



D | C | C

... because good research needs good data

Using DAF and AIDA to scope data management needs

Sarah Jones
DCC, University of Glasgow
s.jones@hatii.arts.gla.ac.uk

Ed Pinsent
ULCC
e.pinsent@ulcc.ac.uk



Funded by:



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.

DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

We'll provide a quick overview to each tool then do an exercise on how to go about scoping requirements.



DC|C

... because good research needs good data

Background to DAF project

“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)



DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

DAF established in response to a recommendation in the Dealing with Data report. This recognised a lack of awareness as to what data is held within HE institutions and how it's being managed.



DC|C

... because good research needs good data

The methodology



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

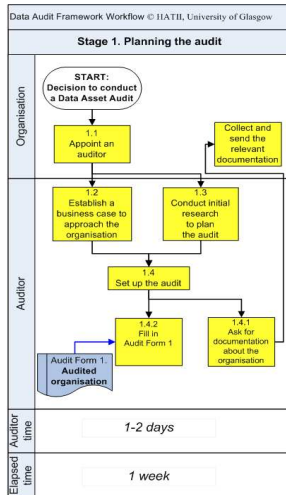


DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

The methodology has four incremental stages, one for planning, one for wrap up and two main audit stages in between that pick up directly on the two aspects in the original recommendation i.e. what's there (inventory stage) and what's happening to that data (assessment stage).



Stage 1: planning



Objective

Determine what you want to find out and prepare work in advance

Process

- Define scope / expected outcomes
- Research organisational context
- Set up survey, interviews, meetings...

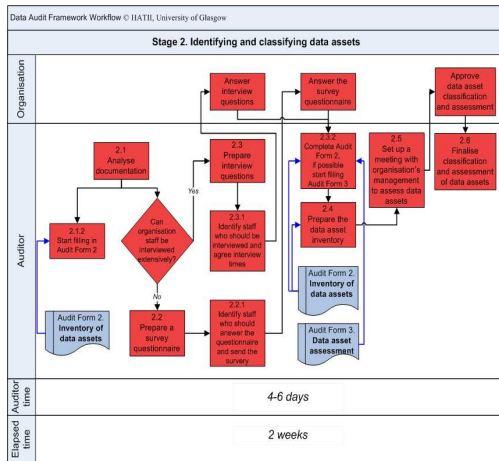


Stage 1 is about scoping out what you intend to do and planning for this. We found it key to restrict scope, but this is quite a hard thing to achieve.

Important to have as clear a focus as possible, to tailor the audit to this and to use it to define benefits.



Stage 2: identifying data



Objective

Create inventory to understand scale of data

Process

Engage researchers to:

- Identify key data assets
- Classify data to restrict scope

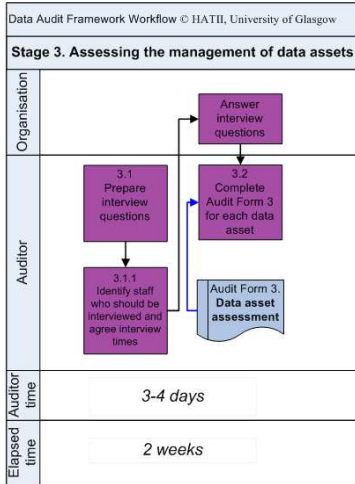


Stage 2 is an initial scoping exercise to get a general handle on the data being created / held. Collect basic details (name / description of data, who's responsible, location) and classify to determine which data you'll look at in more detail later to assess data management.

Classification depends on your goals e.g. a capacity planning exercise for the IR may use restrict the scope to data for which the institution is responsible i.e. that without a designated data centre or other owner.



