Data Audit Framework Methodology

Jones, Ross, Ruusalepp

Release: Version 1.8
Date: 26 May 2009

Draft for discussion
Email comments to info@data-audit.eu

Catalogue entry

Title	Data Audit Framework Methodology
Creator	Humanities Advanced Technology and information Institute (HATII), University of Glasgow
Subject	audit; research data; higher education; further education; data management.
Description	The methodology presented here provides specific guidance on how to plan and execute an audit of research data assets in line with the Data Audit Framework (DAF) conceived by the JISC-funded DAFD project. DAF will enable institutions to find out what data they hold, where it is located and who is (or is not) responsible for it, as well as offering a tool for managing this information and sharing it with other organisations in a controlled environment.
Publisher	HATII at the University of Glasgow
Contributor	Sarah Jones
Contributor	Seamus Ross
Contributor	Raivo Ruusalepp
Contributor	Milena Dobreva
Date	2009.05.26
Type	Text
Format	Adobe Portable Document Format v.1.8
Language	English
Rights	© HATII, University of Glasgow

Citation Guidelines

Jones, S., Ross, S., and Ruusalepp, R., *Data Audit Framework Methodology*, draft for discussion, version 1.8, (Glasgow, HATII, May 2009)

Copyright Notices

Text and diagrams © HATII, University of Glasgow

TABLE OF CONTENTS

PART I BACKGROUND TO THE DATA AUDIT FRAMEWORK

1) Overview

a) Executive summary

The Data Audit Framework Development (DAFD) project was conceived in direct response to recommendations made by Liz Lyon in the seminal JISC-commissioned report Dealing with Data: 'a framework must be conceived to enable all universities and colleges to carry out an audit of departmental data collections, awareness, policies and practice for data curation and preservation.' The aim of the DAFD project is to develop and provide examples for a Data Audit Framework (DAF) adapted to the current needs of data curation activities in UK Higher and Further Education institutions. The project will develop a general solution as well as highlight, through case studies, differences caused by research field and organisation type. DAF will enable institutions to find out what data they hold, where it is located and who is (or is not) responsible for it, as well as offering a tool for managing this information and sharing it with other organisations in a controlled environment.

The DAFD project will distribute a software tool intended to support and facilitate data audit that incorporates a registry component for recording the results of data audits. This tool will provide a powerful mechanism to enable institutions to know what data they have and where it is. The accompanying registry will provide a national perspective of UK research data assets to facilitate future research collaborations and strategic planning.

b) Introduction

The methodology presented here provides specific guidance on how to plan and execute an audit of data assets in line with the Data Audit Framework conceived by the JISC-funded DAFD project. The first two chapters provide background information. Chapter 1 provides details on the scope of the Data Audit Framework and chapter 2 addresses the need for an audit framework, providing several benefits that could be used to formulate a business case. Chapters 3-7 are concerned with the practical work of the audit. Chapter 3 provides details on how to use the Data Audit Framework, chapter 4 covers planning the audit, chapter 5 identifying and classifying assets, chapter 6 assessing the management of data assets, and chapter 7 final reporting and recommendations. Template documents and additional guidance materials for completing the audit are available in the appendices.

c) Glossary of Terms

The definitions below identify specific meanings attributed to common terms within the context of the Data Audit Framework.

auditor	a nominated representative, either from within the department / institution being audited or an external individual, who has been chosen to identify and assess the current value and condition of data assets to make recommendations for their long term management and preservation
appraisal	the process of determining the length of time records should be retained, based on legal requirements and on their current and potential usefulness[1]

---

[1] Richard Pearce-Moses, *A Glosary of Archival and Records Terminology*, (2005), available at: http://www.archivists.org/glossary/index.asp

