Data Audit Framework

The Audit Methodology

Sarah Jones

HATII & DCC, University of Glasgow

DAF Workshop, NeSC, Edinburgh

1st December 2008

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

JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

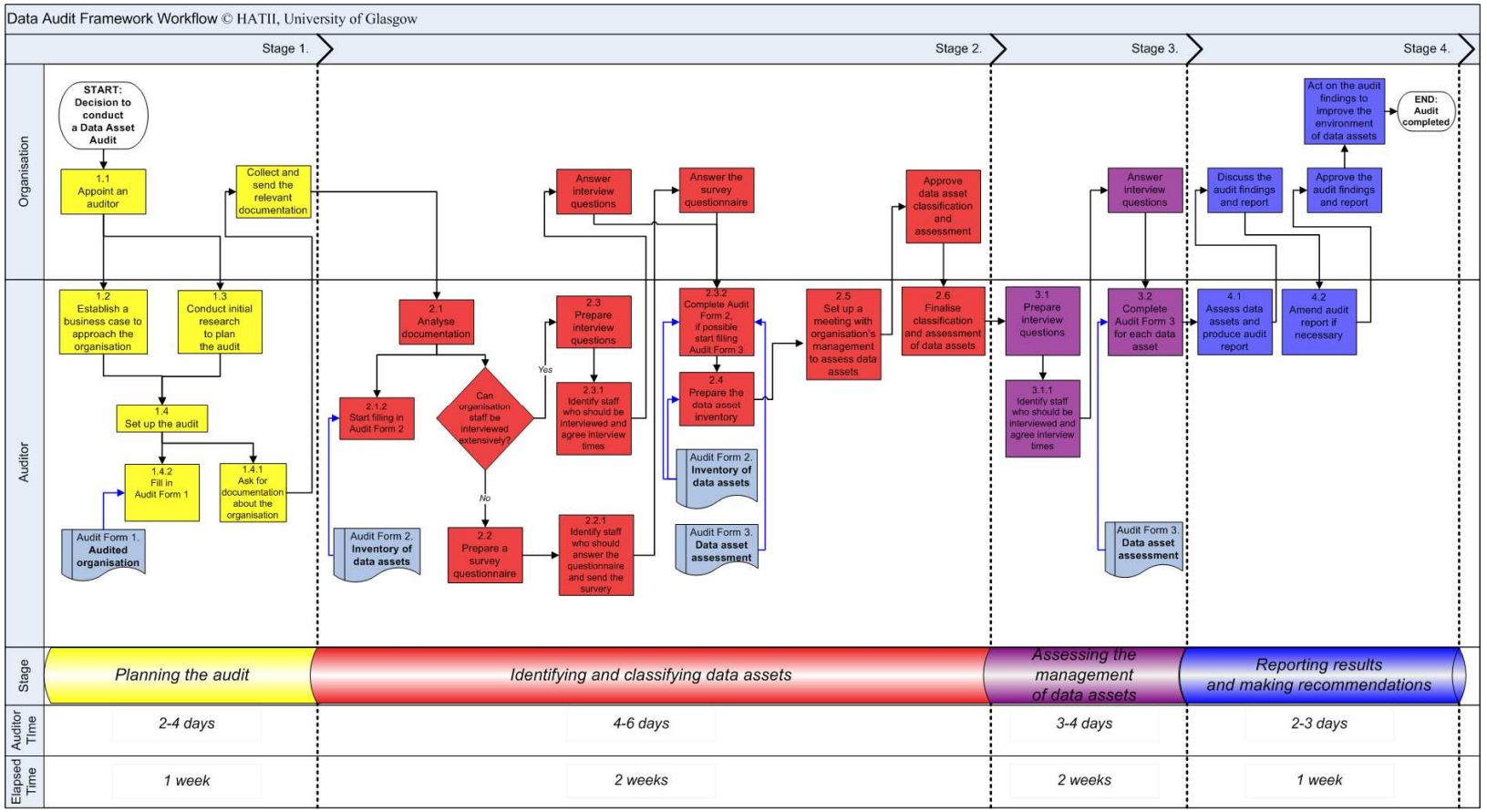
Data Audit Framework

The methodology

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

Data Audit Framework

Workflow and tasks



JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

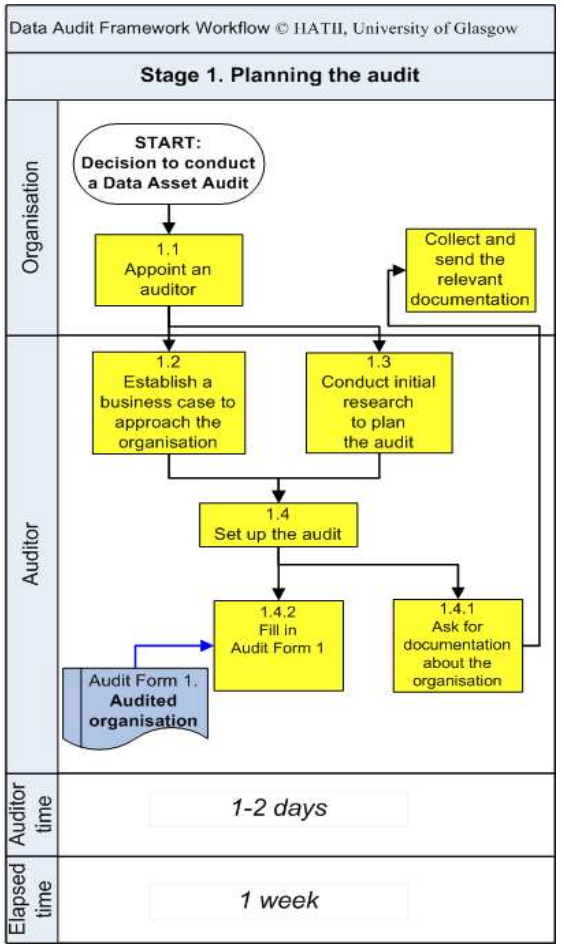
DCC

Data Audit Framework

Stage 1: planning

