Data Audit Framework

a data management toolkit for research led institutions

Sarah Jones, Seamus Ross & Raivo Ruusalepp

HATII, University of Glasgow

CNI Task Force Meeting, Washington DC, USA

8th December 2008

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

Data Audit Framework

Outline

- Background context
- Methodology and tool
- Pilot audit findings
- Conclusions
- Discussion

JISC | University of Glasgow | Humanities Advanced Technology & Information Institute | DCC

Data Audit Framework

The problem

How can organisations realise the value of their research data when it is unclear:

- what data is held;
- where it is located;
- and how it is being managed?

Data Audit Framework

A recommendation

"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)

Data Audit Framework

Scope of work

- 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
  - Imperial College London
  - King's College London
  - University College London

JISC | University of Glasgow Humanities Advanced Technology & Information Institute | DCC

Data Audit Framework

Benefits of audit

Awareness of data holdings

- Capacity planning
- Facilitate data sharing and reuse
- Monitor data holdings and avoid data leaks

Recognition of data management practices

- More efficient of resources and improved workflows
- Ability to manage risks – data loss, inaccessibility, compliance
- Enables the development of a data strategy

Data Audit Framework

The methodology

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

Data Audit Framework

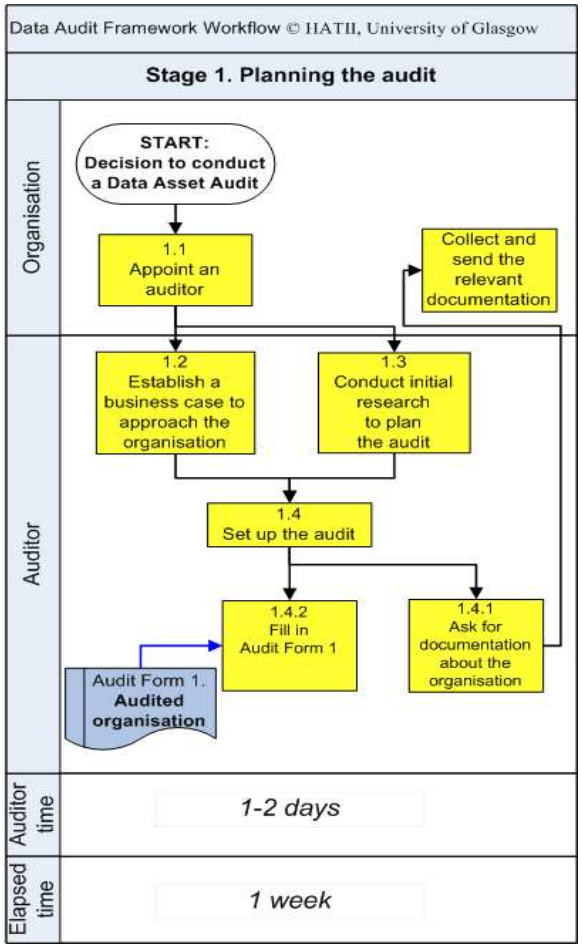
Stage 1: planning

Objective

- Prepare in advance to optimise workflow and time with staff

Process

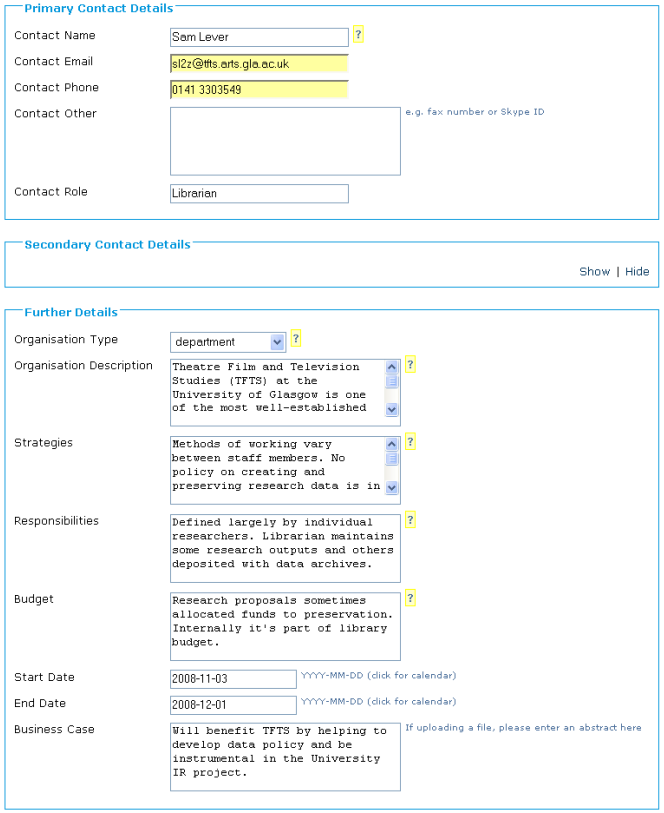
- Define scope / expected outcomes
- Research organisational context
- Set up interviews

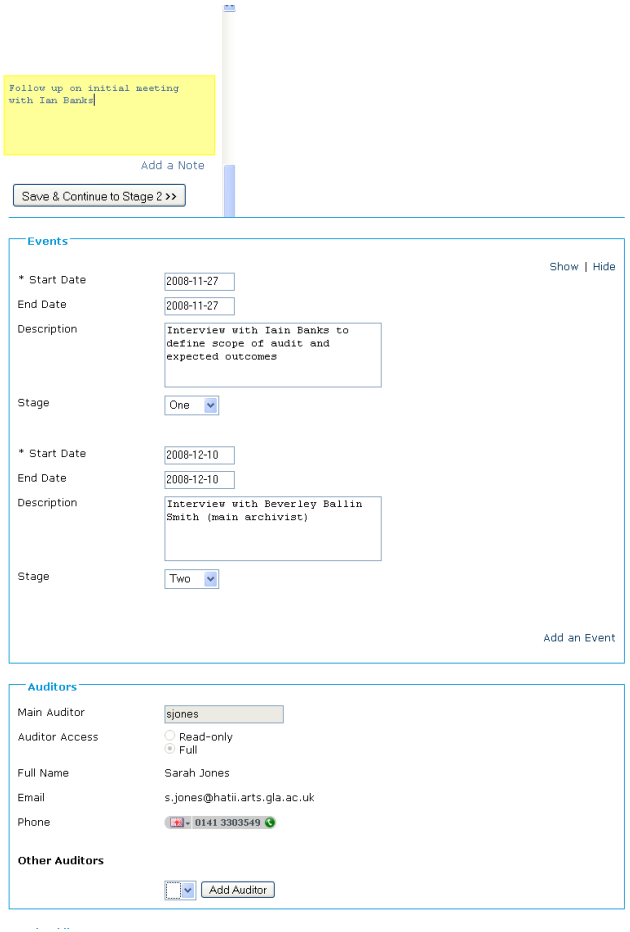


JISC | University of Glasgow | Humanities Advanced Technology & Information Institute | DCC