Data Audit Framework	a framework developed by the JISC-funded DAFD project to identify data assets held within Higher and Further Educational institutions and to explore how they are managed. The framework is structured around audit at departmental or unit level with results being amassed to obtain an institutional or national perspective.
data assets data collections data holdings	broad terms to cover all research data created in Higher and Further Education, including the outputs of funded research projects as well as academics' personal data collections created over the course of their career
department	audits are often described here as being conducted within departments as most of the test audits took place at this level. This is not to suggest that the Framework is only suitable for use at this level; indeed, departmental audits could be scaled up to provide an institution-wide survey or the Framework could be applied directly at such a level using a top-down approach. As such the terms institution and organisation are also used throughout the various sections.
inventory	a detailed list of data assets created by and/or used within an organisation
record	data or information in a fixed form that is created or received in the course of individual or institutional activity and set aside (preserved) as evidence of that activity for future reference[2]
registry	an online system to collect audit results, allowing this data to be federated at institutional or national level. The registry will continue to grow long after the DAFD project ends, enhancing its usefulness as a national reference point.

d)   Scope of the Data Audit Framework

Educational institutions create a wide array of digital assets, including learning objects, electronic journal articles, publications, web resources and datasets. A decision was made between DAFD project management and the JISC to restrict the scope of the Data Audit Framework to research data assets. As such DAF audits are not concerned with administrative assets, such as student databases or inventories of research outputs, nor publications and research papers or web resources. The decision was made on the grounds that existing records management systems and Institutional Repository structures are attempting to address data management issues associated with these other resources. The concept map in Figure 1 provides an overview of the Data Audit Framework.

---

[2] Richard Pearce-Moses, *A Glosary of Archival and Records Terminology*, (2005), available at:
 http://www.archivists.org/glossary/index.asp
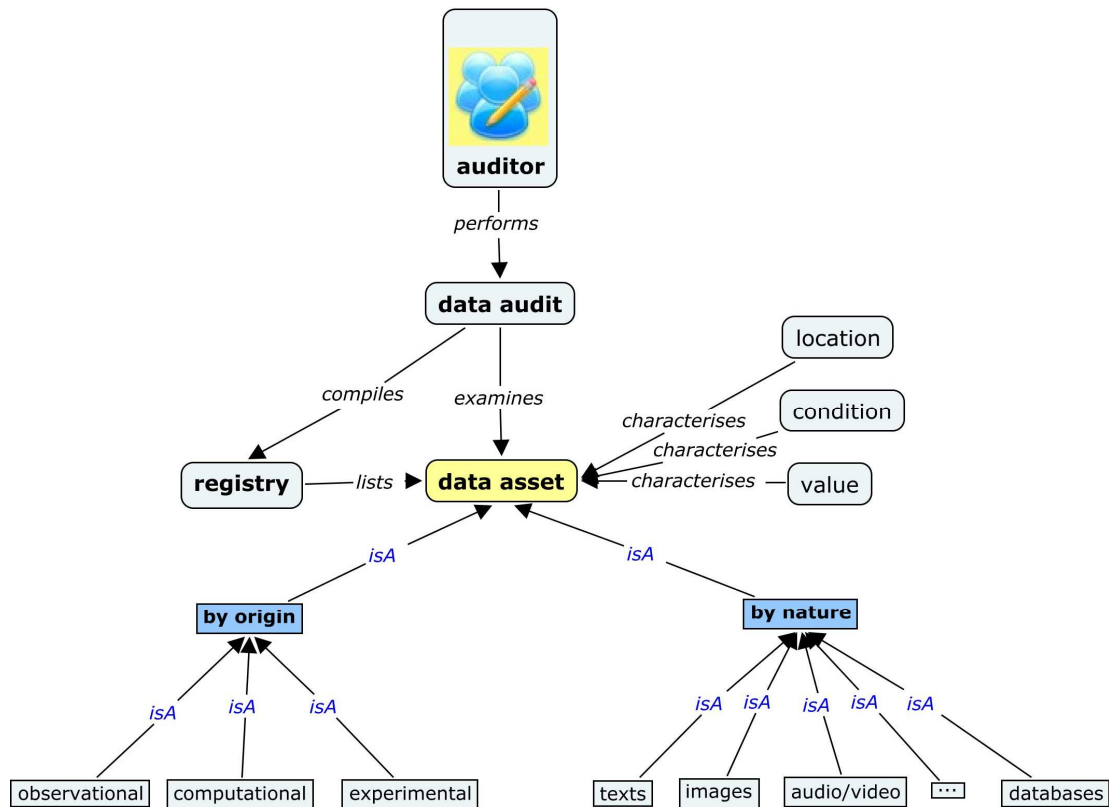
Figure 1: Concept map to demonstrate the scope of the Data Audit Framework[3]

