Preserving repository content: practical steps for repository managers

Open repositories 2010

6-9 July, Madrid, Spain

Miggie Pickton, Steve Hitchcock, Simon Coles,
Debra Morris, Stephanie Meece

Outline

- Preservation and repositories

- The KeepIt project

- Preservation objectives

- The KeepIt training course

- Meeting preservation objectives

- Steps to preservation readiness

Preservation and repositories

- Permanent access to scholarly outputs is at the heart of the open access movement:

  - "Open Access means *immediate, permanent, free online access to the full text of all refereed research journal articles*" [1]

- Long term or permanent access is a stated goal of many repositories…

- … but few repositories have implemented practical action plans for preservation

- Why?

[1] Harnad, S. (2005) The Implementation of the Berlin Declaration on Open Access. *D-Lib Magazine*, **11**(3).
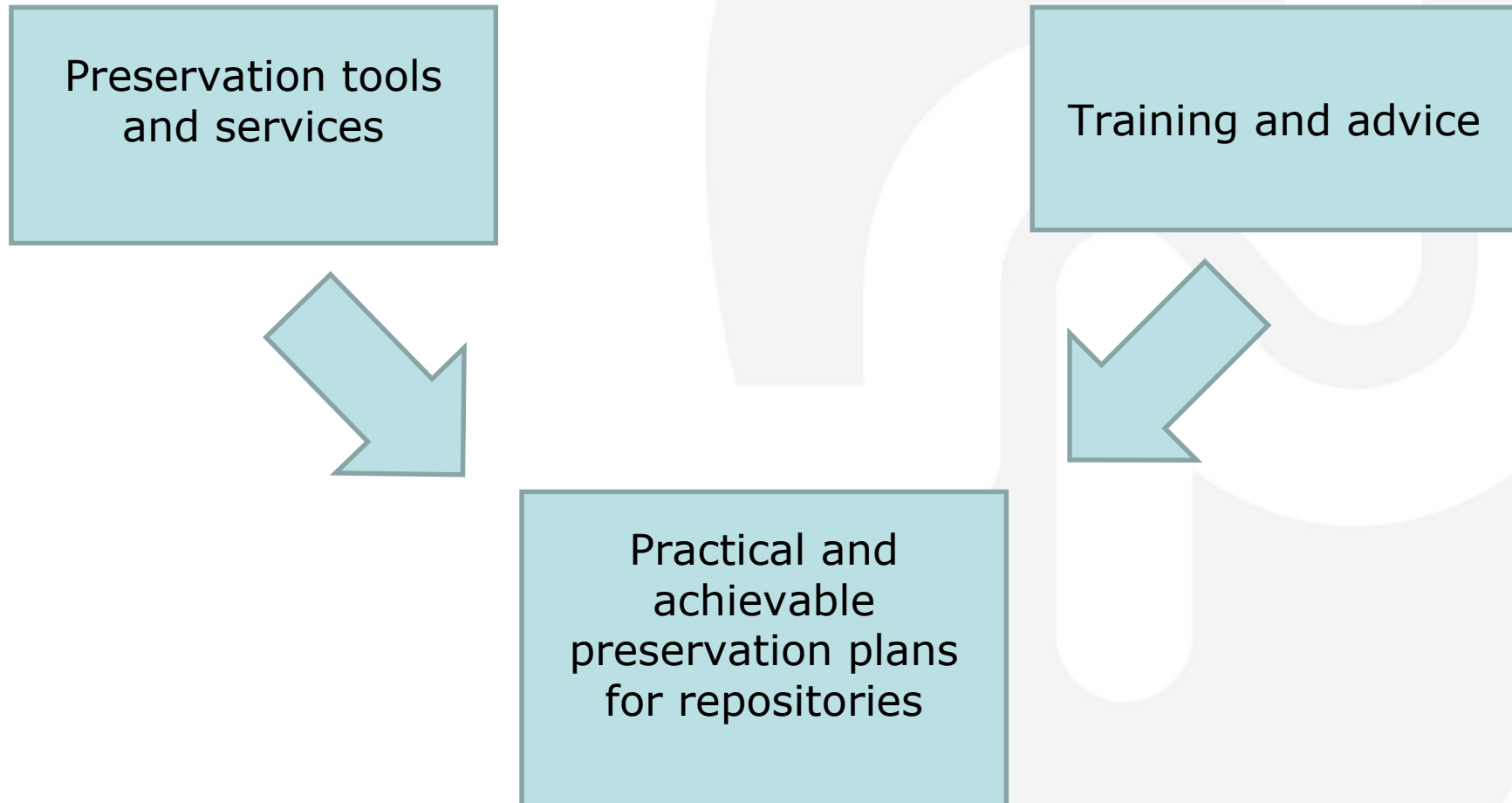
Why so slow to address preservation issues in repositories?