Stage 3: assessing data management



Objective

Identify weaknesses in data management and potential risks

Process

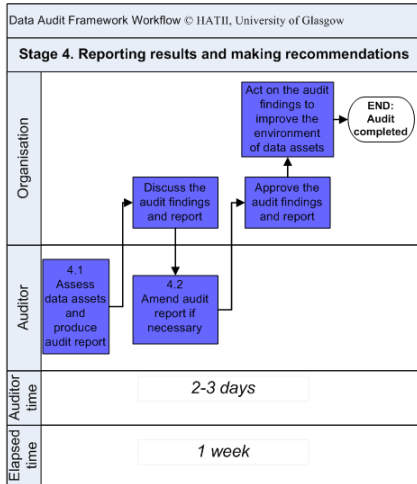
- In-depth assessment of most crucial assets, given purpose of audit
- Discussion on lifecycle of data to assess data management



The data identified in Stage 2 can be a useful vehicle to open discussion into how data are managed throughout their lifecycle to identify weak points or risks. This stage is generally completed by interview as this has been found the best way to engage with researchers to understand working practices / data needs etc.



Stage 4: recommendations



Objective

Recommend changes to improve data management

Process

- Collate audit results
- Analyse data
- Suggest changes to mitigate weaknesses



Stage 4 is to pull together the information collected and provide suggestions for changes that would improve data management.

The way the methodology is presented is linear, however it can be used flexibly. Depending on your aims you may want to focus more on the data management discussions than the inventory or approach them in a different order. We're happy for people to change the approach in DAF or amend the information collected to suit their purpose.



D | C | C

... because good research needs good data

Background to AIDA project

- A tool, based on a proven model, to enable institutions to perform self-assessment of their readiness for digital asset management and digital preservation
- Based primarily on the successful Cornell model for describing institutional readiness and capabilities for digital preservation
- Five Stages and Three Legs



DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

AIDA is also a JISC project, funded under the 1/07 JISC Capital Programme Call, to develop and test digital asset management tools in a variety of institutional settings.

It stands for Assessing Institutional Digital Assets - JISC funded the project and they wanted the results to be useful for UK Higher Education. Like everyone else, Universities have lots of digital assets, but what are they doing to manage them, or even preserve them? How could they find out how well (or how badly) they are doing? AIDA is a measurement tool that will give them some indicators of their preparedness to manage digital assets.

It's based to a large extent on the Cornell models of the five stages and the three legs, which is something we teach at the DPTP.



DC|C

... because good research needs good data

AIDA methodology

- Assess Organisation, Technology and Resources
- Five stages of preparedness: Acknowledge, Act, Consolidate, Institutionalise, Externalise
- Two-tiered approach: assess entire Institution and at Department/collection level



DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

The Toolkit is a comprehensive self-assessment exercise which looks at the 3 Legs in turn, and probes for information about what's going on in the Organisational, Technological, and Resource legs. Within each leg, there are 11 elements (Resources only has 9). To assess yourself, for each element, you would look at one of five brief descriptive scenarios, along with some exemplars and indicators. These match the five stages as closely as possible. Notice also we do it at two levels; one for the entire Institution, and at a second lower level for a Department, or a collection, or a project.



Assessment 1: Organisation Leg

Organisation Element 01: Mission statement	Level of implementation				
	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
INSTITUTION: Indicators / exemplars	Institution relies on departments to identify and manage their own digital assets. <i>"If a mission statement is confirmed not to exist, then it is something we would press for."</i>	Digital asset management is on the agenda. Issues are discussed at senior management level, recognizing issues that will arise as a result of changing practice. Staff, etc. being consulted and options explored. <i>"We have a variety of policies relating to aspects of digital asset management but no unified policy or mission statement on digital asset management." "There is a recognition of the importance of informational assets but a lack of a coordinated response to managing them."</i>	A mission statement is in development. Mission statement exists, but is not widely communicated within the institution. <i>"There are various policies (e.g. records management, intellectual property, information strategy, institutional repository) but they are not always disseminated or enforced."</i>	Mission statement is written and fully reflects institutional commitment to ownership of digital assets, their management and their creation. Statement is communicated internally.	Mission statement is published, accessible to others, and externally recognised. Mission statement is used and copied as an exemplar of good practice nationally or internationally.
DEPARTMENT / COLLECTION: Indicators / exemplars	Departments identify and manage their own digital assets. Low awareness amongst staff and misunderstandings are common.	<i>"Departmental and individual freedom is high, especially in academic departments, so things are dependent on local practice, individual effort and research funding requirements."</i>	<i>"There is an awareness of the value of certain sets of data (e.g. student and staff records) which have strict policies dealing with creation, access and retention."</i>	Mission statement is written and fully reflects a departmental commitment to ownership of digital assets, their management and their creation.	

