

JISC

# JISC Conference 2007



## Repository ecology: EThOS, the new UK e-theses service, national and institutional repository interaction

R. John Robertson

Anthony Troman

Susan Copeland

Paul Needham

Sponsored by



R. John Robertson

JISC-CETIS and Centre for Digital Library Research

University of Strathclyde

[rr@cis.strath.ac.uk](mailto:rr@cis.strath.ac.uk)

Sponsored by



## A repository ecology?



Brewbrooks (2007)

<http://creativecommons.org/licenses/by-nc-sa/2.0/>

<http://flickr.com/photos/brewbrooks/397298796/>

- *The idea of an information ecology*
  - A way of thinking about the relationships between information systems and services
- *Information systems are like ecologies:*
  - Multiple interacting things
  - Dependent on each other
  - Connections are not always obvious to participants
- A repository ecology as a type of information ecology
- Acknowledgements

# An information ecology

“We believe that the ecology metaphor provides a distinctive, powerful set of organizing properties around which to have conversations. The ecological metaphor suggests several key properties of many environments in which technology is used. An information ecology is a complex *system* of parts and relationships. It exhibits *diversity* and experiences continual evolution. Different parts of an ecology *coevolve*, changing together according to the relationships in the system. Several *keystone species* necessary to the survival of the ecology are present. Information ecologies have a sense of *locality*.”

Nardi and O'Day (1999) First Monday 4 (5)

[http://www.firstmonday.org/issues/issue4\\_5/nardi\\_chapter4.html](http://www.firstmonday.org/issues/issue4_5/nardi_chapter4.html)

# Repository ecology: key features (1)

---

## ■ *System*

- changes can affect whole ecology
- local changes not in line with dynamic of the ecology may fail

## ■ *Diversity*

- different kinds of species can work together
- species overlap and duplicate to a degree
- “Monoculture - a fake, brittle ecology - gives sensational results for a short time, then completely fails.”

## Repository ecology: key features (2)

---

### ■ *Coevolution*

- the ecology is constantly changing
  - new things develop
  - existing things continue to develop
  - existing things are used differently

## Repository ecology: key features (3)

---

### ■ *Keystone species*

- critical species needed for ecology to survive; these are often ‘middleware’ - infrastructure and people who make and assist connections

### ■ *Locality*

- ‘name’: what something is used for in a particular location
- ‘habitation’: how this thing sits within a network of relationships with other things

