Data Audit Framework

Sarah Jones

HATII, University of Glasgow

Data Audit Framework

SOME RIGHTS RESERVED

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5 UK: Scotland License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/2.5/scotland/ ; or, (b) send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

Funded by:

JISC e-Science

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)

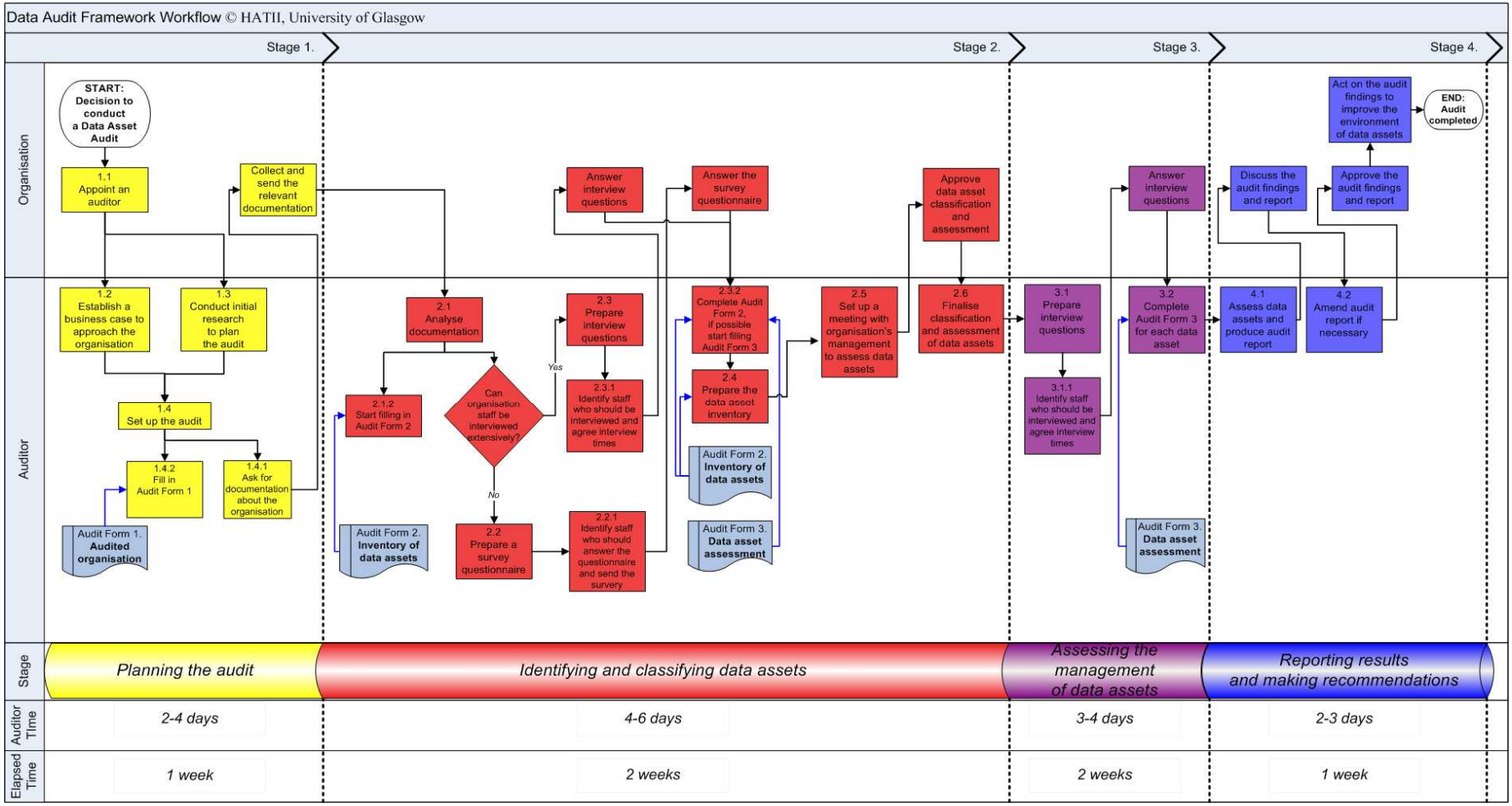
Data Audit Framework

The methodology

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

Data Audit Framework

Workflow and tasks



The inventory

- Name of the data asset
- Description
- Asset manager (responsibility)
- Reference (location)
- Classification
- Comments (general and on classification)

Data Audit Framework

Classification questions

- Is this data central to your research?
- Will the data be useful in the future?
- Are you the intellectual owner?
- Is it documented and in a sustainable form?
- Is it already being preserved elsewhere?