The elements within each leg were derived from a combination of

- 1: *An Audit Checklist for the Certification of Trusted Digital Repositories*. This is the document originally produced by RLG-NARA, which later became *Trustworthy Repositories Audit and Certification (TRAC): Criteria and Checklist*
- 2: *The Cornell University Survey of Institutional Readiness*.
- 3: Summary of RLG-OCLC Framework Component Characteristics.
- 4: Network of Expertise in long-term STORAGE (NESTOR).



ORGANISATION LEG

Organisation Element 08: Metadata management	Level of implementation				
	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
INSTITUTION: Indicators / exemplars	<p>Metadata creation is not managed.</p> <p>No guidance to staff for metadata creation. No policy for external assets' metadata.</p> <p><i>"There is no overall guidance for staff on how to create metadata...no overall strategy and awareness is low."</i></p>	<p>Academic staff are empowered to create basic metadata.</p> <p>Metadata schemas are being invented ad-hoc by individuals or lone departments.</p>	<p>Cataloguing and archival skills available to support programmes for metadata creation.</p> <p>Metadata conforms to standards of interoperability.</p>	<p>Metadata is embedded in institutional workflow, using appropriate standards.</p> <p>Internal and externally-acquired assets are governed by appropriate guidance.</p>	<p>Metadata meets externally recognised standards and is integrated across systems.</p> <p>Metadata management is capable of flexibility and expanding to meet the requirements of all stakeholders and partnerships.</p>
DEPARTMENT / COLLECTION: Indicators / exemplars	<p>Metadata creation is not managed.</p> <p><i>"Some systems use metadata (for example student numbers are the metadata linked to all files relating to that student so that the system can pull them all together)."</i></p>	<p>Metadata schemas are being invented ad-hoc by individuals or lone departments.</p>	<p>Cataloguing and archival skills available to support metadata creation at a collection level.</p>	<p>Metadata is embedded in departmental workflow, using appropriate standards.</p>	<p>Metadata meets externally recognised standards and is integrated with the rest of the Institution.</p>
Possible supporting sources	<p>Written procedures and guidance for all aspects of the management of metadata.</p> <p>Written procedures for managing the metadata of externally-acquired assets.</p>				



DC|C

... because good research needs good data

AIDA Assessment: Organisation leg

Quick summary:

- 1 = ACKNOWLEDGE
- 2 = ACT
- 3 = CONSOLIDATE
- 4 = INSTITUTIONALISE
- 5 = EXTERNALISE

Please refer to the toolkit for further detail and interpretation.

01: Institutional Mission Statement

Assess institution-wide mission statements

1 2 3 4 5

01: Departmental Mission Statement

Assess mission statements which apply to asset collections or department

1 2 3 4 5

Notes on Element 01: Mission statement

DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

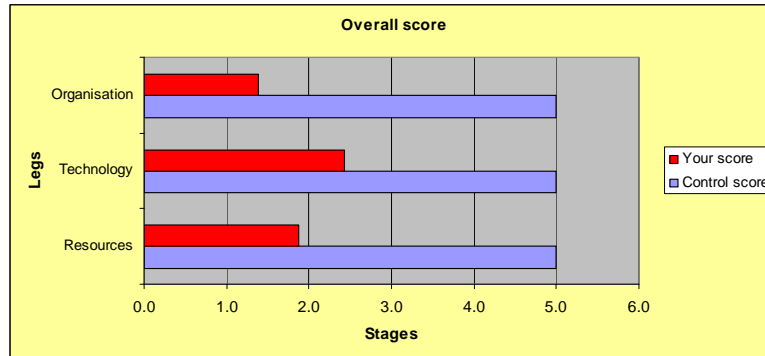
Once you think you've identified where you go, you mark yourself on our online scorecard, which currently exists as a basic GoogleDocs form.