## Benefiting from an ecological analysis

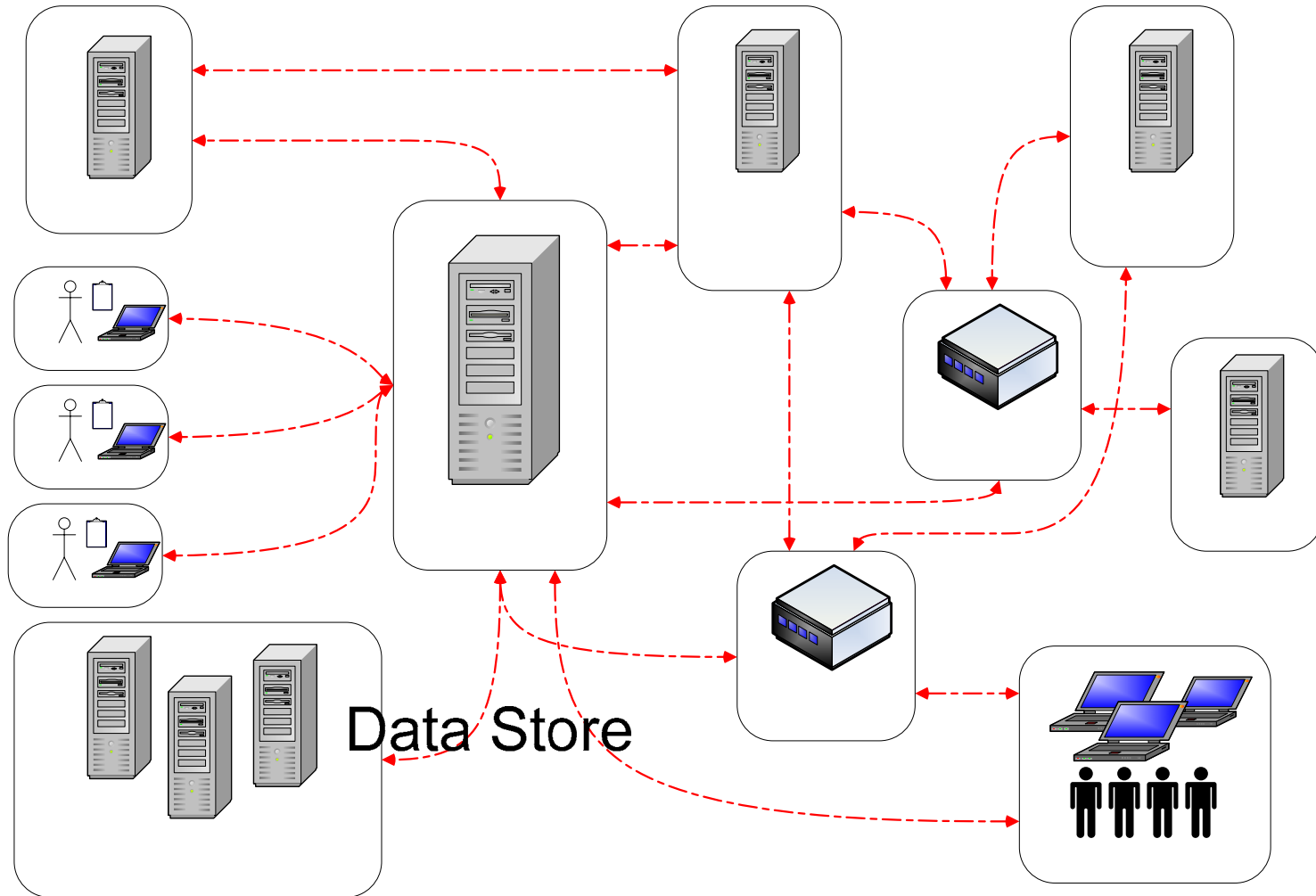
---

### ■ Why is it useful?

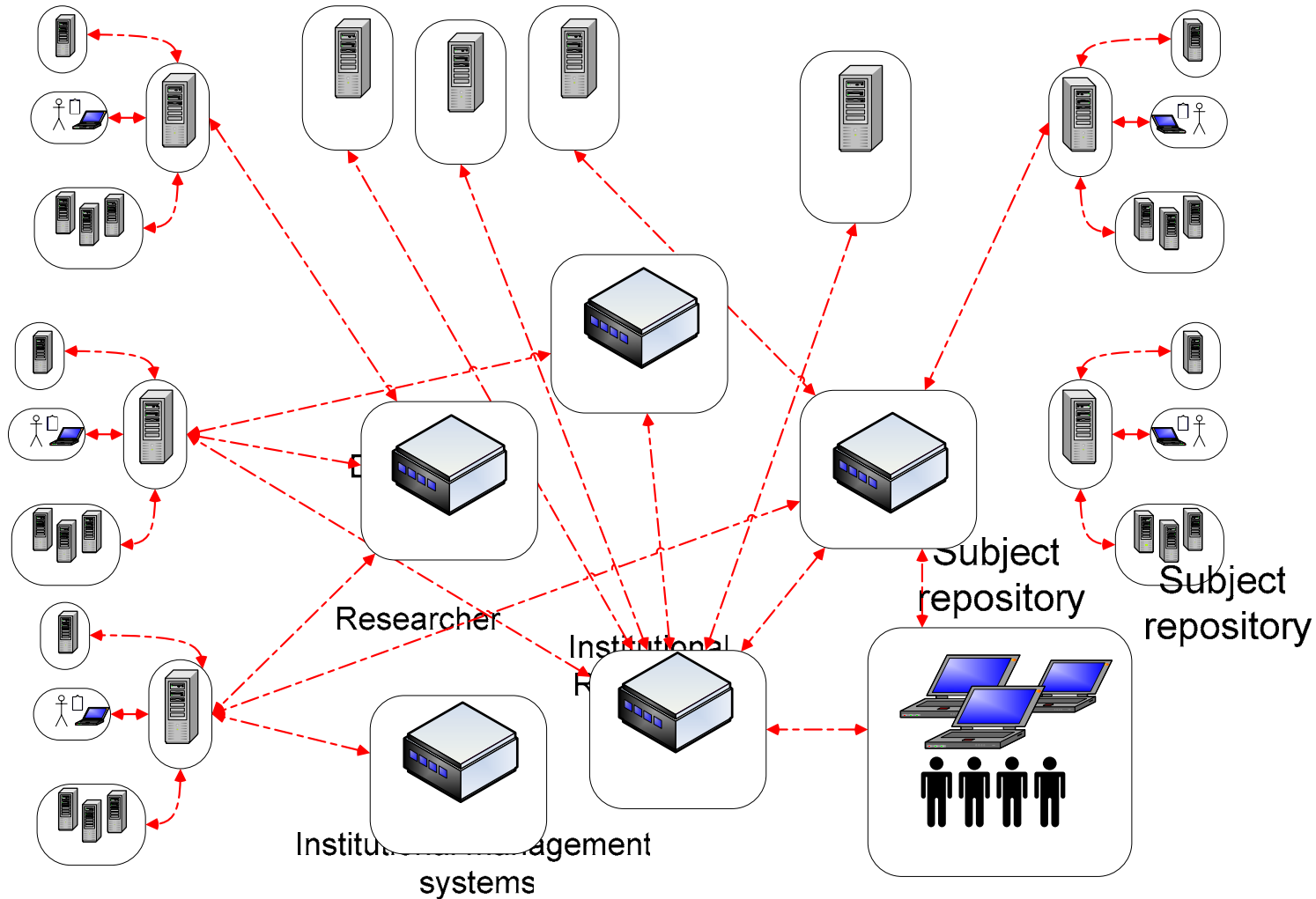
- To an ‘end user’
  - the possibility of richer multifaceted and personalised approaches
- To an administrator
  - the opportunity to looking for efficiencies
  - the opportunity to offer better services
- To developers
  - the opportunity to examine how systems interoperate



## A repository's ecology



## Towards a ecology of repositories and services



- **Key questions for repositories thinking ecologically**
  - What sort of thing (repository or service) is this?
  - What does it relate to (other repositories or services)?
  - What does it depend on?
  - How adaptable is it?
  - What helps it to thrive?

## The EThOS Project

Susan Copeland

Senior Information Adviser (Research)

The Robert Gordon University

[s.copeland@rgu.ac.uk](mailto:s.copeland@rgu.ac.uk)

Sponsored by



## The EThOS project: the opportunity

---

### Repository Ecology

### A Development Opportunity in 2005

Invitation to tender: UK HE e-theses test-bed project

Proposals are invited to undertake a test-bed implementation project to deliver an infrastructure to support the deposit, access and use of research theses for the UK Higher Education (HE) sector. The project will take place over 16-18 months.

[http://www.jisc.ac.uk/fundingopportunities/funding\\_calls/2004/11/funding\\_etheses.aspx](http://www.jisc.ac.uk/fundingopportunities/funding_calls/2004/11/funding_etheses.aspx)

## The EThOS project: the proposal

---

### **A CURL-BL Project to Support e-Theses for UK Higher Education**

“The purpose of EThOS is to deliver, over a period of 18 months, a fully operational, easily scalable and financially viable prototype UK e-theses online service, and supporting infrastructure, that will enable end-users, via one single web interface, to access the full text, in secure format, of electronically stored theses after a selection from a UK Database of Theses”.

(Extract from the EThOS Project Proposal Executive Summary)

### Ecological Issues Addressed by the Project

- What sort of repository / service is to be created?
- How will EThOS relate to other repositories or services?
- What will EThOS depend on?
- How adaptable will EThOS be?
- What will help EThOS to thrive?

## What should EThOS be?

---

### **Learning from previous JISC funded projects: the 'FAIR' Programme**

- The 'Electronic Theses' Project (Consortium led by RGU)
- 'Daedalus' (The University of Glasgow)
- 'Theses Alive!' (The University of Edinburgh)

The need to:

- Maximise the role of institutional repositories
- Create a central hub
- Accommodate the different needs of different institutions
- Take into account the different stages of repository development in H.E.
- Create a flexible service to accommodate changing I.T developments, changing repository content, etc.