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

Prepare in advance to optimise time with staff



JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

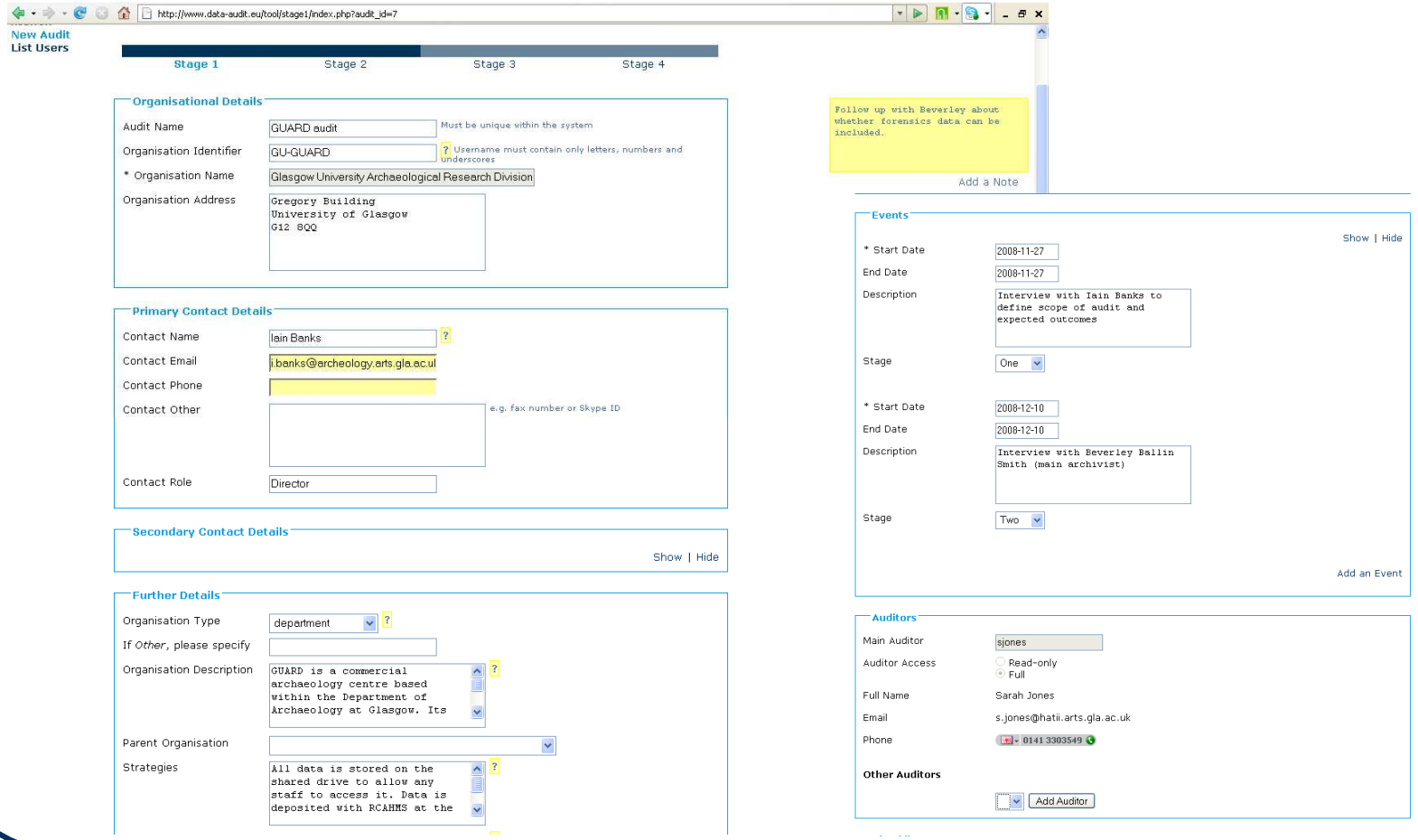
DCC

Audit Form 1

Audit Form 1: Audited organisation	
Organisational details	
Contact details	Name, address, email, phone contact details for organisation being audited / main person
Organisation type	University, Department, Institute, Archive, Data Centre, Computer Service, Research Team etc
Organisation description…..	A short abstract explaining the organisation's remit, main work and areas of activity
Data management details	
Strategies	Details of policies, operational manuals or regulations used, with references.
Responsibilities	Details of staff who have clear responsibility for creating and managing the data assets
Budget	Level of resources available and whether it is sufficient to manage data at the present time
Audit details	
Dates of the audit	Period audit completed in, including any significant dates e.g. main interviews.
Name of the auditor(s)	Name, contact details…
Date of completing form….	Dates.

