Data Audit Framework Development (DAFD) Project

Sarah Jones, HATII

Delos Summer School, 8th – 13th June 2008

JISC

The problem

Lack of knowledge

- what types of data are present within UK institutions?
- how they are managed?
- where they are deposited for long-term preservation?

JISC

The solution

"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

Developing a Data Audit Framework

- DAF Development Project

  (HATII, Glasgow; King's College London; University of Edinburgh; UKOLN, Bath)

- Four pilot implementation projects
  - King's College London
  - University of Edinburgh
  - University College London
  - Imperial College London

JISC

DAFD schedule

April	Develop methodology for collecting data
May-June	Test preliminary methodology through pilot audits Glasgow: archaeology Edinburgh: geosciences King's College: medical UKOLN: engineering
June	Define system requirements & develop prototype
July-August	Implementation and iterative development
September	Release and dissemination

JISC

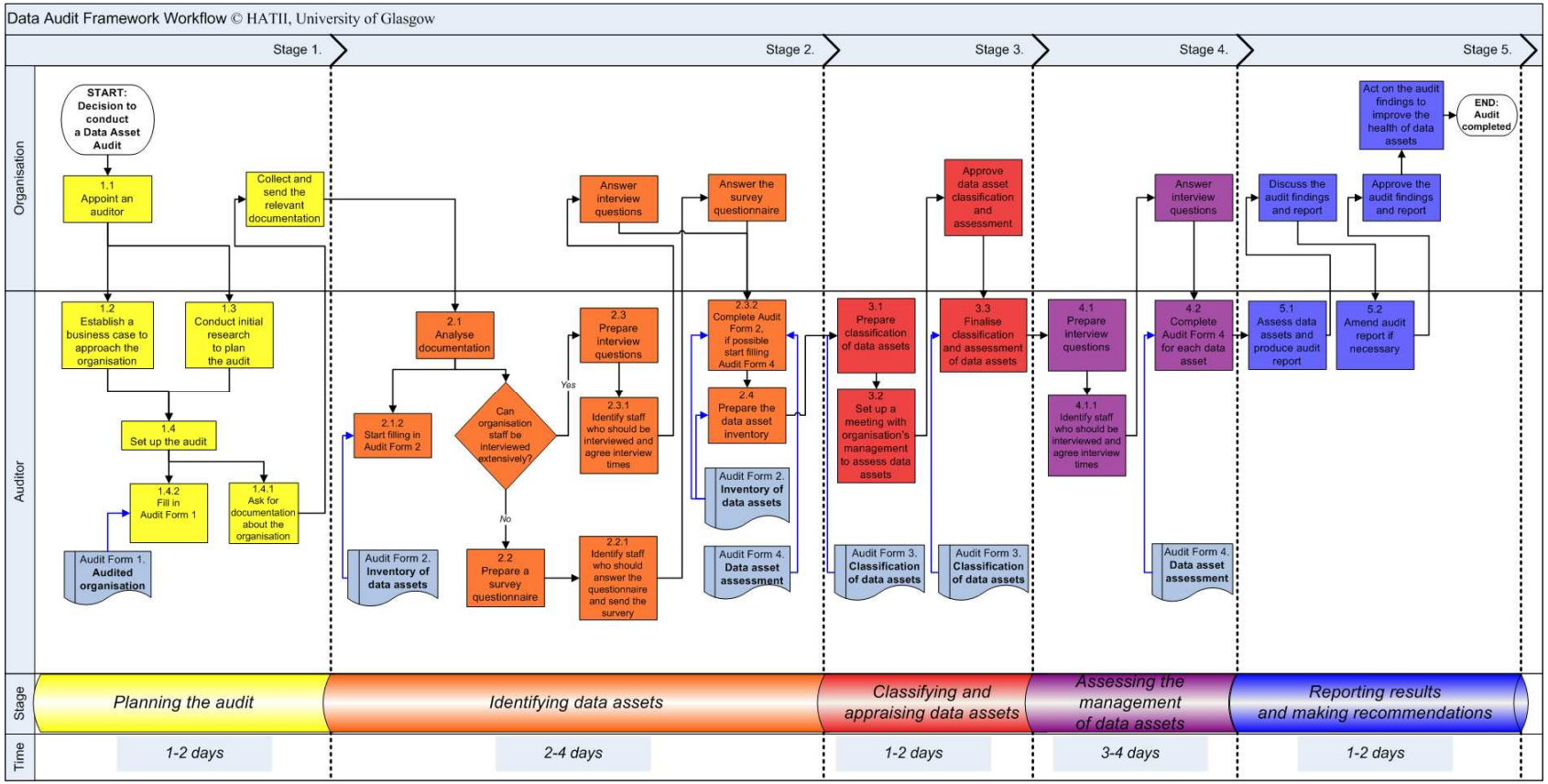
DAF Methodology

Five stages:

- Planning the audit;
- Identifying data assets;
- Classifying and appraising data assets;
- Assessing the management of data assets;
- Reporting findings and recommending change.

JISC

DAF workflow



JISC

Stage 1: Planning the audit

- Selecting an auditor
- Establishing a business case
- Research the organisation
- Set up the audit

Stage 2: Identifying data assets

- Collecting basic information to get an overview of departmental holdings

Audit Form 2: Inventory of data assets				
Name of the data asset	**Description of the asset**	**Owner**	**Reference**	**Comments**
Bach bibliography database	A database listing books, articles, thesis, papers and facsimile editions on the works of Johann Sebastian Bach	Charles Fairall	RAE return for 2007, http://www....ac.uk/...	An MS Access database in H:\Research\Bach\Bach_Bibliography.mdb.

JISC

Stage 3: Classifying and appraising assets

- Classifying records to determine which warrant further investigation

Vital	Vital data are crucial for the organisation to function such as those: ○ still being created or added to; ○ used on frequent basis; ○ that underpin scientific replication e.g. revalidation; ○ that play a pivotal role in ongoing research.
Important	Important data assets include the ones that: ○ the organisation is responsible for, but that are completed; ○ the organisation is using in its work, but less frequently; ○ may be used in the future to provide services to external clients.