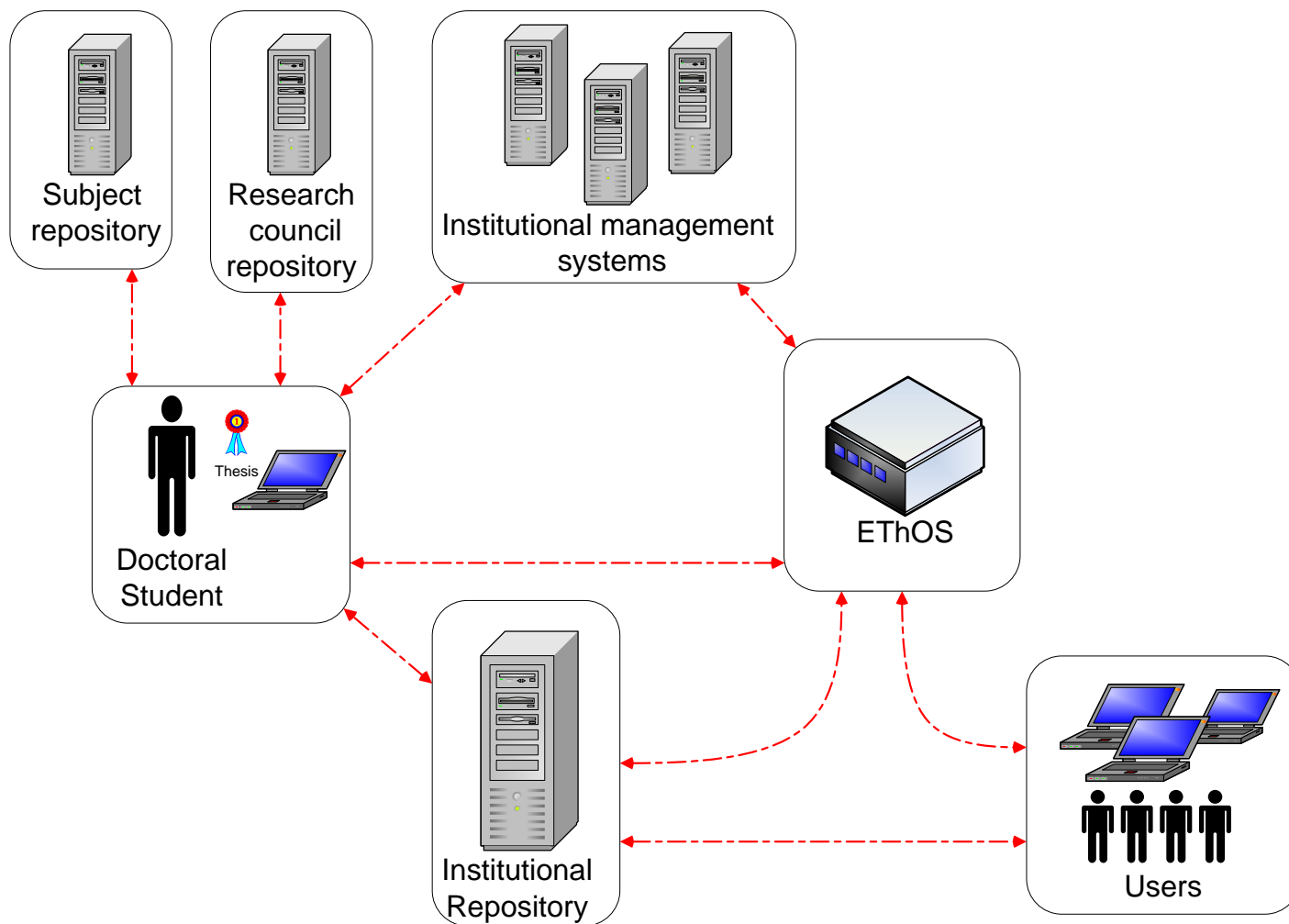
## The EThOS project: relationships

---

**The FAIR programme projects and the EThOS project identified various relationships which would have to be taken into account in the creation of a 'live' service**