Data Audit Framework

Online tool: stage 1

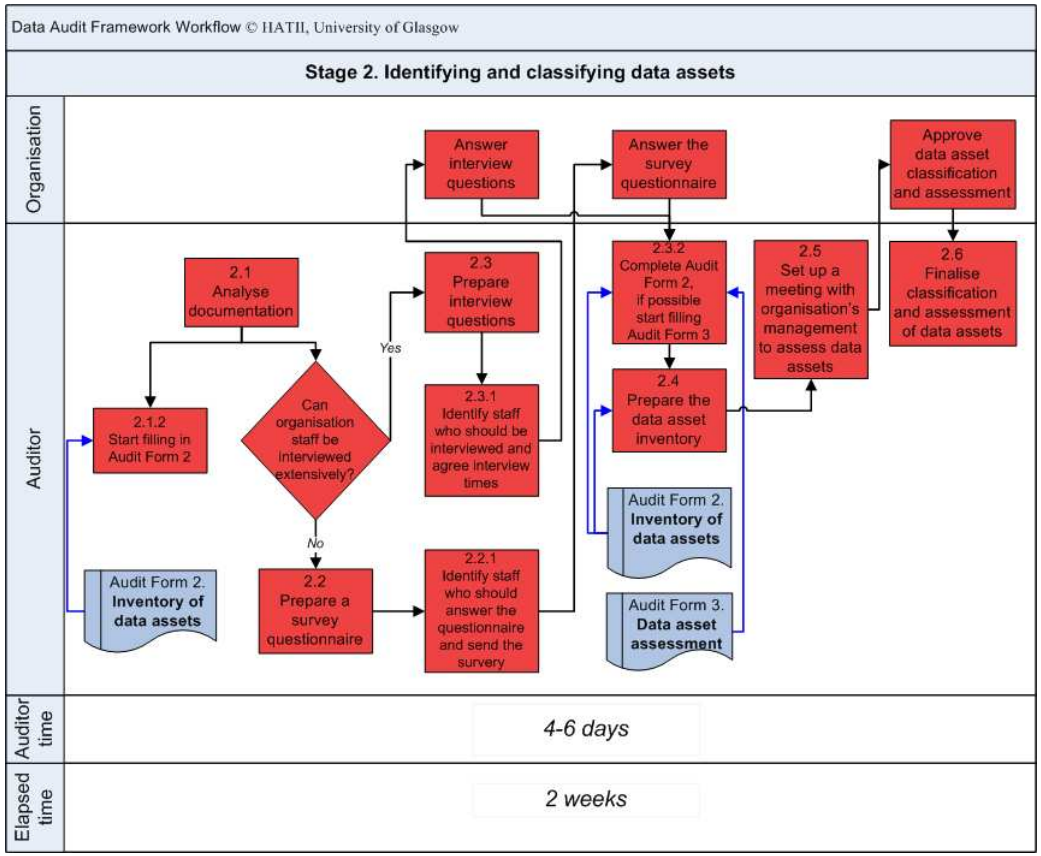


Data Audit Framework

Stage 2: identifying

- Identify and classify assets
- Interviews, questionnaires, research

Create an inventory



JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

DCC

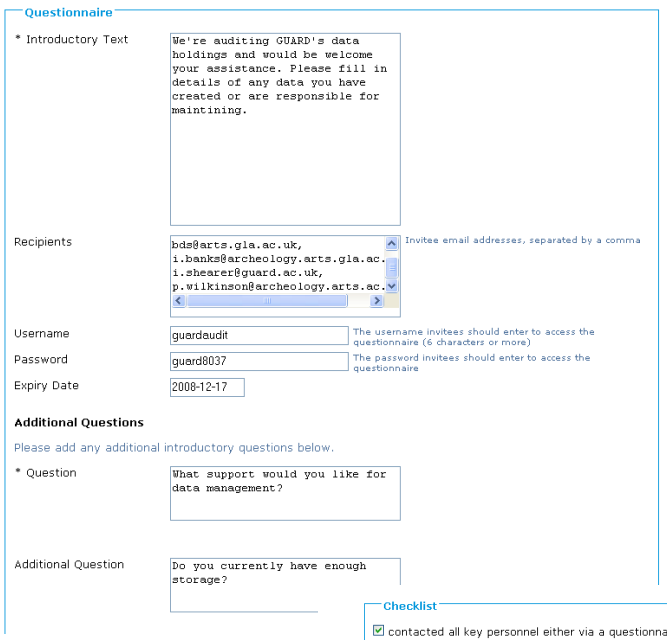
Data Audit Framework

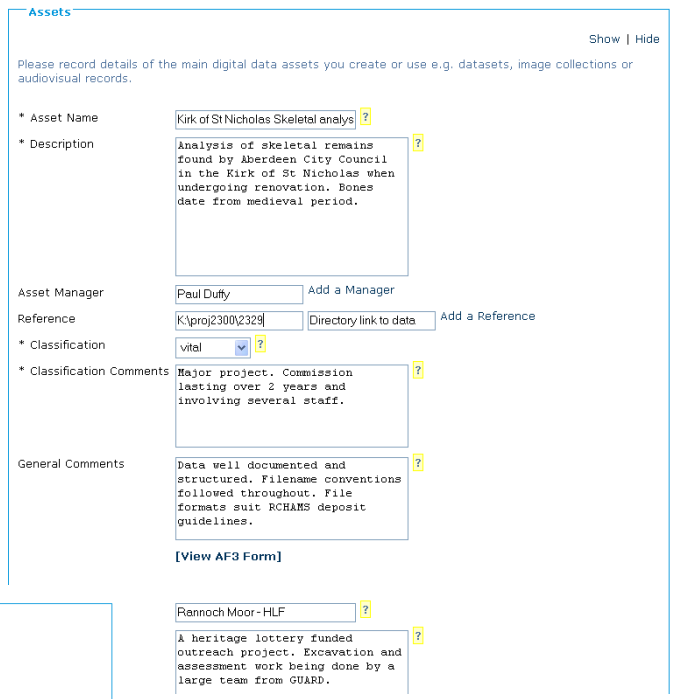
Audit form 2

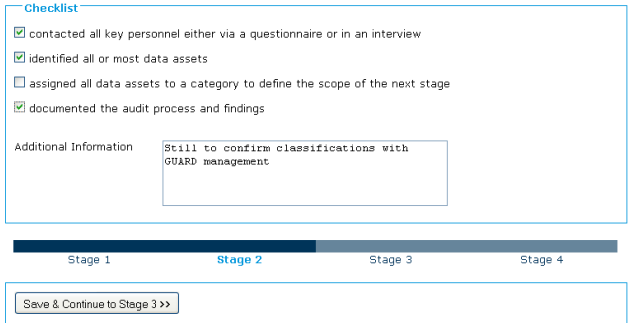
Audit Form 2: Inventory of data assets						
Name of the data asset	**Description of the asset**	**Asset Manager(s)**	**Reference**	**Classification**	**Classification comments**	**General comments**
Official name	*Basic description of content*	*Name and position or role*	*Where reference to asset was found / its location*	*Choice of:* *Vital* *Important* *Minor* *(unless different schema chosen)*	*Reason for classification or Comment.* *Could also include action suggested for each asset*	*Could include creation date, original purpose, type of data asset, or format.*

