# Open archiving in UK universities

Stephen Pinfield

University of Nottingham

**SHERPA**

# 'Open archiving'?

△ 'Open'

    = 'Freely available' as in 'Budapest **Open** Access Initiative'

    = 'Interoperable' as in '**Open** Archives Initiative'

△ 'Archiving'

    = 'Mounting a paper on the web' as in 'self **archiving**'

    = 'Curating and preserving' as in 'digital **archiving**' (and Open **Archival** Information System)

**SHERPA**

# So what am I talking about?

△ Institutional open access OAI-compliant e-print repositories in UK universities

– 'Institutional' – content created by staff and students of an institution

– 'e-prints' – electronic versions of research papers and other similar output, including:

- 'pre-prints' (pre-referred papers)
- 'post-prints' (post-refereed papers)
- conference papers
- book chapters
- reports
- etc.

**SHERPA**

# UK OAI activity before 2002

△ Centralised subject-specific archives

- – e.g. CogPrints
- – RePEc
- – arXiv mirror

△ 'Closed' use of OAI

- – e.g. Resource Discovery Network

△ Experimental institutional repositories

- – e.g. Nottingham
- – Glasgow
- – etc.

SHERPA

# The case for the institutional approach

△ **Encouraging use of e-prints**
  – within a wider range of subject disciplines
  – amongst more individual researchers

△ **Institutions have**
  – resources to subsidise repository start up
  – technical / organisational infrastructures to support repositories
  – an interest in disseminating content

△ **Needs testing in practice**

**SHERP△**

# Institutional work: early lessons

Δ This is not a technical issue

**SHERP**

# Institutional work: early lessons

Δ This is not a technical issue

– Maturity of OAI Protocol

– Availability of OAI-compliant software

- eprints.org
- Dspace?

SHERPA

# Institutional work: early lessons

△ This is not a technical issue

△ Some important collection management issues need investigating

**SHERPA**

# Institutional work: early lessons

Δ This is not a technical issue

Δ Some important collection management issues need investigating

- document type: pre-prints, post-prints, other objects?
- document format: HTML, PDF, postscript, etc?
- digital preservation policies: what will be preserved and how?
- submission procedures: how will files be formatted and then deposited?
- IPR policies: the rights of the author, institution and publisher?
- metadata quality standards: who creates metadata and according to what standards / quality thresholds?

**SHERPΔ**

# Institutional work: early lessons

△ This is not a technical issue

△ Some important collection management issues need investigating

△ Economic models need further exploration

**SHERP**

# Institutional work: early lessons

△ This is not a technical issue

△ Some important collection management issues need investigating

△ Economic models need further exploration

– what are the economics of maintaining repositories in the short and long term?

– relationship to commercial publishers?

**SHERP**

# Institutional work: early lessons

△ This is not a technical issue

△ Some important collection management issues need investigating

△ Economic models need further exploration

△ The biggest challenge is getting content

**SHERPA**

# Institutional work: early lessons

△ This is not a technical issue

△ Some important collection management issues need investigating

△ Economic models need further exploration

△ The biggest challenge is getting content

- how should advocacy be approached?
- what are the incentives / barriers to self-archiving?
- how best can researchers be enabled to self-archive?
- are there discipline-specific issues?

SHERPA

# JISC FAIR Programme

△ JISC – *Joint Information Systems Committee*

△ FAIR – *Focus on Access to Institutional Resources*

△ Aim: "to support the disclosure of institutional assets":

– "To support access to and sharing of institutional content within Higher Education (HE) and Further Education (FE) and to allow intelligence to be gathered about the technical, organisational and cultural challenges of these processes.…"*

– "This programme is inspired by the vision of the Open Archives Initiative (OAI), that digital resources can be shared between organisations based on a simple mechanism allowing metadata about those resources to be harvested into services"*

* http://www.jisc.ac.uk/pub02/c01_02.html

**SHERPA**

# FAIR projects

△ 14 Projects

△ 'Clusters'

- – Museums and images
- – E-prints
- – E-theses
- – IPR
- – Institutional portals

△ Running from Summer 2002 onwards (1-3 year projects)

△ Total funding: £3 million (US $4.7 million)

**SHERPA**

# Not just e-prints

△ Institutional content of various kinds: e-prints, e-theses, e-learning materials etc.

