

Grid-enabling Humanities Data

Mark Hedges

Arts and Humanities Data Service
King's College London

Funded by:



© AHDS

Overview

- What do we mean?
- Particular challenges of the humanities.
- Some approaches.

Funded by:



© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

What do we mean by grid-enabling?

Grid-enabling a data resource



Making the resource available via a grid



Making the resource accessible via
some form of grid middleware

Funded by:



JISC

© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

What do we mean by grid-enabling?

- Shared discovery/access
- Virtualised access
- Data integration
- Collaborative working

Funded by:



JISC

© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

Data on the grid

Implementation often driven by:

- need for fast access to very large, distributed data sets.
- dynamic data sets

Issues:

- Replication, caching
- Data transfer, latency, data propagation

Data formats: Flat files, databases, XML.

Funded by:



© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

An oversimplification ...

- “Hard” sciences – a driver to adopting grids was the need for fast, shared access to large, distributed datasets.
- Humanities – driver is the essential nature of the research data, rather than data size or speed of access

Funded by:



© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

Challenges of humanities data

- Diverse and multi-media/multi-format
- Complex structure
- Highly contextual
- Fuzzy and uncertain
- Interpretative
- Dispersed and isolated
- Rigid yet diverse metadata (if any)

Funded by:



© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

What would we like to do?

- Federation of the data resources
- Location-transparent resource discovery
- Delivery of (integrated) data resources
- Use of delivered data resources
- Semantics of data resources of prime importance

Funded by:



© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

Digital repositories

- Want to represent digital resources so as to reflect complexity and context.
- So: Store using flexible digital repository systems (Fedora (www.fedora.info at AHDS)).
- Want seamless integration between these highly structured repositories.
- So: Integrate repository software with grid middleware.

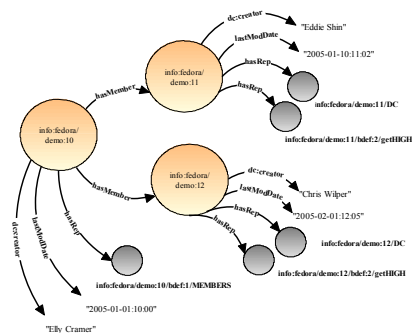
Funded by:



© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

Fedora data model



Funded by:



© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

Repository-Grid Integration

Two broad approaches:

- Grid as virtualised distributed storage.
- Repositories as data resources on grid.

Funded by:



© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

Grid storage for Fedora

- Fedora has been integrated with SRB.
- Currently looking at iRODS (Rule Oriented Data System) integration
- Will be able to make use of the complex metadata stored within Fedora.
- Complex actions encoded as rules built up from atomic services
- Rules integrated with system, yet easily changeable

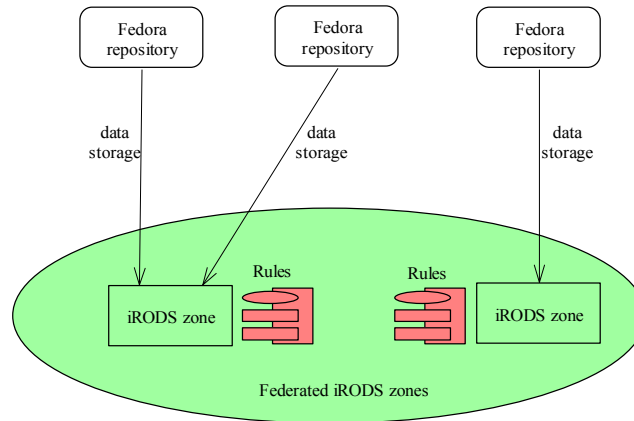
Funded by:



© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

Fedora and iRODS



Fedora repositories as grid resources

Use grid technologies to implement:

- Discovery of distributed Fedora repositories belonging to different administrative domains.
- Discovery of data resources across multiple Fedora repositories
- Delivery and integration of data resources across multiple Fedora repositories
- Grid AuthN and AuthZ mechanisms providing uniform access.

Resource Discovery

- Fedora resources all exposed as web services.
- Use service registries to publish metadata about repository resources.
- Populate registries by a number of means.
- Populate registries by a number of means

Funded by:



© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

Delivery and Integration

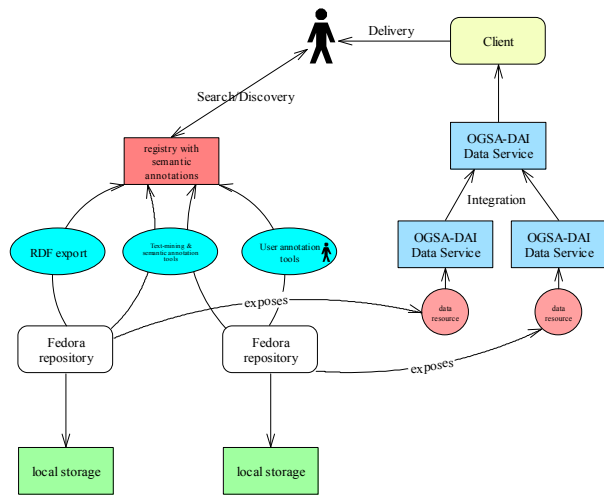
- Discovered data needs to be retrieved, integrated and delivered to clients.
- Looking at use of OGSA-DAI for this task
- OGSA-DAI has an extensible architecture to deal with multiple types of data resources.
- Initially relational and XML databases; projects looking at extending to other types of data resource: GEMS, SEE-GEO, GEESE.
- Highly varied nature of repository contents a significant challenge

Funded by:



© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007



Funded by:
Arts & Humanities Research Council
JISC
© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007

Contact



mark dot hedges at ahds dot ac dot uk

Funded by:
Arts & Humanities Research Council
JISC
© AHDS

NeSC Theme, e-Science in the Arts and Humanities, 2nd July 2007