Data Audit Framework

Online tool: stage 1





Data Audit Framework

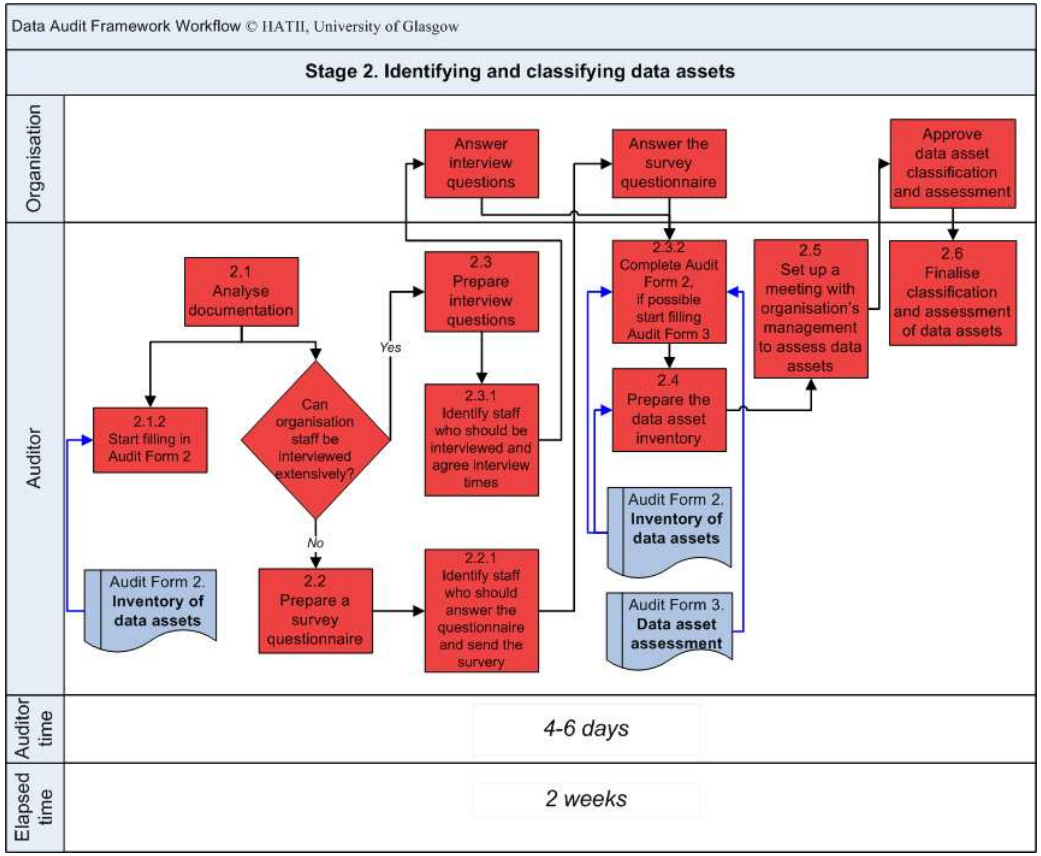
Stage 2: identifying

Objective

- Create an inventory

Process

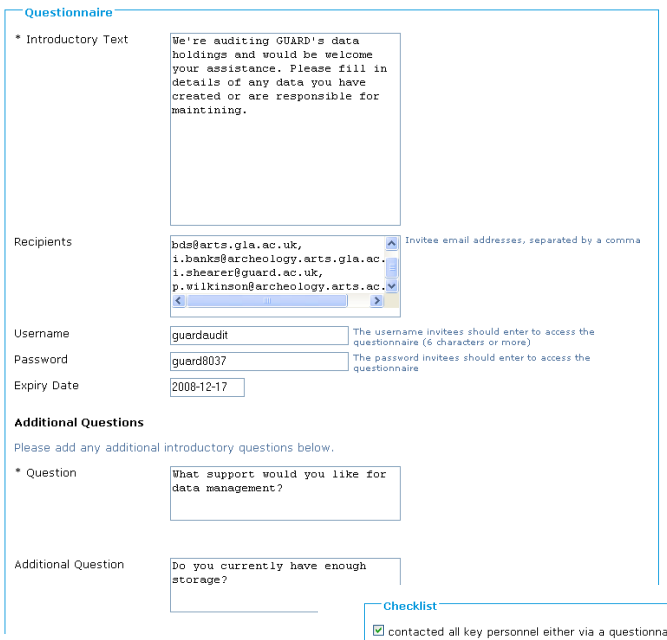
- Identify data
- Classify assets
- Hold interviews, questionnaires, desk-research

