	Network Address (prefix)	Type	Default Path	PhysResc Name
null		generic		[Same]
sdsc-mdal8-fs	mda-18.sdsc.edu:NULL...	file system	/scratch0/srb/SRBVault...	[Same]
sdsc-fs	mda-18.sdsc.edu:NULL...	unix file system	/projects/mdas/srb/SR...	[Same]
ahds-exec-fs-1	192.168.0.103:NULL.N...	unix file system	/raid/1/archive/?USER.?	[Same]
ahds-exec-fs-2	192.168.0.103:NULL.N...	unix file system	/raid/2/archive/?USER.?	[Same]
ahds-fs-2	192.168.0.103:NULL.N...	unix file system	/raid/2/archive/?USER.?	[Same]
ahds-fs	192.168.0.103:NULL.N...	logical		ahds-fs-2
ahds-fs-1	192.168.0.103:NULL.N...	unix file system	/raid/1/archive/?USER.?	[Same]
ahds-fs	192.168.0.103:NULL.N...	logical		ahds-fs-1

Click on a row for more details

Refresh Close

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 Edit View Help

3 Use Resource | Container ahds-fs

Attribute Value

M Holds Some dummy GIFS for testing

M Size Has 500 files

Name	Size	Owner	Timestamp	Repl	Resource
Copy (412) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.04	0	ahds-fs-2
Copy (412) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.05	1	ahds-fs-1
Copy (413) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.03	0	ahds-fs-1
Copy (413) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.04	1	ahds-fs-2
Copy (414) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.06	0	ahds-fs-1
Copy (414) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.06	1	ahds-fs-2
Copy (415) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.07	0	ahds-fs-2
Copy (415) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.07	1	ahds-fs-1
Copy (416) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.08	0	ahds-fs-1
Copy (416) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.08	1	ahds-fs-2
Copy (417) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.09	0	ahds-fs-2
Copy (417) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.09	1	ahds-fs-1
Copy (418) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.10	0	ahds-fs-1
Copy (418) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.11	1	ahds-fs-2
Copy (419) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.11	0	ahds-fs-2
Copy (419) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.12	1	ahds-fs-1
Copy (420) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.16	0	ahds-fs-2
Copy (420) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.17	1	ahds-fs-1
Copy (421) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.18	0	ahds-fs-1
Copy (421) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.18	1	ahds-fs-2
Copy (422) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.19	0	ahds-fs-2
Copy (422) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.19	1	ahds-fs-1
Copy (423) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.20	0	ahds-fs-1
Copy (423) of GIF.gif	396625	srbAdmin	2004-10-01-16.12.20	1	ahds-fs-2

deleted Key 2 = Another text value

Datasets: 454 - Metadata: 2 - Users: 1

Web Client

Explore SRB

Zone Info

userHomeZc
AHDS
connecting2
AHDS
CurrentZone
AHDS

Zones:
AHDS

Collections:

- AHDS
- container
- g1
- home
- srctest
- styles
- trash

Collection: 3
Parent Collection: /AHDS/home/srbAdmin.ahds/shared/Input/Input
Owner: srbAdmin@ahds

Metadata describing entity 3
Field Name : Value Units < Explanation >
Holds : Some dummy GIFS for testing

Move Up
 Ingest File
 Create File
 Register File
 Register Directory
 Register URL
 Register SQL
 Register ORBData
 Register Command
 Make Collection
 Make Container
 Browse Query
 Other Info
 Online Help
 Exit MySRB

/AHDS/home/srbAdmin.ahds/shared/Input/Input/3

Function	Data Name	Creation Time	Owner	Replica Ver NumberNum
Get File	Copy (412) of GIF.gif	2004-10-01-16.12.04	srbAdmin@ahds	0
Get File	Copy (412) of GIF.gif	2004-10-01-16.12.05	srbAdmin@ahds	1
Get File	Copy (413) of GIF.gif	2004-10-01-16.12.03	srbAdmin@ahds	0
Get File	Copy (413) of GIF.gif	2004-10-01-16.12.04	srbAdmin@ahds	1
Get File	Copy (414) of GIF.gif	2004-10-01-16.12.06	srbAdmin@ahds	0
Get File	Copy (414) of GIF.gif	2004-10-01-16.12.06	srbAdmin@ahds	1
Get File	Copy (415) of GIF.gif	2004-10-01-16.12.07	srbAdmin@ahds	0
Get File	Copy (415) of GIF.gif	2004-10-01-16.12.07	srbAdmin@ahds	1
Get File	Copy (416) of GIF.gif	2004-10-01-16.12.08	srbAdmin@ahds	0
Get File	Copy (416) of GIF.gif	2004-10-01-16.12.08	srbAdmin@ahds	1
Get File	Copy (417) of GIF.gif	2004-10-01-16.12.09	srbAdmin@ahds	0
Get File	Copy (417) of GIF.gif	2004-10-01-16.12.09	srbAdmin@ahds	1
Get File	Copy (418) of GIF.gif	2004-10-01-16.12.10	srbAdmin@ahds	0

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

Funded by:


arts and humanities research board



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)

Funded by:


arts and humanities research board



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

Funded by:


arts and humanities research board