Data Audit Framework

Online tool: Stage 2







JISC

University of Glasgow | Humanities Advanced Technology & Information Institute

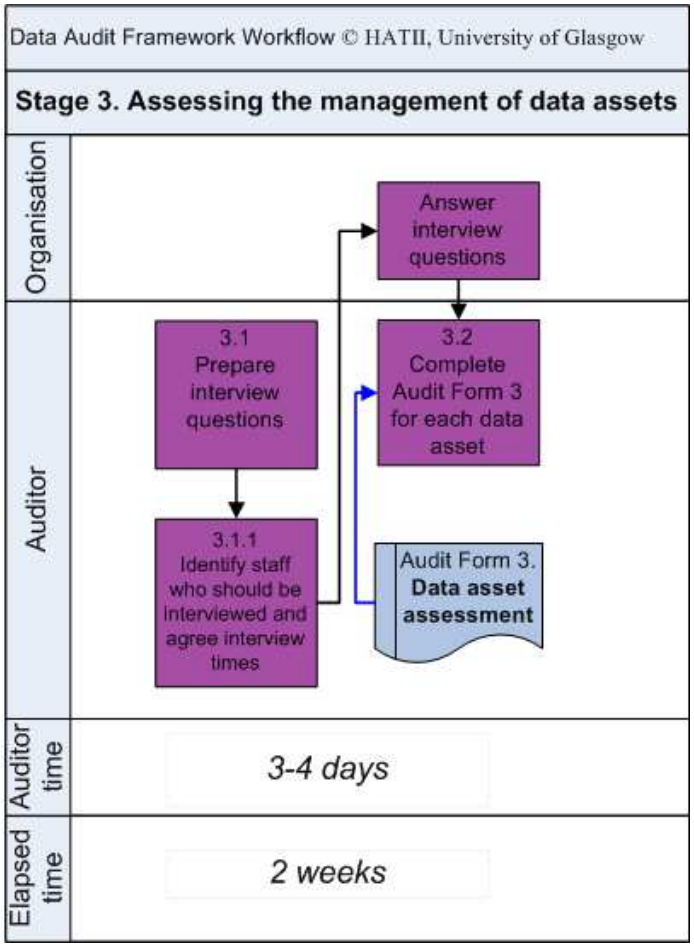
DCC

Data Audit Framework

Stage 3: assessing

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

Identifying weaknesses and risks



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, deposit in 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

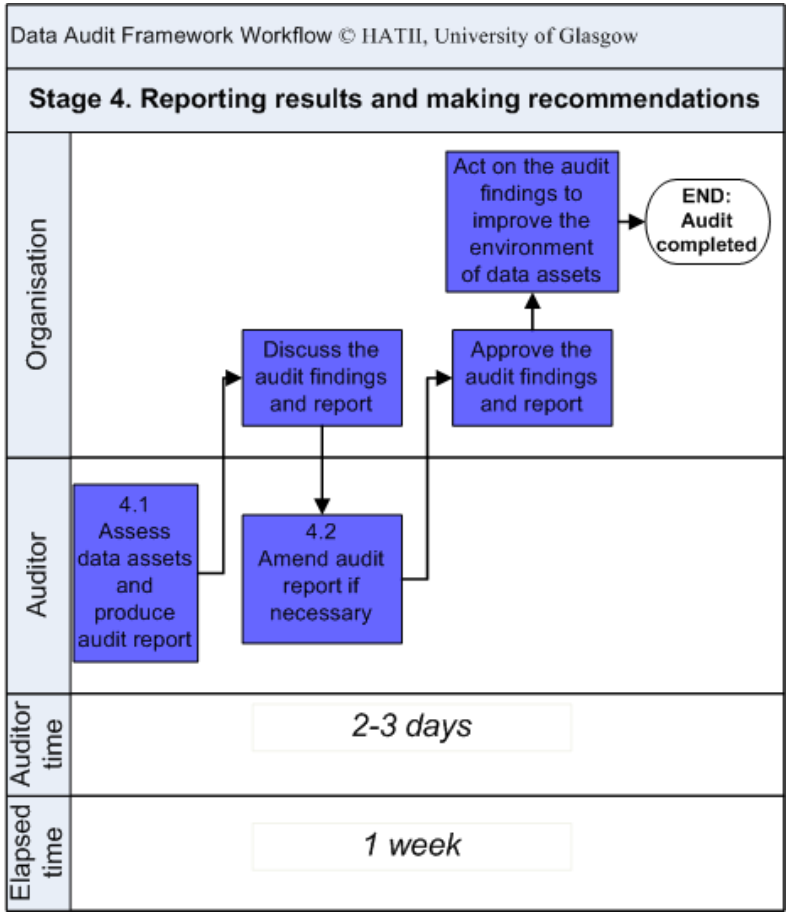
DAF Online Tool: Audit Stage 3 | DAF Online Tool: Audit Form 3: Da...