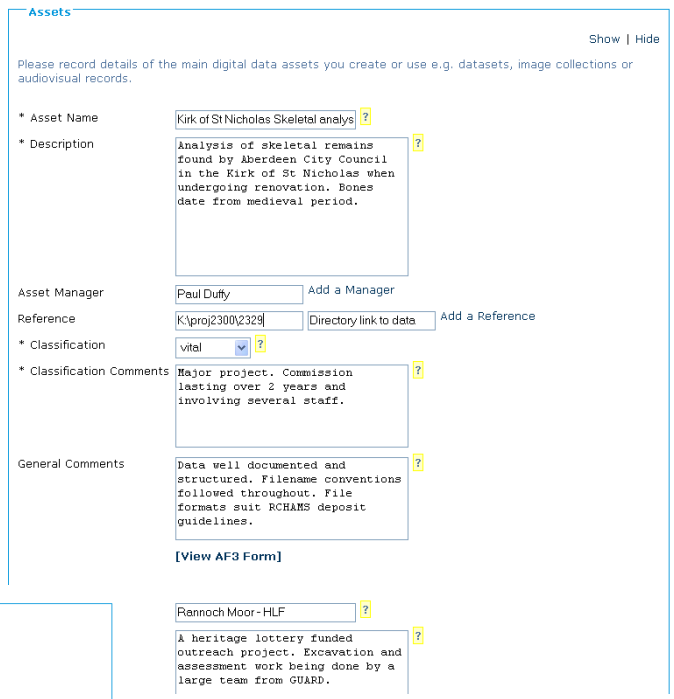


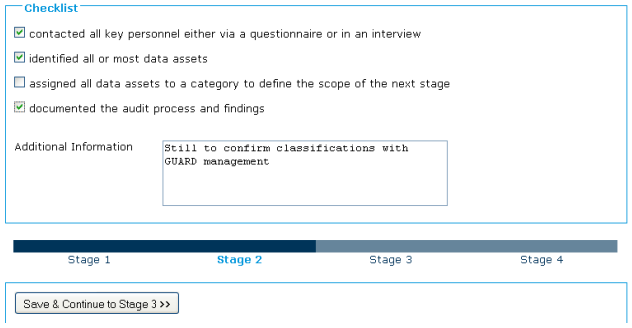
JISC | University of Glasgow Humanities Advanced Technology & Information Institute | DCC