- Preservation tools and services exist, but few have addressed the specific needs of repositories

- In practical terms they often necessitate action that is additional to rather than integral to repository workflow

- Repository content is typically highly varied and complex:

  - Descriptive metadata and file formats used inconsistently

  - Items deposited by those without knowledge or expertise in managing digital assets

- Repository managers are inexperienced in digital preservation and lack **time** and confidence to deal with it.
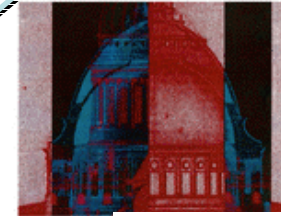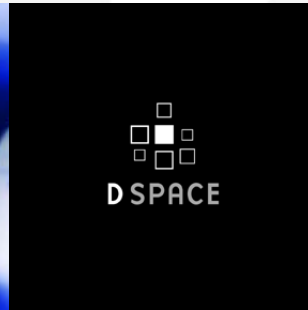
The KeepIt project

Preservation tools and services
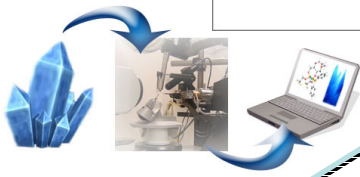
Training and advice

Practical and achievable preservation plans for repositories

KeepIt project exemplars

NECTAR

University of the Arts London

Research

Arts

FedoraCommons™

D SPACE

Learning experience

Teaching

Course material

E-mail Books

Discussion Photos

Web sites Audio/video

Search engines

= learner = teacher

Science
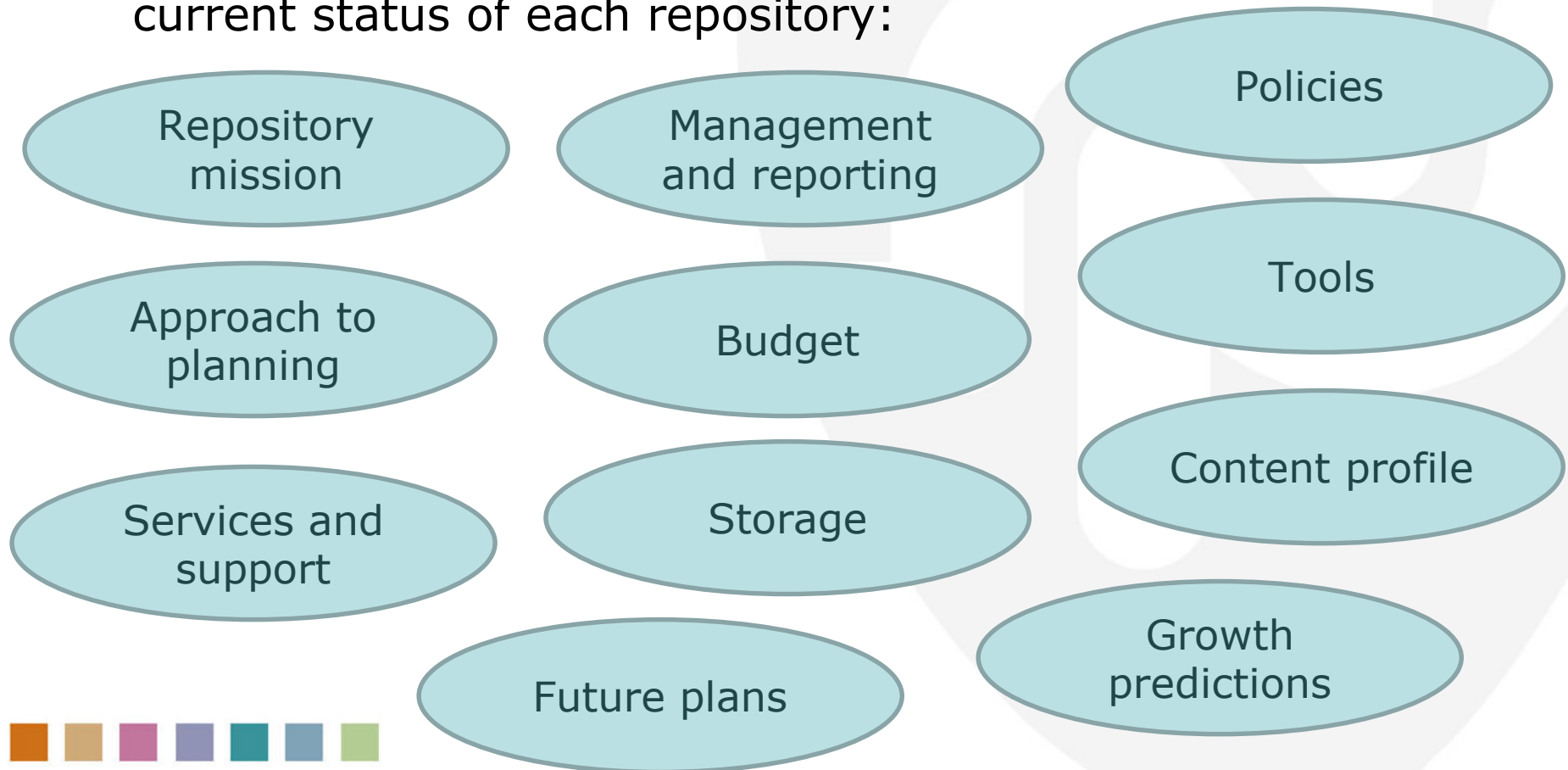
Teaching

eCrystals

EdShare
University of Southampton
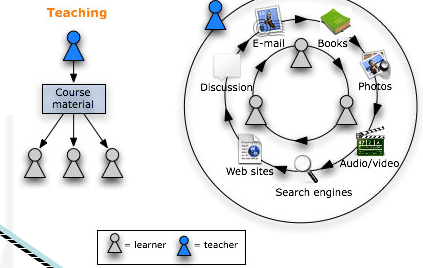
KeepIt exemplars – starting out

- Structured one to one meetings between exemplar managers and the preservation specialist to establish the current status of each repository:

Repository mission

Management and reporting

Policies

Approach to planning

Budget

Tools

Services and support

Storage

Content profile

Future plans

Growth predictions

KeepIt exemplar objectives