* 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	

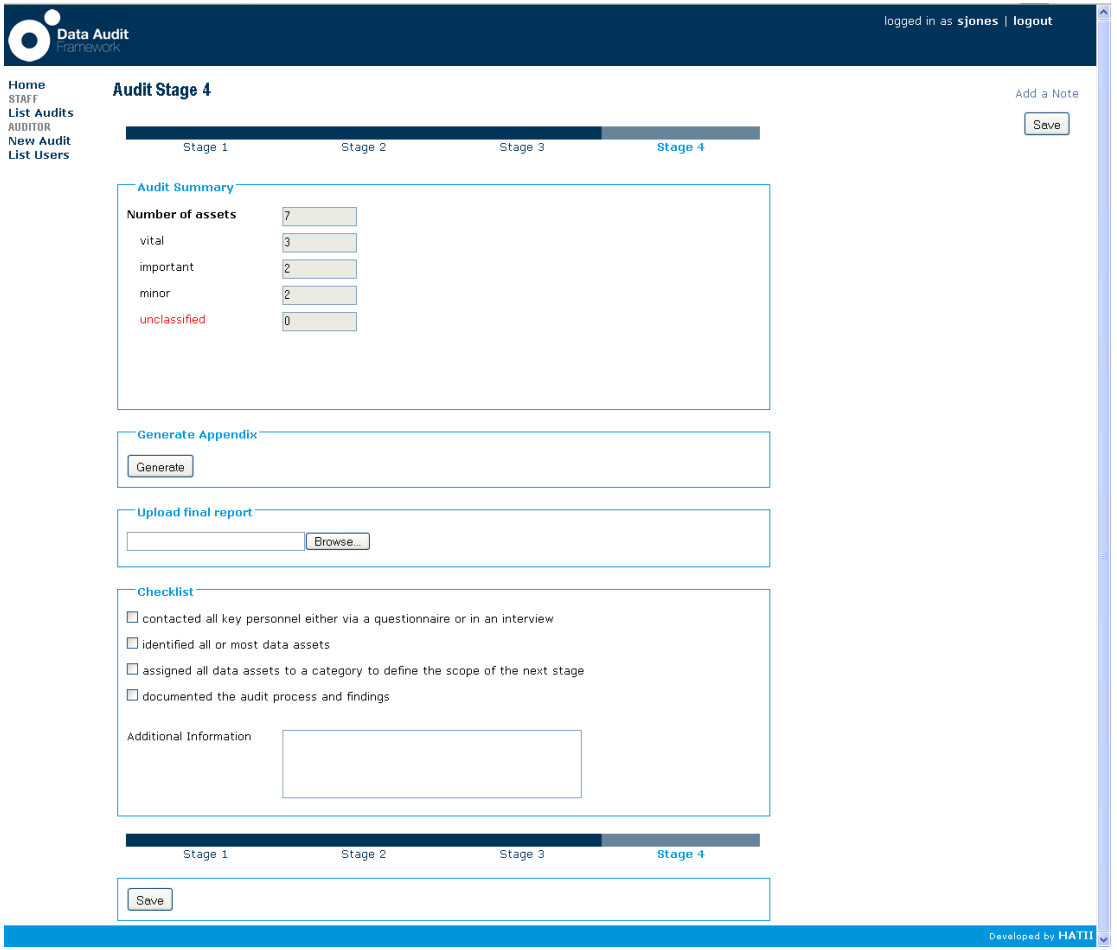
Stage 4: recommendations

- Collate audit results
- Analyse data

Recommend changes to improve data management



Online tool: audit stages



Data Audit Framework

Broad recommendations

- Develop policies
  - Content creation i.e. file naming conventions, directory structure guidelines, recommended formats
  - Curation i.e. storage and backup, access controls, edit rights, processes for checking integrity, deposit with archive
- Guidance
  - Practical guides not just policies
  - Training appropriate to staff needs – begin with PhD students
- Planning
  - Budget for curation and reuse from outset
  - Work with local services such as IR or archive

Data Audit Framework

Useful links

- DAF methodology
  http://www.data-audit.eu/DAF_Methodology.pdf
- Lessons learned reports
  http://www.data-audit.eu/docs/DAF_lessons_learned.pdf
- Sarah Jones s.jones@hatii.arts.gla.ac.uk

Thank you! Any questions?

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