Data Audit Framework

Online tool: Stage 2







JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

Data Audit Framework

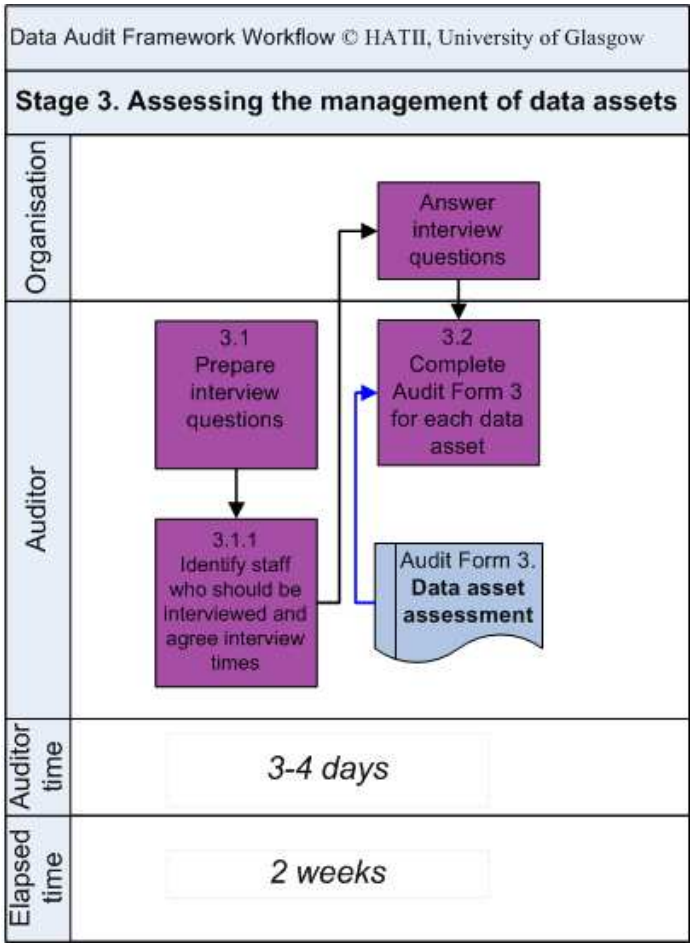
Stage 3: assessing

Objective

- Identify weaknesses in data management and potential risks

Process

- Assess most crucial assets to organisation
- Consider access, reuse, and data management



JISC | University of Glasgow | Humanities Advanced Technology & Information Institute | DCC

Data Audit Framework

Audit form 3A: core

Audit Form 3A: 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	**Data creator(s)**	*Person, group or organisation responsible for the intellectual content of the data asset*
3	**Title**	*Official name of the data asset, with additional or alternative titles or acronyms if they exist*
4	**Description**	*A description of the information contained the data asset and its spatial, temporal or subject coverage*
5	**Subject**	*Information and keywords describing the subject matter of the data*
6	**Creation date**	*The date(s) on which the data was collected or created*
7	**Purpose**	*Reason why the asset was created, intended user communities or source of funding / original project title*
8	**Source**	*The source(s) of the information found in the data asset*
9	**Updating frequency**	*The frequency of updates to this dataset to indicate currency*
10	**Type**	*Description of the technical type of the data asset (e.g., database, photo collection, text corpus, etc.)*
11	**Format**	*Physical formats of data asset, including file format information*
12	**Rights and restrictions**	*Basic indication of user's rights to view, copy, redistribute or republish all or part of the information held in the data asset. Access restrictions on the data itself or any metadata recording its existence should also be noted*
13	**Usage frequency**	*Estimated frequency of use and if known required speed of retrieval to determine IT infrastructure and storage needs*
14	**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*
15	**Archiving policy**	*Number of copies of the data asset that are currently stored, frequency of back-up and archiving procedures*
16	**Management to date**	*History of maintenance and integrity of the data asset e.g. edit rights / security, and any curation activities performed*

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

Data Audit Framework

Audit form 3B: extended

- **Description** title, type, subject, abstract, keywords, language…
- **Provenance** purpose, source, context, dates, modifications….
- **Ownership** creators, curators, rights, usage constraints…
- **Location** references, versions, relations, potential archive...
- **Retention** value, disaster recovery, archive / preservation policy…
- **Management** documentation, audit trail, costs, funding…

Online tool: Stage 3

Assets

	Asset Name	Classification	Date Created (Creator)	AF3 Updated (Updater)	AF3 Form Type
View	M74 assessment	vital	2008-10-19 (sjones)		core
View	Rannoch Moor - HLF	important	2008-10-19 (sjones)		core
View	Water Row Govan evaluation	minor	2008-10-19 (sjones)		
View	Kirk of St Nicholas Skeletal analysis	vital	2008-10-19 (sjones)		core
View	Dun Eistean Archeology Project (DEAP)	important	2008-10-19 (sjones)		

Element set

Element set: Core (Core / Extended) — Please choose an element set

Save

http://www.data-audit.eu/tool/stage3/af3.php

DAF Online Tool: Audit Stage 3 | DAF Online Tool: Audit Form 3: Da...

