Data Audit Framework

Experiences gained from implementing the Data Audit Framework

Cuna Ekmekcioglu

Information Services, University of Edinburgh

Data Audit Framework Launch

1st October 2008, British Academy, London

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

Data Audit Framework

Overview

- Background
- Edinburgh implementation
- GeoSciences pilot audit
  - Methodology
  - General issues
  - Lessons learned
- Further data audits

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

Data Audit Framework

Background

Lack of knowledge:

- What data is held
- Where it is located
- How it is managed

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

Data Audit Framework

Background

A recommendation to JISC:

"*JISC should develop a Data Audit Framework to enable all universities and colleges to carry out an audit of departmental data collections, awareness, policies and practice for data curation and preservation*"

Liz Lyon, Dealing with Data: Roles, Rights, Responsibilities and Relationships, (2007)

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

Data Audit Framework

Background

JISC funded 5 projects:

- DAF Development Project
  (HATII, University of Glasgow; King's College London; University of Edinburgh; UKOLN, University of Bath)
- Four pilot implementation projects
  - University of Edinburgh
  - King's College London
  - Imperial College London
  - University College London

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

Data Audit Framework

Edinburgh implementation

- Information Services: reviewing their support for the research community
- ECDF – Edinburgh Compute and Data Facility
- Edinburgh University Data Library: data curation, management, sharing
- DISC-UK DataShare project: a pilot digital repository of multidisciplinary research datasets
- College of Science and Engineering: investigating data storage requirements

Edinburgh implementation

- GeoSciences pilot audit
- Further data audits across three colleges:
  - Science and Engineering:
    School of Physics and Astronomy, School of Biological Sciences
  - Humanities and Social Science:
    The Moray House School of Education, School of History, Classics and Archaeology
  - Medicine and veterinary Science:
    School of Molecular and Clinical Medicine, School of Biomedical Sciences

Data Audit Framework

GeoSciences pilot audit

The School of GeoSciences

- A leading international centre for research into GeoSciences
- 80 academics, 70 research fellows, 130 PhD students
- Annual research grant and contract income of around £4-6 million
- Staff contribute to one or more of five *Research Groups* and may be involved in inter-University *Research Consortia* and *Research Centres*

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

Data Audit Framework

GeoSciences: research

Five main research groups:

- Earth Subsurface Science
- Global Change
- Human Geography
- Edinburgh Earth Observatory
- Centre for Environmental Change and Sustainability

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

GeoSciences: IT facilities

- Unix machines
- Servers - backed up regularly
- 15 Tb research data on the main server
- Oracle Relational Database Management System
- MySQL
- ArcGIS, ARC/INFO
- MatLab, R, Stata

Data Audit Framework

Audit methodology

http://www.data-audit.eu/DAF_Methodology.pdf

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

Data Audit Framework

1. Planning the audit

- Desk research
  - School website, staff home pages, publications, technical documents etc.
  - Identifying key research staff and projects they are responsible for
- Initial meeting with IT Managers
- Setting up interviews with staff

Data Audit Framework

2. Identifying & classifying

- Conducting semi structured interviews with 35 staff mainly in Human Geography and Global Change research groups
  - Interview duration varied between 30 mins and 2 hrs
- Creating an inventory of 25 data assets
- Classifying these data assets

Data Audit Framework

3. Assessing management

- Assessing the most significant assets in detail, collecting a basic set of data elements based on Dublin Core
- Checking the accuracy of the information collected via emails

Data Audit Framework

4. Reporting back

- A draft report that feeds back on the results of each stage of the audit
- Draft recommendations to the school to improve data management
- Draft recommendations to Information Services to further improve their support to the research community

General issues

- Difficulty accessing the School servers
- Not enough time for a questionnaire
- Inappropriate timing
  - Staff were marking exam papers
  - Some staff were away for field trips
  - I9-20 May: University was closed
  - 21-26 May: Exam Boards
- Difficulty contacting staff via emails
- Interviews were time consuming

Data Audit Framework

Lessons learned

- Outcomes preliminary but positive
- Time needed is longer than initially anticipated but still manageable
- Inventory doesn't always have to be comprehensive but could be a representative sample
- There is little documentation / knowledge of what exists
- There are no standards in creating and managing data assets

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

Data Audit Framework

Lessons learned

- Define the scope and granularity carefully
- Ensure appropriate timing
- Plan well in advance
- Get support from senior management
- Collect as much information as possible in the interviews/surveys

Further data audits

- Six further data audits are being conducted in six schools across three colleges:
  - School of Physics and Astronomy
  - School of Biological Sciences
  - The Moray House School of Education
  - School of History, Classics and Archeology
  - School of Molecular and Clinical Medicine
  - School of Biomedical Sciences

Further data audits

- Planning stage completed
- An online survey is being sent to school contacts for circulation:

  https://www.survey.ed.ac.uk/research_data_short_survey

- Interviews to follow in October

Data Audit Framework

Further information

- Project Director

  Robin Rice: R.Rice@ed.ac.uk
- Project Manager

  Cuna Ekmekcioglu: cekmekci@staffmail.ed.ac.uk
- http://www.data-audit.eu

Thank you! Any questions?

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC