



# Repositories and grids in the humanities

Mark Hedges, Centre for e-Research, King's College London

OGF-Europe Community Outreach Seminar

Digital Repositories - Interoperability Using Grid Technologies

OGF23 - 5 June 2008, Barcelona, Spain



# Overview

- Background: who were/are we?
- Humanities data and research
- Challenges for grids in the humanities
- Digital repositories and grids in the humanities: some models and examples.

# Who were we?



- Arts and Humanities Data Service
- Established 1996, funded until 2008
- Distributed structure: managing executive and specialist subject centres
- Mission: to collect, preserve and distribute digital resources produced by and for arts and humanities research (mainly in the UK)

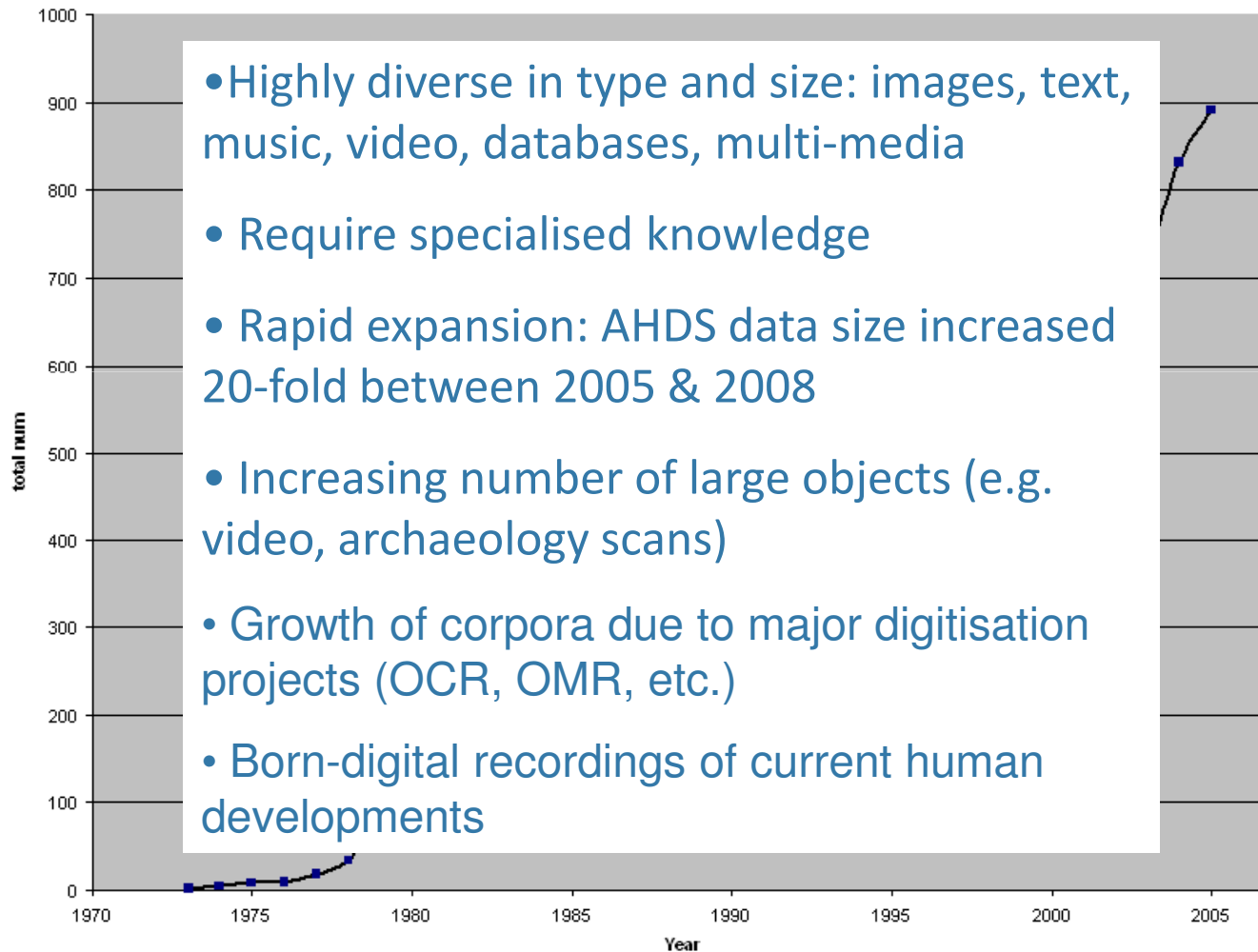


# Who are we?



- Centre for e-Research at King's College London
- Established 2007
- Incorporates staff and expertise of AHDS and other groups such as AHeSSC (Arts and Humanities e-Science Support Centre)
- Continuity, but some change of focus

