Reaching researchers through their data: a Data Asset Framework case study

Repositories Support Project Winter School

9th to 11th February 2011

Miggie Pickton

Outline

- Introduction - The University of Northampton and NECTAR
- The KeepIt project – preservation and repositories
- The Data Asset Framework
- The Research Data Project at Northampton – 4 steps
- Findings and recommendations
- Research data and NECTAR
- Research data and your repository

Introduction - research at UoN

- About The University of Northampton:
  - Achieved university status and research degree awarding powers in 2005
  - Aiming to become "a leading regional, national and international centre for research and knowledge transfer" (from university's Strategic Vision, 2005)
- Increased focus on research and supporting the research community
- 150+ research students, ??? research active staff – numbers are rising
- Like everyone else... thinking about the REF

Introduction - NECTAR

Northampton Electronic Collection of Theses And Research

- Institutional repository set up to '**showcase** and **preserve**' university research
- Developed 2007, launched 2008
- Content to include articles, book chapters, presentations, creative outputs etc
- Authoritative source for university research reporting e.g. the Annual Research Report

NECTAR

- With 2000+ records, the '**showcasing**' was coming along...
  (even if most contained only metadata ☹)
  ... but we hadn't given much thought to **preservation**

- Preservation seen as:
  - Important but not urgent (too many other priorities)
  - A bit difficult and scary
  - But something that we <u>should</u> be doing

- Fortunately, help was at hand...

The KeepIt project

- The JISC-funded KeepIt project aimed to bring together existing **preservation tools and services** with appropriate **training and advice** to enable repository managers to formulate practical and achievable **preservation plans**
- Led by a preservation expert – Steve Hitchcock
- Featured four exemplar repositories:
  - eCrystals (science data)
  - EdShare (educational resources)
  - UAL Research Online (arts)
  - NECTAR (research)
- A further 12 repository managers attended the KeepIt training course

JISC

KeepIt 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
- LIFE³ – 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

Eureka!

- Could see an immediate benefit in several tools, but particularly the **Data Asset Framework** from the Digital Curation Centre
- What is DAF?
  "The Data Asset Framework is a set of methods to:
  - find out what data assets are being created and held within institutions;
  - explore how those data are stored, managed, shared and reused;
  - identify any risks e.g. misuse, data loss or irretrievability;
  - learn about researchers' attitudes towards data creation and sharing;
  - suggest ways to improve ongoing data management."

(Digital Curation Centre, 2009, p.3)

Why conduct a DAF project?

- Little was known centrally about university researchers' data storage requirements, or the research workflow that incorporates the creation and management of data
- No university wide data storage policy or procedure existed
- Research funders are beginning to demand that data as well as published research outputs are made openly available
- In NECTAR, we had available the infrastructure to store and preserve digital data
- *Reaching the researchers*… previous studies had noted that the process of undertaking DAF had been valuable in itself, even if the resulting inventory of data was only partial

Research Data Project – four steps

- The DAF methodology comprises four steps:
  - "Stage 1 is for planning, defining the purpose and scope of the survey and conducting preliminary research.
  - Stage 2 is about identifying what data assets exist and classifying them to determine where to focus efforts for more in-depth analysis.
  - Stage 3 is where the information life cycle is considered to understand researchers' workflows and identify weaknesses in data creation and curation practices.
  - Stage 4 pulls together the information collected and provides recommendations for improving data management."

(Digital Curation Centre, 2009, p.5)

Step 1 – Planning and preliminaries

"Stage 1 is for planning, defining the purpose and scope of the survey and conducting preliminary research."

- Ensure buy-in from senior managers – in Information Services and the research community
- Define the aims and scope of the project e.g. to examine researcher data management practices and the risks associated with these; to raise awareness of good data management practice; to gather evidence to inform policy or future services
- Consider practicalities – who will do what, when and to whom in the project? e.g. Project Board, Project Manager, Project Researcher(s)
- Understand the DAF methodology - learn from previous DAF projects

Step 2 – Overview of research data

"Stage 2 is about identifying what data assets exist and classifying them to determine where to focus efforts for more in-depth analysis."