DC|C

... because good research needs good data

AIDA weighted score



DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

At the end of the exercise you get an overall score, which we tweak slightly with some added weights, because some of the elements are more important than others.



DC|C

... because good research needs good data

JISC data management infrastructure call 07/09

Funding is available for projects:

- **to identify requirements** to manage data created by researchers within an institution or a group of institutions, *and then*
- **to deploy a pilot data management infrastructure** to address these requirements. This infrastructure may be at an institutional level, or closer to the researchers.



DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

The first stage of the JISC call is to identify requirements for data management within institutions. As a practical exercise, we thought it would be useful to think how you would scope these.



| D | C | C

... because good research needs good data

Exercise: how will you scope data requirements?

- What are you trying to find out?
- How will you go about finding that out?
- Key questions to be asked?
- How will you ensure stakeholder buy-in?



DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

What are you trying to find out?

- What is the project aim? Do you have a clear, focused description of what you need to determine before the planned pilot infrastructure could be implemented?

How will you go about finding that out?

- What methods will you use to collect information? Where will you find that information? Who will you need to speak to?

Key questions to be asked?

- What areas do you need to cover? Are there examples you can draw from?

How will you ensure stakeholder buy-in?

- Why is this important? What will it do for people? How are people encouraged to take part?



DC|C

... because good research needs good data

Defining scope: What are you trying to find out?

- Aim of the assessment / scoping exercise?
 - Service gap analysis e.g. Oxford scoping study
<http://www.ict.ox.ac.uk/odit/projects/digitalrepository/>
 - Capacity planning exercise i.e. IRs starting to take data
 - Responding to specific need identified e.g. GUARD audit in Glasgow – n.b. can make scoping and selling benefits easier
- Restricting scope e.g. use sampling (IdMRC, Bath), don't include PhDs (Edinburgh, GeoSciences), set time period (GUARD, Glasgow)
http://www.data-audit.eu/docs/DAF_lessons_learned.pdf



DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

We found with DAF that there are various reasons why people may want to do an audit and the kind of information you want to collect and areas to focus on differ with each. As they stand, the forms don't offer much flexibility but we're happy for people to tailor them to suit their needs.

There is a lot of data out there so we found it necessary to restrict the scope and use sampling. Examples are provided on how we did this



DC|C

... because good research needs good data

Methods: How will you go about finding that out?

- Desk based research

- ✓ Good for initial planning stage and to collate background information
- ✗ Remote access to data a challenge and could be hard to understand. TIP: use PhD students

- Questionnaires

- ✓ Good for collecting basic overview and to obtain wide participation
- ✓ Can be useful to identify potential interviewees
- ✓ Potential to use wiki to collect info e.g. ANDS – allows respondents to adapt survey to add other info
- ✗ Uptake can be low – best if pushed by internal advocate
- ✗ Make sure software meets needs – <http://www.disc-uk.org/docs/SouthamptonDAF.pdf>. TIP: Use Bristol Online Surveys

- Interviews

- ✓ Provide best quality information – ability to develop questions to follow up on key points
- ✓ Allow you to gauge better the awareness of data issues
- ✗ Very time consuming – TIP: one person to interview and one to note-take (Imperial College)

TIP: Use lifecycle model to guide discussion – <http://www.disc-uk.org/docs/DAF-Oxford.pdf>



DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

Questionnaires and interviews were the most useful methods for different stages / purposes.

Interviews were really the only way to understand how people work and get good quality information on user needs.



DC|C

... because good research needs good data

Coverage: Key questions to be asked?