e)   Caveats

Given the project scope and that the data audit tool should be applicable to all universities and colleges we cannot prescribe a fixed top-down method, whereby we determine exactly how the auditor should proceed. We have provided various options and guidance on how to complete the audit but DAF relies on the department or institution and its auditor defining the best course of action.

---

[3] The concept map uses two taxonomic groups for data assets ('by origin' and 'by nature'). These groups follow the classification suggested in the USA National Science Board's (NSB) report 'Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century', published by the National Science Foundation in September 2005. The report is available at: http://www.nsf.gov/pubs/2005/nsb0540/start.jsp

2)   The Need for a Data Audit Framework

a)   Why Audit Data Assets?

To effectively manage data holdings and fully realize their potential, an organisation must first be aware of the location, condition and value of its assets. Conducting an audit will provide this information, raising awareness of collection strengths and data issues to improve overall strategy. An audit will highlight duplication of effort and areas that require additional investment, allowing an organisation to put its resources to best use. It will also highlight inadequacies in data creation and curation practices, suggesting policy change to lessen the risks faced. An organisation that is knowledgeable about its data puts itself in a position to maximise the value of its collections through continued use. Broadly speaking, auditing data brings three main benefits:

- prioritisation of resources which leads to efficiency savings;
- ability to manage risks associated with data loss and irretrievability;
- realizing the value of data through improved access and reuse.

Efficiency savings

Failure to share information on data holdings can result in duplication of effort and general inefficiency. It's only from a position of knowledge that organisations can make informed decisions as to how best to utilise resources for data management. Auditing data assets provides such information, identifying valuable data assets so efforts can be directed towards these. This ensures resources are not wasted managing assets unnecessarily, for example if they are already being curated elsewhere or legally should not be kept. Data can be disposed of appropriately or moved to more cost-effective offsite storage.

The information gained through auditing holdings also assists in forward planning as it gives a more realistic idea of infrastructure and storage requirements. Institutional repositories can be informed of the type and quantity of data being created so appropriate capacity and skills can be secured. Creating a centralised system for managing data leads to further efficiency savings, as curation actions can be carried out on all relevant datasets at once. Indeed, workflows can be improved throughout the whole curation lifecycle by drawing on the recommendations provided at the end of the audit.[4] Records of data assessments also act as an audit trail of management decisions, providing evidence of the level of resource dedicated to data management and justification for the selections made.

Risk management

Organisations need to be aware of data holdings and have appropriate data management structures in place in order to be able to manage risks. The implications of poor data management and quality are widespread and potentially damaging, financially, legally, and in terms of reputation. There are many scenarios in which data could put an organisation at risk. The natural flow of staff turnover can lead to assets being orphaned: the member of staff who created or was responsible for the data may leave before a new asset manager is assigned. This reduces the value of the asset, its usability and its chances of surviving in the long term. Mismanagement of data can also reduce its value, either by failing to update content so the data remains current; creating poor or misleading metadata that

---

[4] The DCC curation lifecycle model maps processes and activities involved throughout the lifespan of digital objects and is available at: http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf

hinders resource discovery; or non-adherence to correct handling, preservation and disaster recovery procedures that leads to data loss. Misuse of data is a major concern, as research data often incorporates sensitive or personal details. If such data is used in an inappropriate way or for a purpose to which subjects did not give consent, questions will be raised as to why appropriate data management protocols were not in place and penalties may ensue.

Proper maintenance of data assets protects against the cost of legal action, so it's necessary that organisations are aware of their assets and how they are being managed. Underlying all risk management is knowledge: knowledge about your assets; where they are; what is happening to them; what potentially detrimental things could happen to them; the likelihood of such occurrences; and most importantly, an understanding of the implications associated with such risks. Armed with such information, organisations can implement policies appropriate to their circumstances.