Data Audit Framework

An example

Name of Asset	Description	Asset Manager	Reference	Classification	Comments
The Gazetteer for Scotland	Collection of text, images, geo maps, Featuring details of towns, villages & glens from the Scottish Borders to the Northern Isles.	Bruce Gittings in collaboration with David Munro	http://www.geo.ed.ac.uk/scotgaz/gaztitle.html	Vital	Funded by Royal Geographical Society Data is stored in the School of Geo Sciences' Oracle database and updated regularly.
Habitable cars	Collection of videos (QuickTime movie clips) and transcripts.	Senior Research Fellow	Mentioned in interview on Thursday May 8th.	Vital	The project is funded by ESRC. Videos are catalogued with keywords and stored on external HDD.
Labour outsourcing in the European ferry sector	Interview data (audio files and transcripts)	Lecturer	Mentioned in interview on Monday 26 May 2008.	Important	NUFFIELD funded. Data is kept on the school server as well as on USB storage and PC hard disk. Confidentiality issues

Data Audit Framework

Asset management

- ID
- Data creator(s)
- Title
- Description
- Subject
- Date
- Purpose
- Source
- Updating frequency
- Type
- Format
- Rights and restrictions
- Usage frequency
- Relation
- Back-up and archiving policy
- Management to date

Data Audit Framework

An example

No	Parameter	Comment
1	ID	GS_01
2	Data creator(s)	Bruce Gittings in collaboration with David Munro (The Royal Scottish Geographical Society)
3	Title	The Gazetteer for Scotland
4	Description	Geographical encyclopedia featuring details of towns, villages, bens and glens from the Scottish Borders to the Northern Isles.
5	Subject	Geography, history, Scottish Borders, Northern Isles
6	Date	1995
7	Purpose	Teaching tool, also open to public use
8	Source	http://www.geo.ed.ac.uk/scotgaz/Sources.html
9	Updating Frequency	Weekly
10	Type	Online database
11	Format	Geo maps, images, sound, video, text files
12	Rights and restrictions	Unless explicitly stated otherwise, the content of these web pages is the copyright of the editors of the Gazetteer for Scotland.
13	Usage frequency	Daily
14	Relation	None
15	Backup and archiving policy	Data is stored in the School of GeoSciences' Oracle database and updated regularly by Bruce Gittings
16	Management to date	Backups done as stated above. Content controlled by Bruce Gittings.

Practical exercise: scenario

Imagine you are an early career researcher in Glasgow University's Economics Department and have been awarded a one year ESRC grant to investigate social and economic factors affecting a nation's Olympic success. You are still in the early stages of your project so have only collated a small amount of data.

The collections advisor of the UK Data Archive has recently been in touch to offer guidance on best practice in data creation, storage and management to ensure deposit with them at the end of the grant runs smoothly. This brought to light the importance of maintaining documentation during the project so you have decided to assess data already collated to see where information is lacking and identify data for deposit.

Data Audit Framework

The exercise

Using the Data Audit Framework excel spreadsheets and the data provided in the Olympics zip file, complete the following steps:

1. Record each data asset in the inventory attributing an identifiable name, description, asset manager and reference.
2. Classify each asset as vital, important or minor, based on whether it needs to be maintained to enable completion of your research or whether you are required to deposit it. A note should be made to explain your classification and suggest actions to follow e.g. deposit with UKDA / dispose of.
3. Consider vital assets in more detail to see whether you have all the necessary information to hand to ensure their long-term preservation. If information is lacking, make a note of what kind of documentation should be maintained during the remainder of your research project.

Data Audit Framework