# AHDS Collections





Museum of London  
Archaeological Archive

New Survey of London Life  
and Labor, 1929-1931

London College of Fashion:  
The Woolmark Company

Imperial War Museum

Designing Shakespeare



# Contrasts

- “Hard” sciences: requirements driven by size of data and high-throughput computation needs; fast access to large distributed data sets; simulations.
- Humanities: requirements driven by the nature of the research material and research processes. Data sizes (relatively) small and computation requirements (generally) limited.

# Humanities data

- Qualitative human-centric data that needs novel methods of selection
- Diverse: lack of standard formats and interfaces
- Semantics barrier: complexity and context dependency of research material
- Fuzzy, incomplete (and incompletable), inconsistent, inaccurate

# Challenges for the humanities

- Linking up (“grid enabling”) scattered & diverse data
- Linking up people: overcoming the traditional reluctance of (some) scholars to collaborate
- Representations and computational methods to deal with complex, incomplete, fuzzy, inconsistent data
- Provenance: recording the scholarly process and justifying scholarly conclusions

# Humanists' engagement with grids (in UK)

- Small projects funded by AHRC, JISC, EPSRC
- Larger projects in A&H e-Science programme
- Mainly data-centric: joining up data
- Access Grid for performance-based research and collaboration in the arts: joining up people
- Some potential for computational grid (arising from nature of humanities data)
- Virtual Workbenches/Virtual Research Environments

# Repositories, grids and ...

## Models for integration

## Examples of integration

# Models for repository-grid

repository / grid / utilities

## I. repository rests on grid

- data perspective:  
repository mounts a data grid
- the underlying "grid" could be any utility,  
including Amazon 3S, EC2 or GData



## IIa. repository ties in grid

- digital objects are distributed in grid
- processing objects is distributed



## IIb. repository embeds utilities

- insular embedding of dedicated infrastructure utilities

## III. gridified repository

- repository services are distributed
- ad hoc repository federations, etc.
- repository *is* a repository utility, consisting of a number of  
dedicated utilities