Access and reuse

Unless organisations are aware of their data holdings it will prove difficult to adequately promote them to encourage reuse. Appropriate levels of information on the data, the organisation and the context of creation are collected in the audit. Such metadata facilitates resource discovery and allows future users to trust the information provided. If this metadata is interoperable, conforming to controlled vocabularies and taxonomies, it will also allow broader data sharing. Audits help to ascertain the value of assets and identify resources that are underused. Little known datasets can then be more actively promoted to raise awareness and encourage reuse. Knowledge of data holdings also promotes synergies within organisations as researchers become more aware of each others' work and more likely to collaborate. Keeping data available in the long-term enables organisations to undertake comparative research, such as longitudinal studies. Promoting access and reuse in this way will have significant benefits in terms of research rating and reputation.

As data is increasingly relied upon to act as an evidential base for analytic work, reassurances of its persistent availability and reusability are sought. Some funding councils mandate data deposit in a trusted repository or require it remains available for a number of years after publication so research findings can be verified. In order to adhere with such regulations organisations need to implement and practice appropriate data management policies; data audit provides a method of ensuring institutional readiness to comply. Moreover the Data Audit Framework toolkit could be used as a preceding step to other audits such as ISO270001: Information Security Management, acting as an overview to lead into more detailed work to help organisations develop robust data strategies.

Auditing data assets and improving data management systems inevitably requires an initial investment of time and resources. The benefits to be gained in terms of efficiency savings, improved risk management, and the ability to fully exploit the value of data should outweigh these initial costs. The reason why an organisation decides to audit its data will vary with each case. Several benefits have been outlined here that could be used to justify undertaking a data audit. Chapter 4.e.ii provides more detail on preparing a business case and tailoring benefits to meet specific organisational needs.

b)   How the Data Audit Framework can Help

Organisations need guidance as to what information should be collected in a data audit, where they can find the information and how they might exploit the information and data once they have collected it. DAF has been developed with these needs in mind. The inventory collects basic information to raise awareness of data holdings and improve knowledge of data issues. The asset

manager is recorded so orphaned works can be identified and investigated further. This aspect may also highlight poor data management practices, for example if one person is responsible for all data, placing the organisation at risk if this person were to leave. The classification process can belie weaknesses in data management. Organisations with large quantities of minor or little used data held locally are unlikely to be making best use of resources. The overall inventory will highlight other inefficiencies such as data being stored by several researchers, which increases storage and back-up costs as well as making it difficult to ascertain which copy is most current.

Information collected in the basic data asset management form (Audit Form 3A) will provide additional benefits. More detailed information is collected on the data asset such as the author, subject, date and rights. This will be crucial for making data available for reuse. Background contextual information is collected such as the original purpose, source and usage frequency. These aspects will assist the organisation in attributing value. Data being actively used or added to may be rated more highly. The source could be central for some domains, such as archaeology or performing arts where data can not be recreated due to ephemerality. Information is collected at this stage on data management to highlight weaknesses. The data type and format will note any collections that are complex to curate and preserve. Finally the 'back-up and archiving policy' and 'management to date' elements will demonstrate to what extent the organisation is engaging with data management. A more extended element set is also provided for cases where additional information is available, or where a more extensive examination of data management practice is desired. This set of information will allow more detailed recommendations to be provided to help ensure the audited organisation can obtain significant efficiency savings, implement better data management and improve reuse.

The framework is designed to be used at both departmental and institutional level without dedicated or specialist staff and with limited investment of time or effort. The audit forms seek to collect only the necessary level of information required to make informed decisions how best to manage data assets in the long term. Optional fields are included for circumstances when extra data is readily available, however the audit does not require onerous amounts of data to be collected unnecessarily. The audit addresses five core questions:

1. What data assets currently exist?
2. Where are these assets located?
3. How have these been managed to date?
4. Which of these assets need to be maintained in the long term?
5. Do current data management practices place these assets at risk?