- Institutional repositories and the central hub
- Theses and other material types within IRs
- Theses within EThOS and theses within subject repositories
- The role of the British Library & the National Library of Wales
- The role of EThOS in the UK and the Networked Digital Library of Theses and Dissertations (NDLTD) internationally

## A student's view of the EThOS ecology



**The EThOS project identified critical success factors, and tested a prototype e-theses service based on a central hub at the British Library, with reference to:**

- Cultural and administrative issues – advocacy, changes to institutional policies and procedures, content, etc.
- Legal issues – addressing copyright, IPR, and plagiarism concerns
- Technical issues – within institutions, at the BL, between IRs and the BL
- Financial issues – set up and running costs (at institutional level and at service level)

# How adaptable can EThOS be?

**The EThOS project recognised the need to create a service that takes account of the diverse requirements of stakeholders and the changing nature and content of theses**

- Some UK HE institutions have established IRs, some require advice on establishing an IR, some choose to 'buy-in' commercial IR support, some consider it is not cost effective at present to maintain their own repository
- Institutions use different software for their IR e.g. DSpace, Eprints, Fedora, in-house, etc.
- Institutions require different levels of metadata to be provided
- Some institutions allow theses to be deposited in electronic format, some encourage it, some mandate it
- Some institutions wish to participate in EThOS but to keep the full text of their theses within their own IR, some wish provide the full text to the central hub and rely on the BL for a complete service, some wish to include the full text in their IR but to have the assurance that a 'back up' copy is kept at the BL.

# The EThOS Project: the results

---

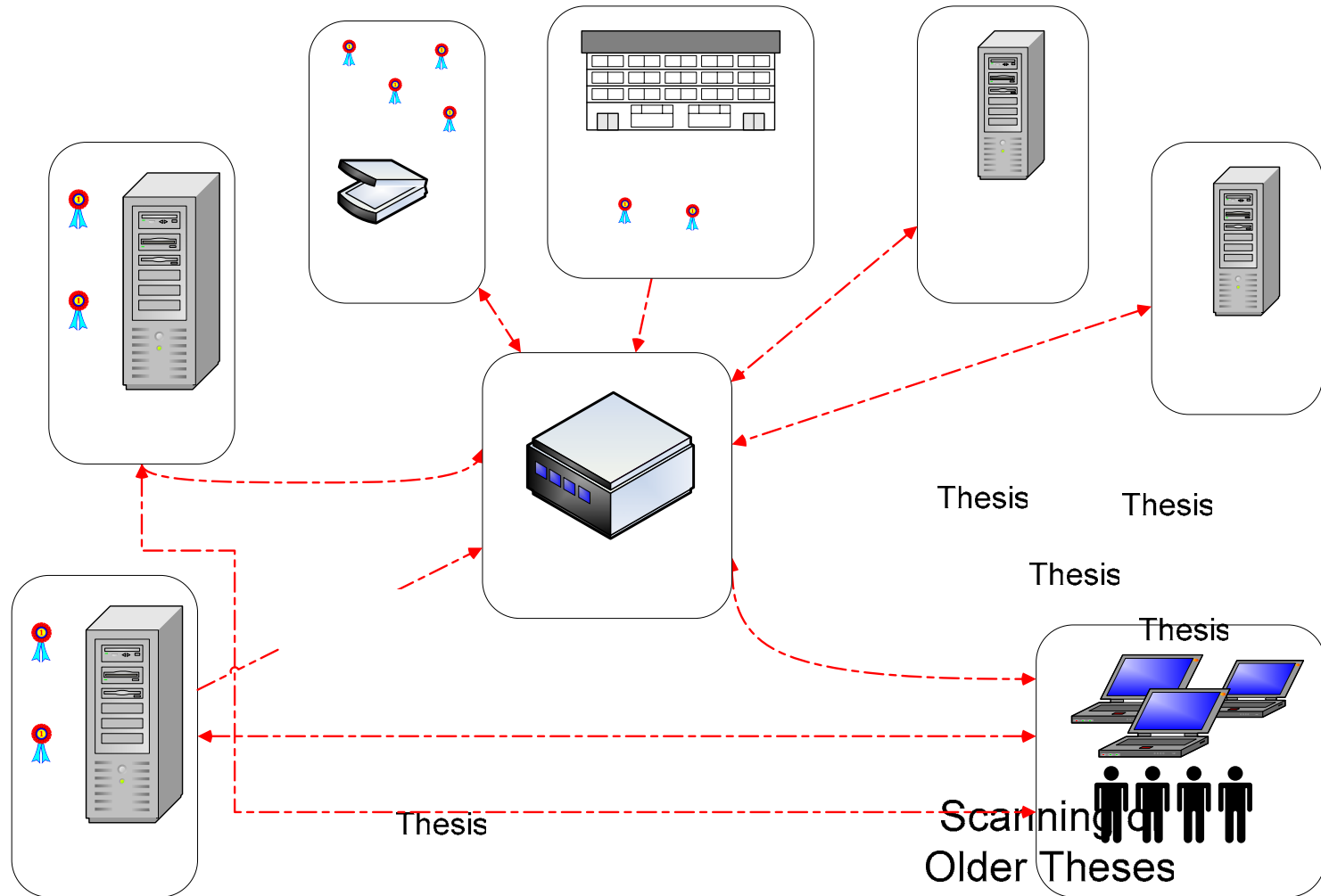
**The EThOS project aimed to create and test an e-theses service that is robust and sustainable yet also capable of adapting and evolving to meet changing user needs**