- HaIRST *(Harvesting Institutional Resources in Scotland Testbed)* – Lead: University of Strathclyde
- DAEDALUS *(Data providers for Academic E-content and the Disclosure of Assets for Learning, Understanding and Scholarship)* – University of Glasgow

△ Plus e-theses, images and cultural objects

**SHERPA**

# Addressing the problems

- △ RoMEO *(Rights MEtadata for Open archiving)*
  - University of Loughborough
  - investigating rights issues for e-prints
  - developing guidelines for adding rights information to OAI-compliant metadata

- △ TARDis *(Targetting Academic Research for Deposit and Disclosure)*
  - University of Southampton
  - investigating overcoming technical, cultural and academic barriers to institutional repositories
  - developing working model of multi-disciplinary institutional repository

**SHERP△**

# Two other examples

Δ SHERPA *(Securing a Hybrid Environment for Research Preservation and Access)* – Data Provider

Δ ePrints UK – Service Provider

**SHERPA**

# SHERPA

- △ **Initiator**: CURL *(Consortium of University Research Libraries)*
- △ **Development Partners**: Nottingham (lead), Edinburgh, Glasgow, Leeds, Oxford, Sheffield, British Library, York, AHDS (Arts and Humanities Data Service), plus others
- △ **Duration**: 3 years, November 2002 – November 2005
- △ **Funding**: JISC and CURL
- △ **Aims**:
  - to set up a series of institutional OAI-compliant e-print (pre and post-print) repositories using eprints.org software
  - to investigate key issues in populating and maintaining e-print collections, including advocacy in the research community
  - to work with Service Providers to achieve acceptable standards for metadata exchange and the dissemination of the content
  - to investigate OAIS-compliant digital preservation
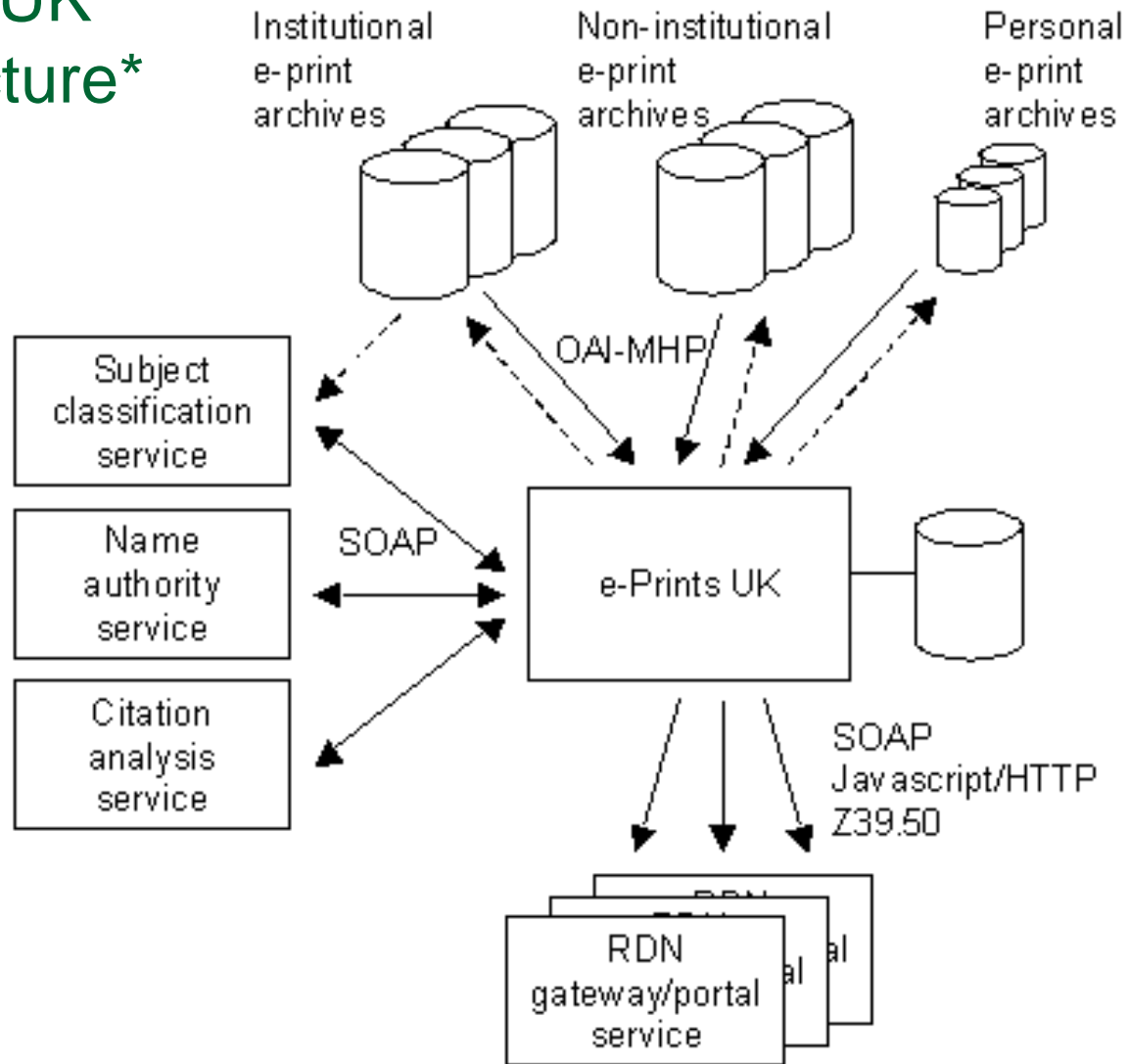  - to disseminate learning outcomes and advocacy materials, including providing detailed advice to others