The information collected by DAF is sufficient to provide a clear overall picture of organisational data collections, current data curation and preservation policies and staff awareness of data issues. Organisations armed with this information can make changes to improve existing data management. DAF provides a simple method of collecting and using this information. The following chapter outlines how to use DAF and details additional sources of support.

PART II THE DATA AUDIT FRAMEWORK TOOLKIT

3)   Guidance for using the Data Audit Framework

a)   How to Use the Data Audit Framework

The guidance in this methodology is based around audits being conducted at research group, unit or departmental level. The Framework could also be applied at institutional level, either by scaling up departmental audits or approaching the audit on an institution-wide basis. The success of the data audit will depend on the strength of the collaboration between the auditor and the organisation being audited. Chapter 4.e.i will discuss how to select an auditor. The ability of this individual or team to perform a complete and accurate assessment depends on the willingness of the organisation being audited to share information and the co-operation of key personnel. It is crucial that a positive relationship is fostered from the early stages.

There are four main stages to the Data Audit Framework:

1. Planning the audit;
2. Identifying and classifying data assets;
3. Assessing the management of data assets;
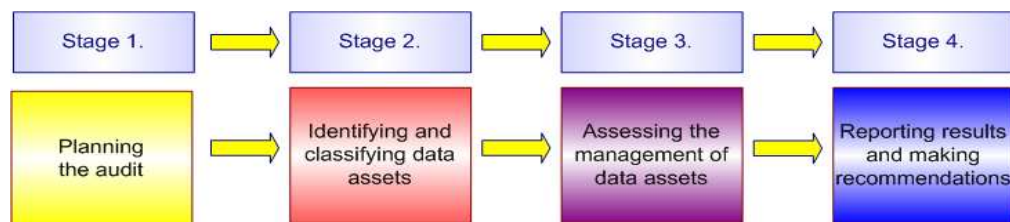4. Reporting findings and recommending change.

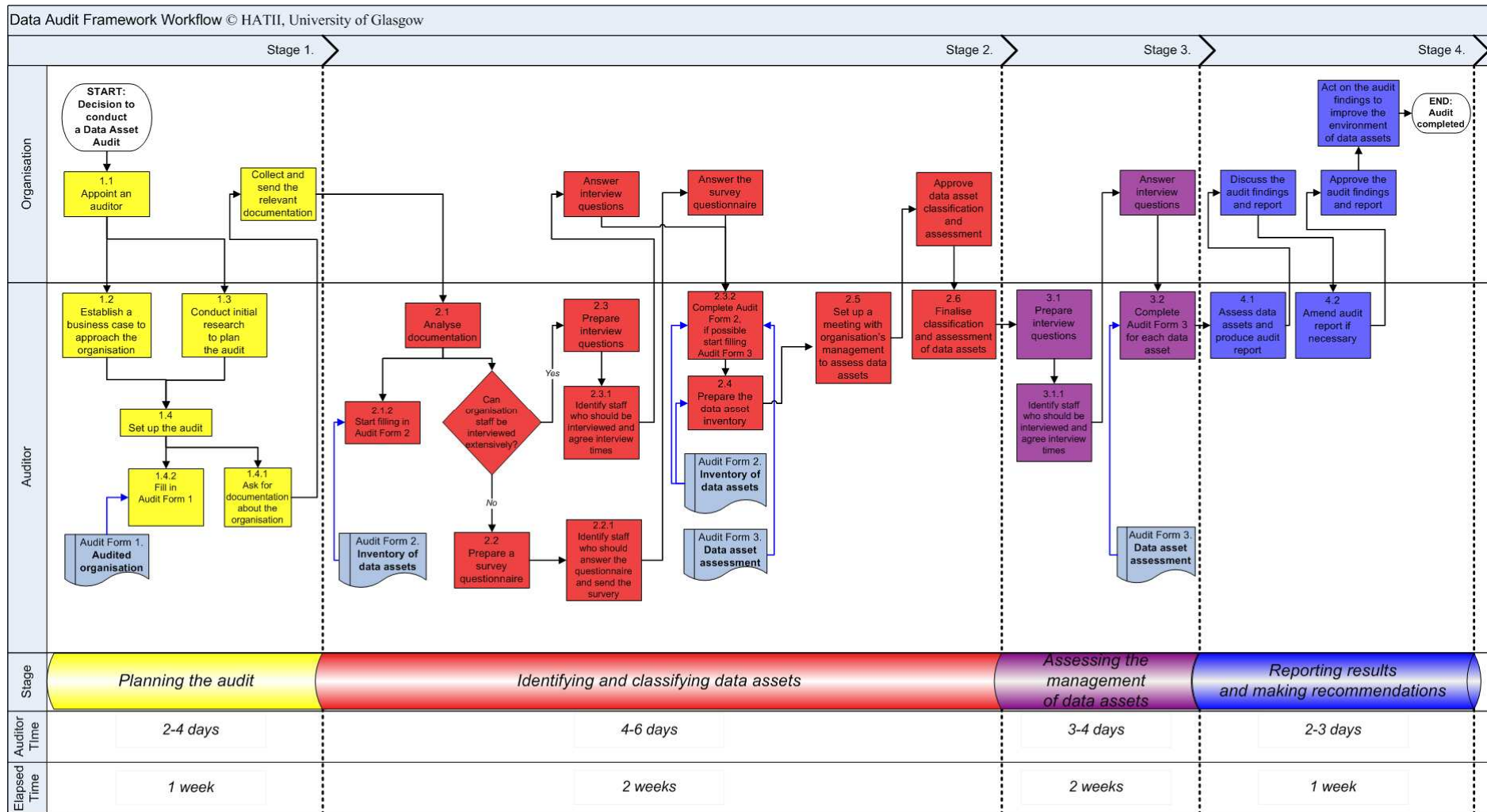**Figure 2:** The four stages of the Data Audit Framework