- The EThOS Toolkit provides advice and guidance to allow institutions to participate fully in the EThOS service. (It is available online and will be updated to accommodate future changes to the service)

<http://ethostoolkit.rgu.ac.uk/>

- The EThOS Toolkit provides advice in a way that ensures maximum value is obtained through participation in EThOS while minimum restrictions are placed on institutions (e.g. through the provision of a metadata core set)
- The EThOS project built on the work of earlier JISC funded e-theses projects, and international developments, but also considered new elements such as the potential value retrospective digitisation
- The project demonstrated that the prototype e-theses service could become a reality – and it identified further developmental issues which will to be addressed in the forthcoming EThOSnet project

## The EThOS ecology



## The EThOS Project: “what helped it thrive?”

---

**As participants involved in the creation and testing of a prototype, the partners in the EThOS project benefited from the knowledge that a successful project should lead to the creation of an improved thesis service within the UK and an improved level of dissemination of UK research output globally.**

The project thrived partly because:

- It built on firm foundations as 3 earlier JISC funded projects had independently reached similar conclusions about aspects of the way forward for e-theses in the UK
- The project team had a clear focus on what the new service had to provide – but a flexible approach to how that would be achieved
- Where relevant, the findings of research relating to the development of other types of repositories was adapted and incorporated into the e-theses prototype
- JISC, CURL, the BL and the UK HE community in general are supportive of the development of the ‘live’ service.

## The EThOS Central Hub

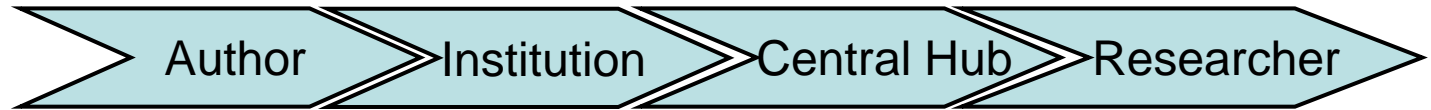
Anthony Troman

British Library Product Development Manager

[anthony.troman@bl.uk](mailto:anthony.troman@bl.uk)

Sponsored by





## ■ The EThOS Central Hub...

- ...is a 'one stop shop' giving access to ALL UK Theses
  - e-theses harvested from Institutional Repositories
  - Paper held in Institutional Libraries
  - Microfilm held at the British Library
- ...digitises paper and microfilm theses and delivers them to the Researcher in the format they require (download, printed, bound, CD/DVD) and stores them in electronic form for future supply – because Researchers want the content of those theses
- ...makes the work of UK PhD authors available to Researchers the world over and demonstrates the quality of research work undertaken by the Institution and in the UK as a whole
- ...respects the Intellectual Property Rights of the Author and the Institution by appropriate licensing of the theses it supplies and recording who a thesis has been supplied to.
- ...offers a viable and sustainable business model while also meeting the requirements of UK HE
  - Open Access
  - Flexible relationships supporting the participation of ALL Institutions regardless of size or output
  - Cost-recovery – every penny generated goes back into the service

