



# Data Audit Framework Development (DAFD) Project

---

Sarah Jones, HATII

Delos Summer School, 8<sup>th</sup> – 13<sup>th</sup> 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?



## 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



## 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



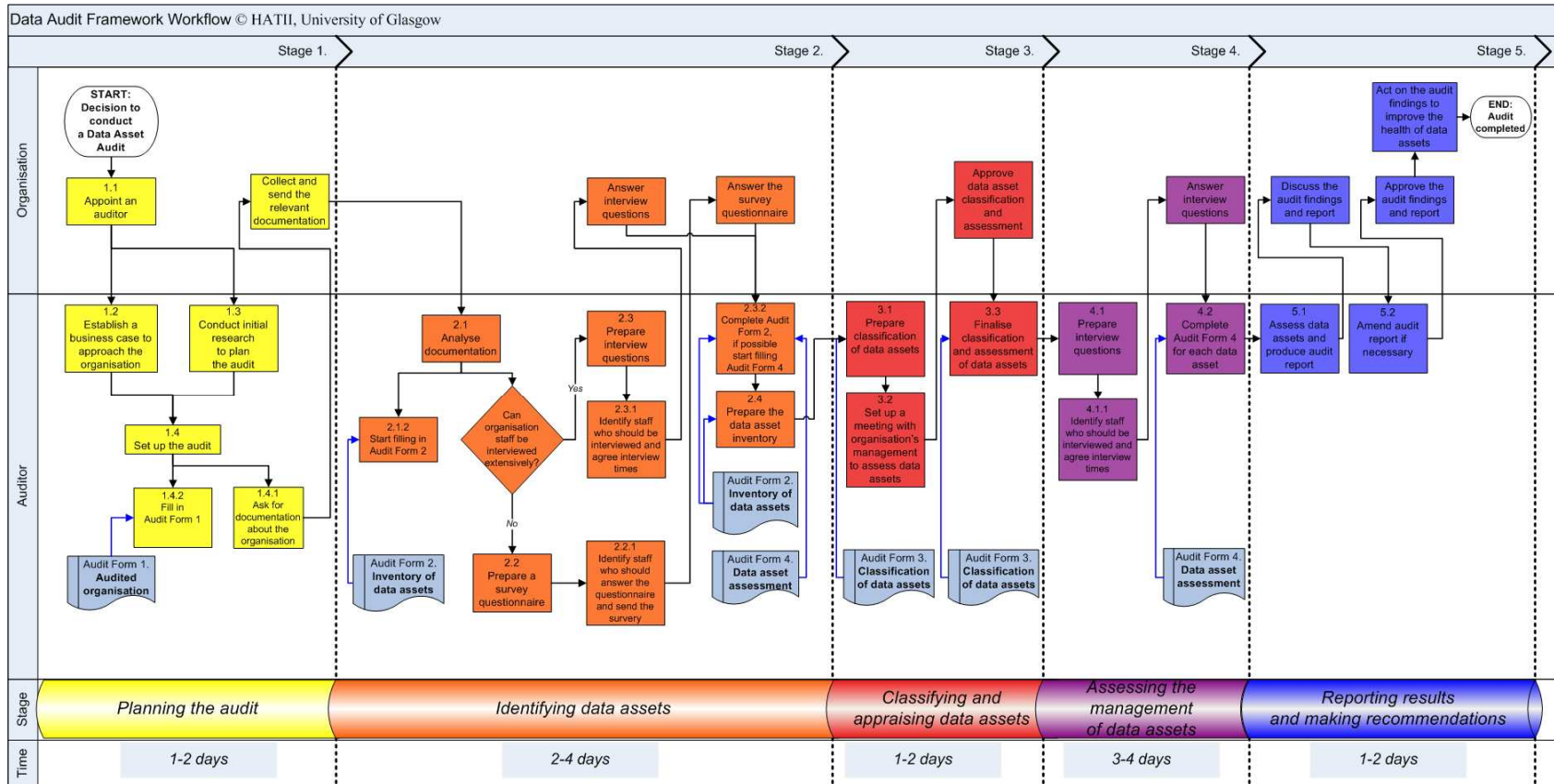
# DAF Methodology

---

Five stages:

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

# DAF workflow





# 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

<b>Audit Form 2: Inventory of data assets</b>				
<b>Name of the data asset</b>	<b>Description of the asset</b>	<b>Owner</b>	<b>Reference</b>	<b>Comments</b>
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, <a href="http://www....ac.uk/">http://www....ac.uk/...</a>	An MS Access database in H:\Research\Bach\Bach_Bibliography.mdb.



## 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: <ul style="list-style-type: none"><li>○ still being created or added to;</li><li>○ used on frequent basis;</li><li>○ that underpin scientific replication e.g. revalidation;</li><li>○ that play a pivotal role in ongoing research.</li></ul>
Important	Important data assets include the ones that: <ul style="list-style-type: none"><li>○ the organisation is responsible for, but that are completed;</li><li>○ the organisation is using in its work, but less frequently;</li><li>○ may be used in the future to provide services to external clients.</li></ul>
Minor	Minor data assets include those that the organisation: <ul style="list-style-type: none"><li>○ has no explicit need for or no longer wants responsibility for;</li><li>○ does not have archival responsibility e.g. purchased data.</li></ul>



## 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



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

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



## Stage 5: Report and recommendations

---

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



# Pilot audits – lessons learned

---

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



# 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