- Arrange meetings with research leaders to gain broad understanding of research practices in our six Schools and support for the project
- Design and pilot online questionnaire survey covering ownership of research data; types and formats of data; storage; security; backups; data sharing; funder requirements; open access to data
- Make survey live – offer incentives for participation and for agreement to interview

Step 3 – Data and the research lifecycle

"Stage 3 is where the information life cycle is considered to understand researchers' workflows and identify weaknesses in data creation and curation practices."

- Conduct one-to-one semi-structured interviews with research active staff and research students
- Follow up and expand on survey responses – determine individuals' data management practices and service needs
- Focus on one specific data object e.g. an audio file containing an interview or the output of a lab-based experiment; complete a standard metadata form
- Engage the researcher in discussion of the role of data in their own research lifecycle and seek their views on future policy and services (including deposit of data in NECTAR)

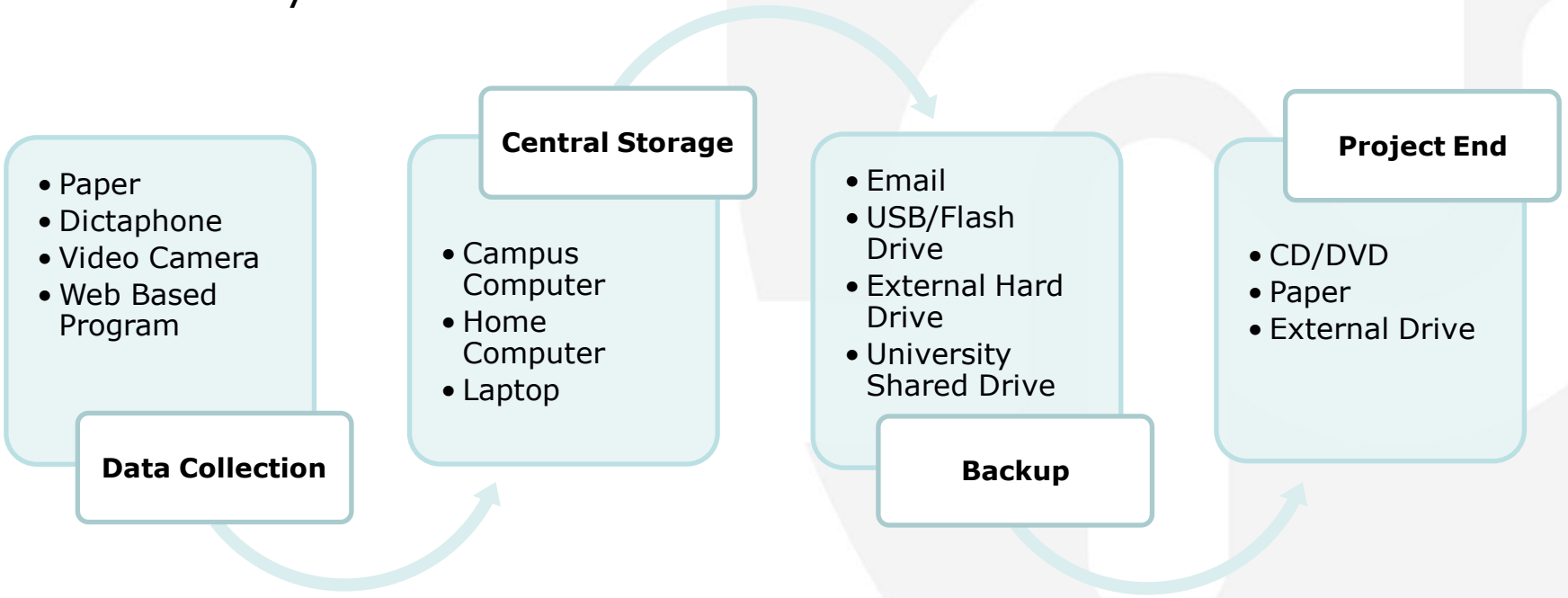
Step 4 – selected findings (1)

"Stage 4 pulls together the information collected and provides recommendations for improving data management."