## Key question: What sort of thing (repository or service) is this?

---

- The EThOS Central Hub is a service offered by The British Library on behalf of UK HE
  - Serves the Researcher
    - One stop shop, etc
    - Delivery options (download, printed, bound, CD/DVD)
  - Serves the Supplier (the Institution)
    - Maximises access to the Institution's theses
    - Harvests e-theses from IRs with minimal impact on Institutional policies and procedures or the Institutional Repository – maximum benefit for minimum effort
    - Digitisation on a cost-recovery basis
    - Supplies digitised theses back to the originating Institution
    - Preserves supplied e-theses & digitised paper theses
    - Saves manual effort (and therefore cost) in the Institution as more and more theses are available direct to the Researcher from the IR and the Central Hub
    - Saves space in the Institution's Library – digitised theses can be stored away from immediate access
  - Business Model
    - Open Access (UK HE requirement) but offers Institutions options in their relationship (next slide)
  - Intellectual Property Rights
    - Recognises & respects the IPR of the Author and the Institution
    - Researcher accepts usage terms & conditions on ordering thesis
    - Author licence supplied by toolkit

- The major costs of the service are digitisation, so why digitise?
  - Very few e-theses (but growing)
  - Researchers want the content of paper theses – 80% of current service sales are on theses up to 13 years old
  - To give the service a ‘critical mass’ and further encourage the move to e-
  - To put in place the means to deliver a fully e- service in the future
  - Maintaining the current loss making service is non-viable
- Institutional Relationships
  - Open Access Sponsor (Advance Payment):
    - Annual sponsorship paid in advance to cover the cost of digitisation of the institution’s theses ordered ‘on-demand’ by Researchers
    - Guaranteed number of theses digitised from paper and microfilm per annum
    - Guarantees income for the cost-recovery service, so guaranteeing the service
    - Level of sponsorship based on JISC banding
  - Open Access Associate (PAYG):
    - No annual sponsorship, digitisation paid for on a piecemeal basis
  - Associate:
    - Institution supplies theses to the service, first researcher pays for digitisation
  - Collaborator:
    - Supplies metadata only – all orders are routed to the institution
- Diversity = options for Institutions = inclusive of Institutions of all sizes

### ■ Other repositories

- Harvesting e-theses from and supplying digitised theses to Institutional Repositories
- The Institution's physical library - supply of paper theses for digitisation
- Wider initiatives
  - Euro-portals
  - NDLTD (International portal)
  - Search engines (Google, etc.)

### ■ *Co-evolves* with supplying Institutional Repositories

- Data transfer interfaces from IRs
  - Evolving standards, etc.
  - Data transfer design means minimal impact on Institutions on upgrade
- Business model
  - Alternative sources of funding (e.g. government)
  - Etc.

## Key question: What does it depend on?

---

- Continued production of doctoral theses
- Continued involvement of Institutions
  - Cost-effective
  - Efficient
  - Fit for purpose
  - Value of money
  - Etc.
- Continued, relevant service to Researchers
  - As above
- Although the Central Hub is a *keystone species* in the EThOS ecology, the following are NOT:
  - The British Library – the Central Hub could be hosted elsewhere, but can they be ‘trusted’ to uphold the rights of the Author and Institution and offer cost-recovery services?
  - JISC – vital role in the evolution of EThOS and enablers via initial funding and experience
  - CURL - enablers via initial funding and representatives of key suppliers (sponsors)
  - Business Model – there are other business models, but none which offer Open Access to all e-theses (including digitised)