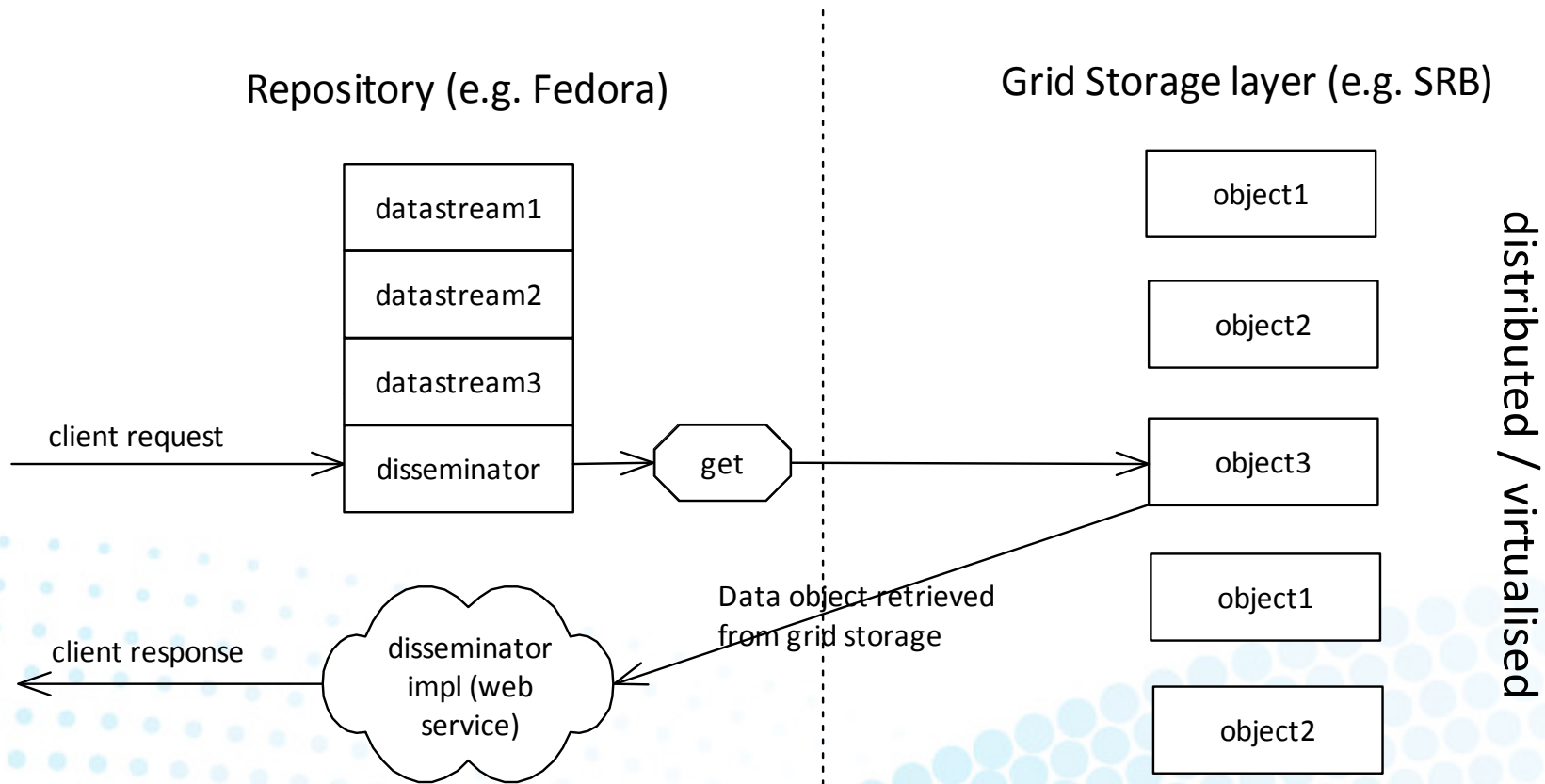


Andreas Aschenbrenner, Mark Hedges

# Repository + grid storage

- Repository/digital library rests on distributed, virtualised storage layer
- Digital objects distributed in data grid
- Simple model – doesn't really exploit grid
- Examples:
  - Academia Sinica, Taiwan: DSpace + SRB
  - AHDS, UK: Fedora + SRB
  - Archer, Australia: Fedora + SRB
  - Fedorazon, UK: Fedora/EPrints + Amazon S3