- 80 researchers responded to the survey and 16 agreed to take part in the follow-up interviews; all Schools were represented
- Some common behaviours identified e.g. overwhelming use of Microsoft software for creating documents and spreadsheets (.doc/.docx and .xls/.xslx files); .jpeg preferred for images
- Greater variation in software and hence file types used for databases, audio and video

Step 4 – selected findings (2)

Data storage needs, behaviours and vulnerabilities vary through the research lifecycle:



A few researchers had previously lost data but most performed regular backups to avoid this.

Step 4 – selected findings (3)

- Researcher views on **open access to data**:
  - 56% of participants agreed that they would like a university repository to store their research data, but not necessarily to offer open access
  - Responses varied by School (Business and Education most in favour, Health and Social Science most against)
  - Examples were given of funders who expressly forbade sharing of data
  - Most researchers had not applied for funding from a body that required open access to research data

Step 4 – recommendations

- Nine recommendations made, covering:
  - Reporting to senior research managers and leaders
  - Creation of research data policy (and procedure to support it)
  - Clarification of ownership of research data
  - Training and guidance (a role for Information Services)
  - Dissemination of findings

(Full results and recommendations are described in the project report – see Alexogiannopoulos *et al*., 2010)

Research Data Project – follow-up

- The Research Data Project report has been presented to Research Committee and disseminated via NECTAR and the DAF website
- The Research Committee has formed a Research Data Working Group to develop a research data policy and procedures to support this. Their proposal is currently out for consultation among the research community.
- A session on data management is about to be introduced into the mandatory research student induction week
- University Records Manager actively involved (good research data management supports his role in dealing with FOI and EIR requests (JISC, 2010))

The Research Data Project and NECTAR

- We had hoped that researchers would have welcomed the opportunity to deposit their data in NECTAR, but the response was luke-warm
- … but we may yet get another repository (a hive?) for research data
- The project gave us the chance to have much more meaningful and in-depth discussions with individual researchers – allowing us to learn more of their needs and to promote our services (including NECTAR)
- The survey and discussions flagged up the full range of research outputs that could potentially end up in NECTAR – valuable information for repository preservation planning.

Research data and your repository

Your new Vice Chancellor has asked you to lead a project to capture all university research data in the institutional repository.

What would you do?

Suggest two ‘quick wins’ that would result in success for your project.

References

- Alexogiannopoulos, E., McKenney, S. and Pickton, M. (2010) *Research Data Management Project: a DAF investigation of research data management practices at The University of Northampton.* Northampton: University of Northampton. Available from: http://nectar.northampton.ac.uk/2736/ [Accessed 31.01.2011].
- Digital Curation Centre (2009) *Data Asset Framework: Implementation guide*. Available from: http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf [Accessed 31.01.2011].
- JISC (2010) *Freedom of Information and research data: Questions and answers.* Available from: http://www.jisc.ac.uk/publications/programmerelated/2010/foiresearchdata.aspx [Accessed 31.01.2011].
- Research Councils UK (2009) *RCUK Policy and code of conduct on the governance of good research conduct: Integrity, clarity and good management*. Available from: http://www.rcuk.ac.uk/documents/reviews/grc/grcpoldraft.pdf [Accessed 31.01.2011].

Acknowledgement

We are grateful to the JISC for funding the KeepIt project; to the Graduate Boost programme for supplying the two project researchers, Sam McKenney and Edward Alexogiannopoulos; and to Sarah Jones and Martin Donnelly of the Digital Curation Centre for their help and support with the DAF and DMP Online tools.

JISC