* Data Creator 1	Alan Leslie (project management)	
* Data Creator 2	Alastair Becket (fieldwork and auth	
* Data Creator 3	John Arthur (trench survey and illus	Add a Creator
* Data Title 1	Birkhill, Coalburn evaluation	Add a Title
* Description	c.590 photographs with photo list; and several concordances / trench tables pertaining to an archaeological evaluation of a c 30 ha site proposed for	
* Subject	evaluation; archaeological remain	Keywords separated by semicolon
* Creation Date 1	2008-04-14 — Start of evaluation which ran for 1 week	
* Creation Date 2	2008-05-20 — Date of final report	Add a Date
* Purpose	Commissioned by the Coakley Group in response to a condition within the planning permission granted by South Lanarkshire Council.	
* Source	fieldwork; historic maps; previous work e.g. desk-based assessment;	
* Updating Frequency	daily during fieldwork then closed to edits / updates after write-up	
* Data Type 1	photographs	Add a Type
* Data Format 1	doc	
* Data Format 2	jpg	Add a Format
* Rights & Restrictions	Accessed internally during creation and write-up. Client could stipulate access restrictions but liklely to be made publicly available	
* Usage Frequency	Daily during creation and write-up then sporadic. This project may be reopened and conclusions revisted as local council not convinced there is	
* Relation	Desk-based assessment conducted	