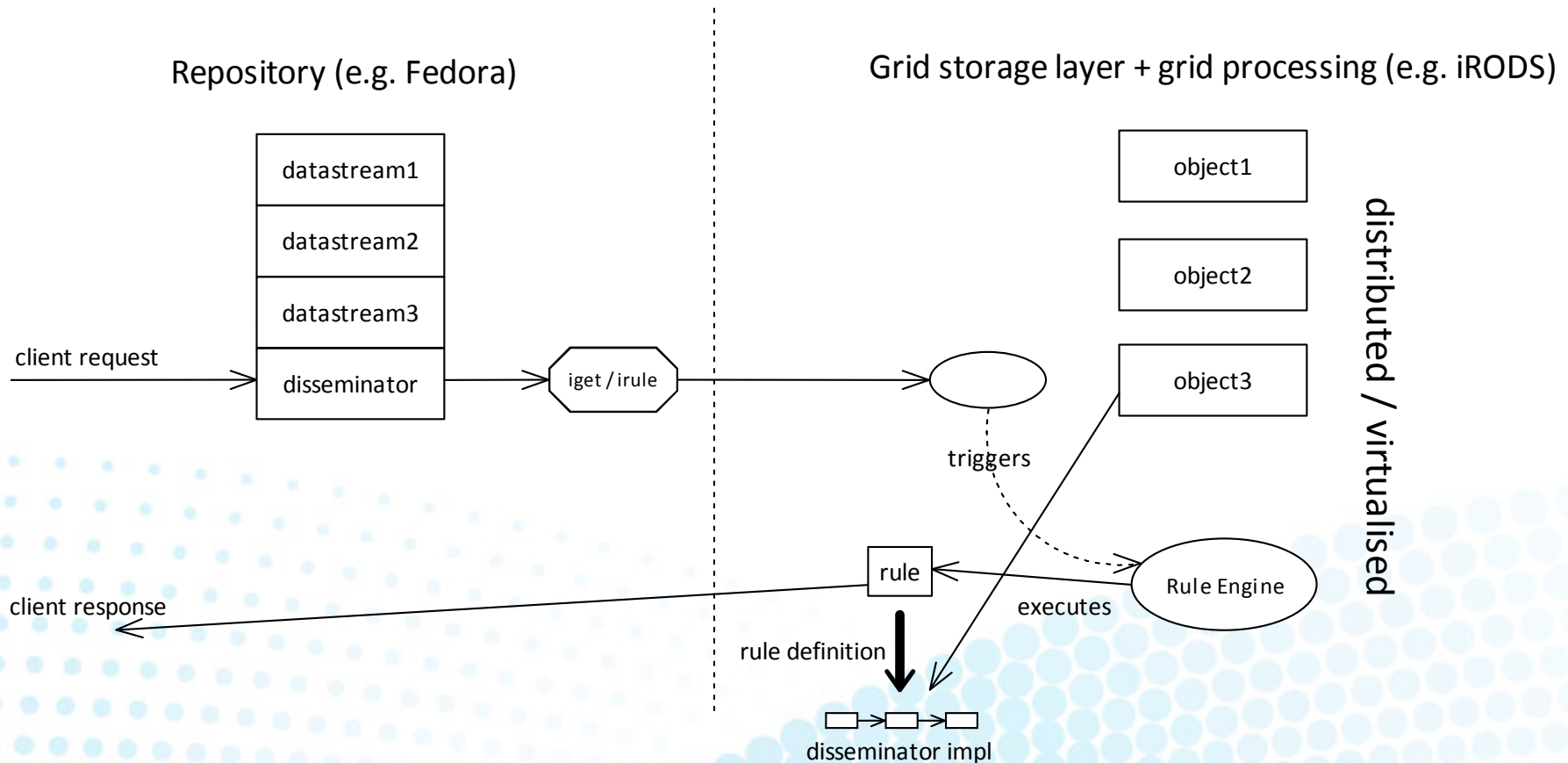
# Repository + grid storage



# Repository + grid storage + grid processing

- Digital objects distributed in grid
- (Some) processing distributed in grid
- Execution of data-side workflows
- Grid still layered beneath repository
- Examples:
  - King's College London
  - SHAMAN

# Repository + grid storage + grid processing



# Repository + grid storage + grid processing

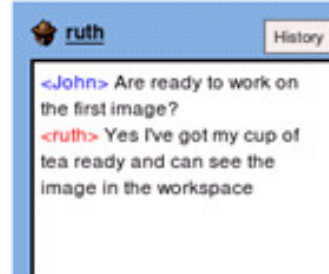
- Workflows can be executed close(r) to data
- Processing/workflows can be sent to most appropriate nodes for execution
- Support for automated curation/preservation (essential for large quantities of highly structured material)

# Repositories as grid resources

- Repository data distributed & virtualised
- Repository services distributed & virtualised
- Repository is itself a (grid) resource, containing other (grid) resources – virtualised, structured data
- Approaches: OGSA-DAI, semantic web (e.g. musicSpace), OAI-ORE, ...

# Virtual Workbenches/Virtual Research Environments

- Virtual Workspace for the Study of Ancient Documents (Oxford)
  - Degraded/damaged documents on papyrus, wood and stone
  - Provide tools to work with them
- Virtual Vellum (Sheffield)
  - Access to digitised manuscripts (integrated with grid storage)
- TextGrid (Germany)
  - Grid-based environment for collaborative text processing
  - Proposed repository integration