- Common themes:

  – Tools and procedures e.g. to deal with different file formats

  – Costs e.g. for supporting business plans and funding bids

  – Organizational issues e.g. institutional and user concerns, advocacy, training and documentation

- Plus repository-specific objectives e.g.

  – responsibilities of content creator and repository (EdShare)

  – Facilitating (automating) preservation actions (eCrystals)

Eprints preservation toolkit

- Developed for EPrints 3.2.

- Incorporates:

  - Storage (including cloud storage)

  - Format management (characterisation)

  - Risk assessment

  - Preservation planning

  - Preservation action e.g. migration

- ... enabling a complete preservation workflow

The KeepIt training course

Module 1, Organizational issues, audit, selection and appraisal
School of ECS, University of Southampton, 19 January 2010

Module 2, institutional and lifecycle preservation costs
School of ECS, University of Southampton, 5 February 2010

Module 3, Primer on preservation workflow, formats and characterisation
Westminster-Kingsway College, London, 2 March 2010

Module 4, Putting storage, format management and preservation planning in the repository
University of Southampton, 18-19 March 2010

Module 5, Trust, of the repository, of the tools and services it chooses
University of Northampton, 30 March 2010

Course tools
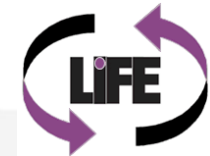
- Data Asset Framework (DAF) - identify, locate, describe and assess research data assets

- Assessing Digital Institutional Assets self assessment toolkit (AIDA)