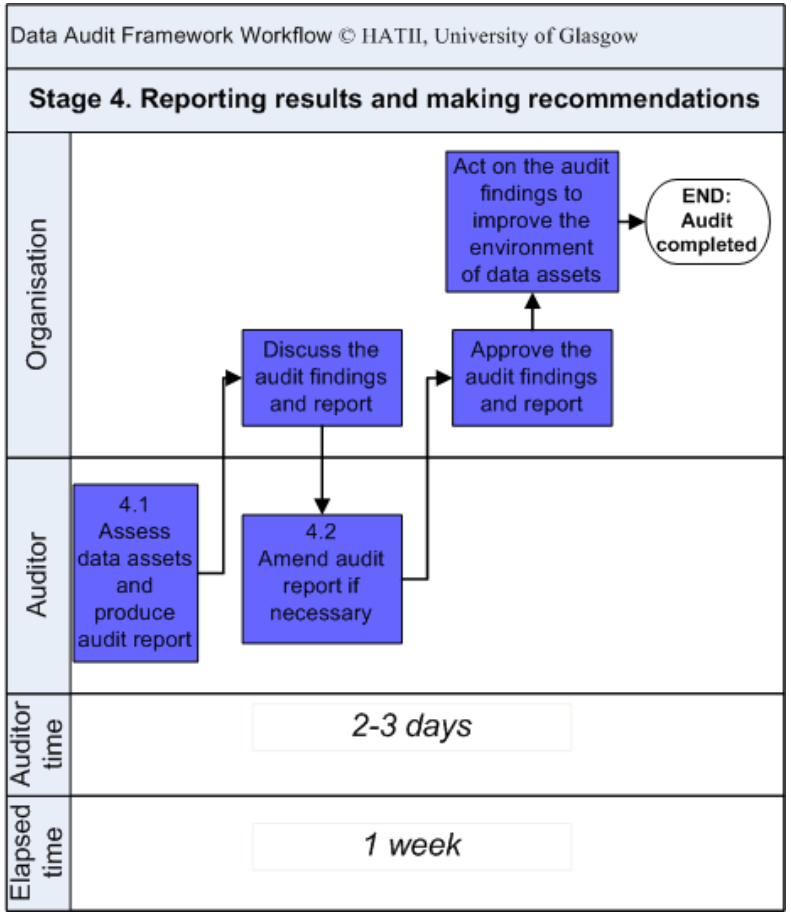
Stage 4: recommendations

Objective

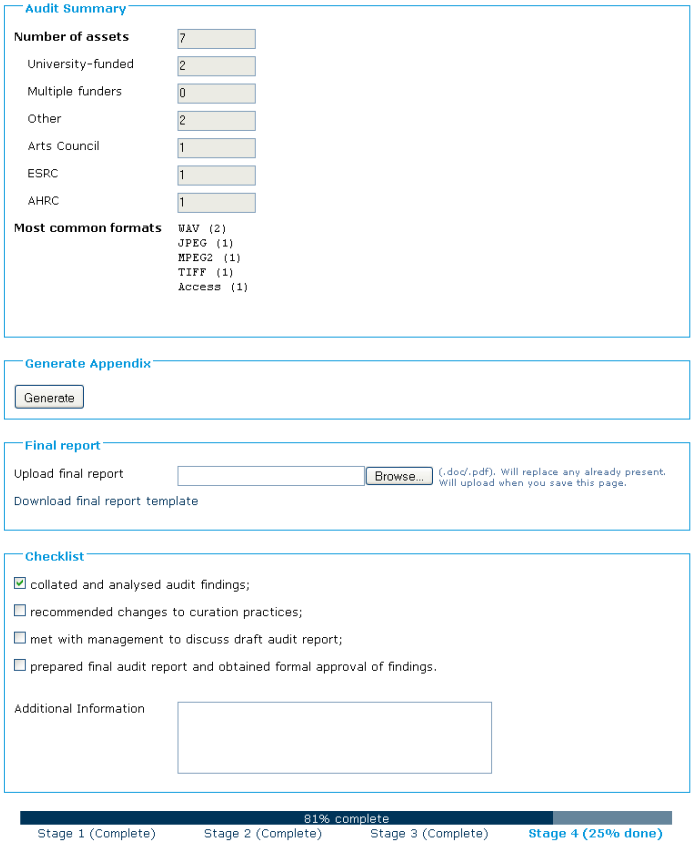
- Recommend changes to improve data management

Process

- Collate audit results
- Analyse data



Online tool: stage 4



Audit Summary

Number of assets	7
University-funded	2
Multiple funders	0
Other	2
Arts Council	1
ESRC	1
AHRC	1
Most common formats	WAV (2) JPEG (1) MPEG2 (1) TIFF (1) Access (1)

Generate Appendix

Generate

Final report

Upload final report — Browse... (.doc/.pdf). Will replace any already present. Will upload when you save this page.

Download final report template

Checklist

- [x] collated and analysed audit findings;
- [ ] recommended changes to curation practices;
- [ ] met with management to discuss draft audit report;
- [ ] prepared final audit report and obtained formal approval of findings.

Additional Information

81% complete

Stage 1 (Complete) | Stage 2 (Complete) | Stage 3 (Complete) | **Stage 4 (25% done)**

Department of Theatre Film and TV Studies (TFTS): Audit Form 1

Organisation details

Address

Theatre Film and Television Studies Department,
University of Glasgow,
The Gilmorehill Centre,
University Avenue,
Glasgow G12 8QQ

Description

Theatre Film and Television Studies (TFTS) at the University of Glasgow is one of the most well-established departments for these subjects in the UK. We have a strong team of over twenty teaching and research staff who specialise in diverse areas of performance and media arts. Several taught courses are offered. Through these students build up critical knowledge, creative approaches and confidence in their own skills. An excellent subject library situated on-site provides access to web resources, course literature, a large video collection and theatre production materials. The Department places a strong emphasis on research and benefits greatly from financial support offered by previously alumni, most notably Dame Margaret McArthur. Several widely renowned research projects stem from McArthur Fellowships as well as from other research council funding secured by TFTS. These include Changing Approaches to Theatrical Tuition, Glasgow Cinemas and Shakespearean performance database.

Strategies

Methods of working vary between staff members. No policy on creating and preserving research data is in place at present. Best practice guides from the Performance Data Archive (PDA) are used on research council projects and some staff have drawn these processes into their everyday work.

Responsibilities

Defined largely by individual researchers. Librarian maintains some research outputs and others deposited with data archives.