# DARIAH

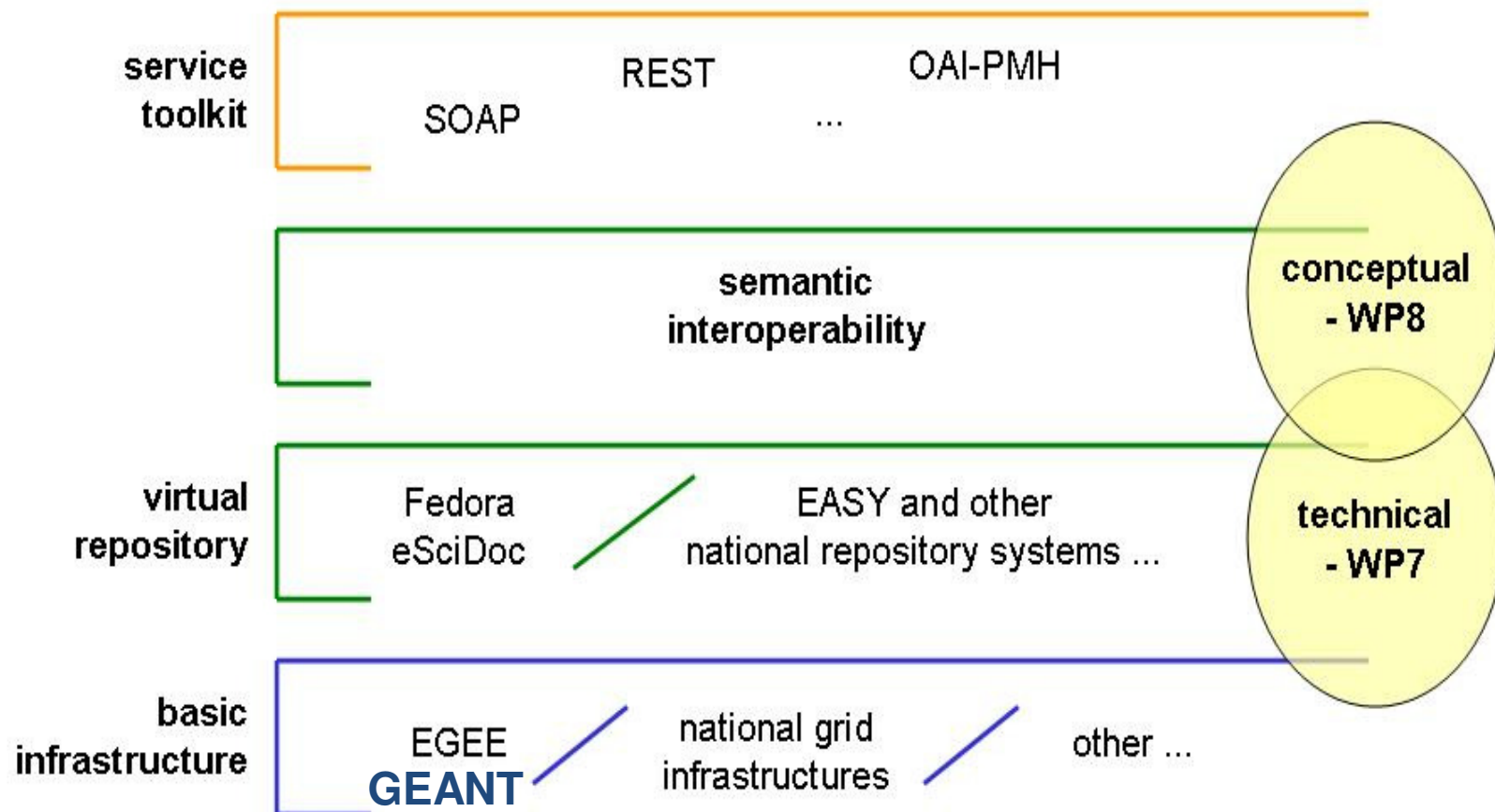


## Digital Research Infrastructure for the Arts and Humanities

Vision: To facilitate **long-term access to,**  
and **use of,** all European humanities &  
cultural heritage **digital information**



# Technical Architecture



# Conclusion

- Challenge is in the (nature of the) data
- Humanists are using digital repositories (among other techniques) to represent/model information
- Humanists are using grids (among other techniques) to collaborate & link up “stuff”
- But: grid not the only option, or necessarily the best for all situations



# Shameless Plug



EPSRC-funded e-Science Network:  
Digital Repositories in the e-Science World  
[www.dresnet.net](http://www.dresnet.net)

First event:  
Special Session  
on repositories





Thank you