Minor	Minor data assets include those that the organisation: ○ has no explicit need for or no longer wants responsibility for; ○ does not have archival responsibility e.g. purchased data.

JISC

Stage 4: Assessing management of assets

- Once the vital and important records have been identified they can be assessed in more detail

- Level of detail dependent on aims of audit
  - Form 4A – core element set
  - Form 4B – extended element set

JISC

Audit Form 4A: Data asset management (Core element set)

No	Parameter	Comment
1	**ID**	*A unique identification assigned by the audior or organisation to each data asset*
2	**Title**	*Official name of the data asset, with additional or alternative titles or acronyms if they exist*
3	**Description**	*A description of the information contained the data asset*
4	**Subject**	*Information and keywords describing the subject matter of the data asset*
5	**Purpose**	*Reason why the asset was created, intended user communities or source of funding / original project title*
6	**Coverage**	*Intellectual domain or subject area covered by the information in the data asset. Spatial and temporal coverage*
7	**Source**	*The source(s) of the information found in the data asset*
8	**Author**	*Person, group or organisation responsible for the intellectual content of the data asset*
9	**Date**	*The date on which the data asset was created or published*
10	**Updating frequency**	*The frequency of updates to this dataset to indicate currency*
11	**Language**	*The language(s) of the data asset content*
12	**Type**	*Description of the technical type of the data asset (e.g., database, photo collection, text corpus, etc.)*
13	**Format**	*Physical formats of data asset, including file format information*
14	**Rights**	*Basic indication of the user's rights to view, copy, redistribute or republish all or part of the information held in the data asset*
15	**Relation**	*Description of relations the data asset has with other data assets and any any DOI ISSN or ISBN references for publications based on this data*

Stage 5: Report and recommendations

- Summarise departmental holdings
- Profile assets by category
- Report risks
- Recommend change

JISC

Pilot audits – lessons learned

- Timing
- Defining scope and granularity
- Merging stages
- Data literacy

JISC

Conclusion

- Outcomes very preliminary but positive
  - Experience confirms data audit is needed
  - Time needed is longer than initially anticipated but still manageable
  - Results will support various other data projects

JISC