Budget

Research proposals sometimes allocated funds to preservation. Internally it's part of library

Data Audit Framework

Example recommendations

Devise and implement a data policy

- Use a consistent approach to file naming and directory structures
- Store data appropriately – not on external HDDs, USB sticks, CDs…
- Maintain data integrity (checksums, access restrictions, edit rights)

Develop practical guidance and training (from PhDs / ECR)

Budget for curation and reuse from outset

Work with local services such as IR or archive

Data Audit Framework

Pilot audits

- School of GeoSciences at University of Edinburgh
- Innovative Design and Manufacturing Research Centre (IdMRC) at University of Bath
- Glasgow University Archaeological Research Division (GUARD) at University of Glasgow

Data Audit Framework

Main findings

- Lots of data being created
- Few policies for data creation, storage and management
- Researchers unsure where to begin
- Unaware of available support
- Often no place of deposit or funds for preservation

International Journal of Digital Curation

http://www.ijdc.net/ijdc/article/view/91/109

JISC | University of Glasgow | Humanities Advanced Technology & Information Institute | DCC

Data Audit Framework

Implementation tips

- Scope audit carefully
- Timing needs to be appropriate
- Find an advocate in the department
- Collect information at once where possible
- General discussion to build rapport, communicate purpose of audit and understand organisation

Lessons learned: http://www.data-audit.eu/docs/DAF_lessons_learned.pdf

Data Audit Framework

Conclusions

Premise of DAF project is founded

- Lack of knowledge of what data is held
- Poor understanding of how to curate it

Key areas for future work

- Support on policies, standards and best practice guidance
- Skills lacking – basic training & professional workforce needed
- Develop robust, sustainable UK infrastructure for curation
- Recognition of data and active push for sharing and reuse critical

Data Audit Framework

Next steps for DAF

- Support and training
- Implementation projects → redevelopment
- Results feed into other data research
  - UKRDS, Data Skills study
  - DCC policy study
  - Glasgow and Edinburgh University curation strategy

Data Audit Framework

Further information

- Jones, Sarah, Ross, Seamus & Ruusalepp, Raivo, *The Data Audit Framework: a toolkit to improve data management in research led institutions* http://www.bl.uk/ipres2008/ipres2008-proceedings.pdf
- Jones, Sarah, Ball, Alex & Ekmekcioglu, Çuna, *The Data Audit Framework: a first step in the data management challenge* http://www.ijdc.net/ijdc/article/view/91/109

http://www.data-audit.eu

Sarah Jones: s.jones@hatii.arts.gla.ac.uk

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

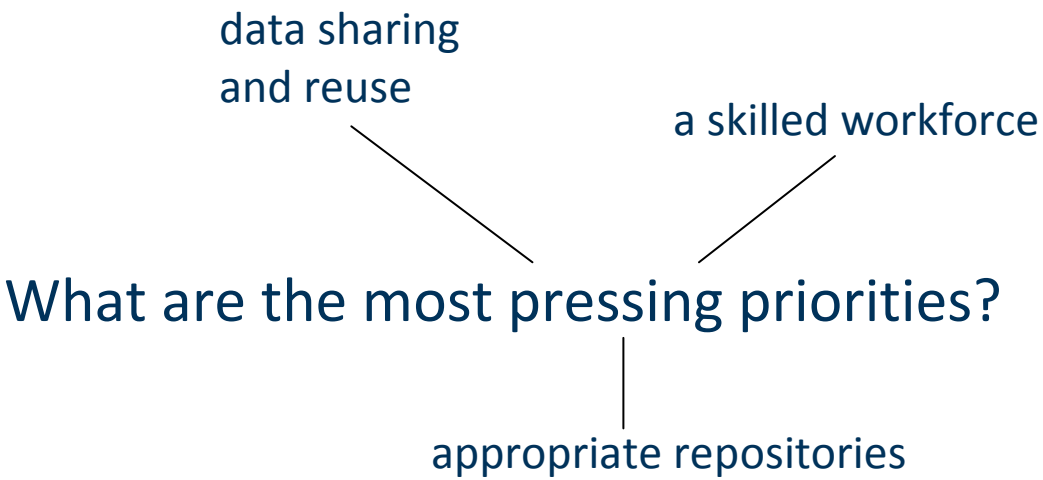
DCC

Data Audit Framework

Discussion…

Why curate?

What role should researchers assume in the curation lifecycle?

How can we overcome ownership and control issues to make data open?

What is required of co-ordinating bodies such as the DCC?