- Keeping Research Data Safe (KRDS) – benefits and costs of a repository

- LIFE[3] – predictive costing tool for digital content

- Eprints preservation toolkit

- DROID & JHOVE – file format identification and characterisation

- PREMIS - data dictionary for preservation metadata

- Plato - preservation planning tool from PLANETS

- Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) – repository risk assessment and reporting

Meeting preservation objectives - eCrystals

- All four exemplars will upgrade to Eprints 3.2 and implement the Eprints preservation tools

- **eCrystals:**

  - will add their two main file formats (CIF and CML) to DROID as part of an investigation into the automatic validation and verification of content;

  - have published initial cost data as part of the Keeping Research Data Safe (KRDS2) study;

  - are working to understand and develop the relationship between a research data repository and the host institution or research community in terms of migration of preservation plans.

Meeting preservation objectives - EdShare

- has identified the most prevalent file types in the repository and in the university's VLE (Blackboard) and is working with IT colleagues to generate statistical information;

- is planning some comparative analysis with other institutional learning and teaching repositories (how similar/dissimilar is EdShare?);

- is exploring the specific institutional concerns with respect to preserving resources for learning and teaching – what does this mean for institutional beneficiaries/stakeholders e.g. teachers, educational leaders, students?

- is contributing to the 'Southampton Learning Environment' – a framework for supporting, delivering and enhancing learning and teaching across the whole university community.

Meeting preservation objectives - NECTAR

- has upgraded to Eprints 3.2 and is ready for the Eprints preservation tools to be installed;

- has undertaken a Research Data Project using the DAF methodology to establish how researchers at Northampton create, use and manage their data;

- is using the results of this survey to inform policy, guidance and preservation planning for the repository;

- has spread the preservation message by involving colleagues with expertise in technical, metadata and collection management areas in appropriate elements of the KeepIt training course.

Meeting preservation objectives – UAL Research Online

- is dealing with the preservation implications of a range of image, audio and video formats, and beginning to consider archived websites;

- is using the DRAMBORA self-assessment tool to identify the activities and assets of the repository and to identify, assess and calculate the associated risks;

- which will inform guides to digital preservation for researchers, repository staff and senior managers;

- but Stephanie notes: "We all know that we very much need to assess, manage and minimise risk, but preservation tasks tend to fall into the 'Important' category, not the 'Urgent' one. It's easy to spend six months attending to 'Urgent' work, and never get to any of the 'Important' bits."

KeepIt wins

- Pursuing a 'preservation' agenda has delivered wider benefits to the repository managers:

  - Increased our own understanding of the content of our repositories

  - Enabled greater engagement with content providers

  - Raised awareness (of the repository as well as digital preservation) among colleagues and managers

  - Provided the opportunity for us to reflect on our repository's current status and think strategically about its future

- Convinced us that it <u>is</u> possible to formulate practical and achievable preservation plans.

Steps to preservation readiness

1. **Know the institutional context**. A preservation plan must meet the needs of the institution and its stakeholders;

2. Develop **preservation policy** appropriate to your institution's and users' needs;

3. Make a **business case** for preservation;

4. Identify an appropriate preservation **metadata** schema to describe your institution's types of output;

5. Identify **tools** to support preservation planning and decision-making;

6. Consider **storing** repository content in multiple locations;

7. **Promote** the preservation services offered by the repository. Build trust among your user community.

Acknowledgement

We are grateful to the JISC for funding the KeepIt project and the earlier Preserv and Preserv2 projects upon which it draws.

JISC

Further information

- Miggie Pickton, **NECTAR** Repository Manager, University of Northampton, miggie.pickton@northampton.ac.uk

- Steve Hitchcock, **KeepIt** Project Manager, University of Southampton, sh94r@ecs.soton.ac.uk

- Simon Coles, **eCrystals** Repository Manager, University of Southampton, S.J.Coles@soton.ac.uk

- Debra Morris, **EdShare** Repository Manager, University of Southampton, D.Morris@soton.ac.uk

- Stephanie Meece, **UAL Research Online** Repository Manager, University of the Arts, London, s.meece@arts.ac.uk

Or check out the KeepIt project blog at:

http://blogs.ecs.soton.ac.uk/keepit/