Examples from DAF implementations

- DAF audit forms - http://www.data-audit.eu/DAF_Methodology.pdf pp41-51
- Edinburgh questionnaire - http://ie-repository.jisc.ac.uk/283/1/edinburghDAFfinalreport_version2.pdf pp27-28
- Imperial questionnaire - <http://ie-repository.jisc.ac.uk/307/> pp24-29
- Oxford interview framework - <http://www.disc-uk.org/docs/DAF-Oxford.pdf> pp14-15
- Southampton questionnaire - <http://www.disc-uk.org/docs/SouthamptonDAF.pdf> pp7-15 and generic interview schedule pp19-29

Related data studies

- Australian University survey, investigating data management - http://www.apsr.edu.au/orca/investigating_data_management.pdf
- Oxford University projects - <http://www.ict.ox.ac.uk/odit/projects/digitalrepository/> and <http://eidcsr.blogspot.com/>
- Interview questions for Glasgow University preservation policy study <http://www.hatii.arts.gla.ac.uk/research/prespolicy.html>
- Purdue data interviews http://docs.lib.purdue.edu/lib_research/81/
- SCARP project <http://www.dcc.ac.uk/scarp/>



DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

There are lots of examples of how people have used DAF or done similar data survey work. Questionnaires and interview approaches have been made available. Please reference if you use them.



AIDA scope: organisation leg

- Policies and mission statement
- Sharing and continuity
- Audit and monitoring
- Metadata management
- Contractual agreements
- IPR and rights management
- Disaster planning



These are lists of the things we're looking for within each leg.

Sharing and continuity: systematic management of digital data that allows resources to be re-used and repurposed; and remain available for this use. (Not all assets need to remain available if they are scheduled for destruction).

Audit and monitoring: audit trails of user behaviour, and using that information to demonstrate certain things about your activities. Relating the user experience to the actual assets through user evaluation; are you spending money creating assets that are not used?

Metadata: management of metadata as you create assets internally, but also as you ingest them from outside the organisation.



| D | C | C

... because good research needs good data

AIDA scope: technology leg

- IT environment / infrastructure
- Appropriate technologies in place
- Storage, backup, security mechanisms
- Integrity, data corruption
- Obsolescence, format management
- Critical processes
- Security of environment
- Institutional repository management





DC|C

... because good research needs good data

AIDA scope: resources leg

- Business planning
- Resource allocation
- Risk analysis
- Business transparency
- Sustainability of funding
- Staff skills, numbers and training





| D | C | C

... because good research needs good data

Audience engagement: Ensure stakeholder buy-in?

- Senior management approval
- Internal advocate
- Sell the benefits to the individual and institution
 - Imperial College – <http://ie-repository.jisc.ac.uk/307/> pp19-20
 - University of Oregon business case for DAF audit - <http://libweb.uoregon.edu/inc/data/faculty/datainventorybizcase.pdf>



DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

Senior management approval and internal advocates have both been found to get the ball rolling well – see GU policy study and GUARD audit.

Defining benefits is crucial – Imperial College and Oregon have provided examples of their approaches.



DC|C

... because good research needs good data

AIDA outcomes

- AIDA could work for any organisation with digital assets
- Asset management isn't just preservation
 - Control of creation and disposal
 - Sharing
 - Monetising
- AIDA might help strengthen your case



DC101 Lite for JISC data management call - #datman, July 15th 2009, Leeds

AIDA is targeted at HFE but it could feasibly work for any organisation that has digital assets to manage. The exemplars speak of things used in Higher Education, but this could be fixed. It could even work alongside a records management survey; electronic records are digital assets too.

Asset management isn't just preservation; management can include e.g. stop wasting money on creating these expensive assets we don't need, or share these useful assets which we do need, or throw away those assets which have expired.

An AIDA result may help your Institution in some way because it strengthens your case when you ask for more resources and funding.



D | C | C

... because good research needs good data

Thanks!

Further questions and follow up:

- DCC helpdesk for this call - <http://www.dcc.ac.uk/helpdesk/#dataman-helpdesk>
- DAF website – <http://www.data-audit.eu>
- Sarah Jones - s.jones@hatii.arts.gla.ac.uk
- AIDA website – <http://aida.jiscinvolve.org/>
- Ed Pinsent – e.pinsent@ulcc.ac.uk



DC101Lite for JISC data management call - #datman, July 15th 2009, Leeds