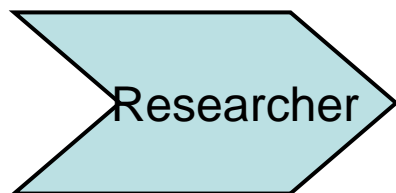
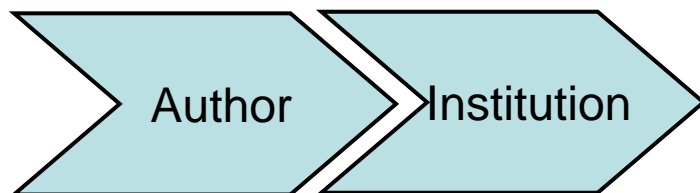
### ■ Technology

- Modular design of the Central Hub means individual elements can be updated piecemeal
- Admin interface allows easy adjustment of variable data (sponsorship fund update, digitisation charges, supply timescales, etc.)
- Harvesting interfaces to IRs are developed to standards and will *co-evolve* with IRs

### ■ Business Model

- Very adaptable – the Institution's relationship to the Central Hub is under the control of the Institution...
- But if the service does not attract enough sponsors, particularly in the early years, it will not be viable
- It will adapt over the years - as more and more theses are digitised there are fewer and fewer needing to be digitised and costs decrease significantly

## Key question: What helps it thrive



### ■ Supply of material

- A service is only as good as the breadth and quality of material it offers
- UK research is high quality and well regarded worldwide
- Good will and involvement of Institutions and Authors

### ■ 'Sales'

- Offer a quality product (see above) to a market which wants it
- Analysis of worldwide trends in demand for e-theses shows that they are in great demand

### ■ Remain relevant

- The developed service meets the requirements of the Google generation (“What have you got? When can I have it? How much will it cost?”)
- The technology can be updated and improved in a piecemeal manner thanks to modular design
- The Business Model can adapt to changing circumstances over the years

## Repository ecology: the next step?

R. John Robertson

JISC-CETIS and Centre for Digital Library Research

University of Strathclyde

[rr@cis.strath.ac.uk](mailto:rr@cis.strath.ac.uk)

Sponsored by



## A repository ecology – a way forward



Brewbrooks (2007)

<http://creativecommons.org/licenses/by-nc-sa/2.0/>

<http://flickr.com/photos/brewbrooks/397238796/>

# Developing the repository ecology

---

- EThOS' illustration of the value of thinking ecologically
  - As a way of addressing repository management and workflow issues
  - As a way of communicating and talking about such issues
- Future work
  - Looking at different repository domains; both for domain specific ecologies and wider ecologies in which domain specific features interact, and will look at boundary issues.
    - RRT are preparing a report for JISC about the repository ecology
    - Presenting this concept at OAI5

### ■ Information Ecology

- Bonnie A. Nardi, and Vicki L O'Day, First Monday Vol 4 No 5 May 3, 1999. Information ecologies: using technology with heart. Chapter Four: Information ecologies. [http://www.firstmonday.org/issues/issue4\\_5/nardi\\_chapter4.html](http://www.firstmonday.org/issues/issue4_5/nardi_chapter4.html)
- Thomas H. Davenport, Information ecology, OUP, 1997

### ■ Repository Ecology and related work

- R. Heery and A. Powell, Digital Repositories Roadmap: looking forward <http://www.ukoln.ac.uk/repositories/publications/roadmap-200604/>
- Rachel Heery and Sheila Anderson, Digital Repositories Review, UKOLN and AHDS, 2005 (Final version) [http://www.jisc.ac.uk/uploaded\\_documents/digital-repositories-review-2005.pdf](http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf)
- R. J. Robertson and J. Barton, "Optimising Metadata Workflows in a Distributed Information Environment", Digital Repositories: Interoperability and Common Services, The Foundation for Research and Technology, 11-13 May 2005 , Hellas (FORTH), Heraklion, Crete, 2005, <http://delos-wp5.ukoln.ac.uk/forums/dig-rep-workshop/robertsonbarton.pdf>
- J. Barton and R.J. Robertson, "Developing a metadata lifecycle model" Workshop at CoLIS 5, June 2005, <http://mwi.cdlr.strath.ac.uk/Colisworkshop.htm>
- Kerry Blinko and Neil McLean, 'A 'Cosmic' View of the Repositories Space (Wheel of Fortune)', 2004, <http://www.rubric.edu.au/extrfiles/wheel/main.swf>