**SHERPA**

# ePrints-UK

△ **Partners**: RDNC (Resource Discovery Network Centre) at King's College London, UKOLN (UK Office for Library and Information Networking) at Bath, OCLC; universities: Southampton, Leeds, Bristol, Heriot Watt, Birmingham, Manchester Metropolitan, Oxford, Nottingham, UMIST

△ **Duration**: Summer 2002 – Summer 2004

△ **Aims**:

– To set up as a Service Provider gathering metadata from institutional, disciplinary and personal Data Providers

– To enhance records, via web services, with:

- Automatic subject classification
- Authority headings
- Citation analysis resulting in OpenURL citations

– To deliver search interfaces through RDN Hubs

**SHERPA**

# ePrints-UK architecture*



Institutional e-print archives

Non-institutional e-print archives

Personal e-print archives

OAI-MHP

Subject classification service

Name authority service

Citation analysis service

SOAP

e-Prints UK

SOAP
Javascript/HTTP
Z39.50

RDN gateway/portal service

* reproduced with permission

# Advocacy

- Researchers are not interested in the 'serials crisis' *per se*
- Impact and access barriers are key problems for researchers
- Emphasis should be put on the benefits for:
  - the researcher: *e.g. wide dissemination, rapid dissemination, ease of access, interoperability, value added services*
  - the institution: *e.g. raising profile of institution, asset management, long-term cost savings*
  - the research community: *e.g. freeing-up communication*
- Researchers' concerns should be taken seriously *e.g. quality control, IPR, undermining journals, work load*
- Emphasis on complementing not necessarily replacing journals
- Does not *have* to mean pre-prints
- 'The library will do the work'

**SHERPA**

# Mountains still to climb

△ Engaging the interest of researchers

△ Translating interest into action

△ Sustaining interest and action

**SHERP**

# http://www.sherpa.ac.uk

Stephen.Pinfield@Nottingham.ac.uk

**SHERPA**