Figure 3: The Data Audit Framework Workflow

The stages are designed to run sequentially as shown in the simple overview in Figure 2 and the full workflow in Figure 3. Depending on the level of information provided before the auditor begins work on-site, the initial task of analysing documentary sources in stage 2 may begin during stage 1. The overall audit is expected to take between 11-17 days in terms of auditor man hours. The actual time will depend on the size of the department or unit being audited, the number of vital data assets it has and how readily information is made available. Approximately half of this time (5-8 days) will need to be spent on-site. The experience of the first trial audits has shown that around 2 months of elapsed time is needed for each audit. A large amount of this time falls within the planning stage as the audit is set up. The elapsed time during stages 2 and 3 reflects the time needed to schedule interviews. As a general rule of thumb it takes around 2-3 weeks from initially requesting an interview to the event taking place. The auditor will require full access to internal documents and systems during the entire audit period (including elapsed time).

The time requirements above are based on audits taking place at unit or departmental level rather than at university level, as it was thought variations in research and data management policy would necessitate audits at this level. The results of departmental audits could be federated to facilitate an institution wide audit. Alternatively a shallower audit could be conducted across an entire institution to obtain a picture of holdings and data management. This could involve an institution appointing a single auditor to go out to various departments. The process of scoping the audit is discussed below to explain how the Framework can be adapted to small units as well as larger departments or schools.

The three initial test audits run in May and July 2008 took place in departments of varying size with very diverse data collections. As such the scoping and level of granularity adopted varied considerably. The audit at the University of Edinburgh took place in the School of Geosciences, a large research centre with over 150 members of staff and 130 PhD students. Data assets created by PhD students were eliminated from the scope and a decision was made to interview staff based on their involvement in the 5 research groups. Over 35 members of staff were interviewed and an inventory of 25 data assets was created. While this is by no means a comprehensive survey, the later interviews started to provide information already collected, suggesting the most important data assets had been recorded. At Glasgow University meanwhile the audit was conducted in GUARD (Glasgow University Archaeological Research Division), a small commercial unit employing 33 members of staff. Although GUARD is much smaller than the School of Geosciences, its data holdings are very extensive. Around 60 projects are funded each year that create significant digital data assets. The scope was set to create a comprehensive inventory for 2005-2008, however after recording all data assets created over a period of 18 months a decision was made to sample the remaining data. This was due to projects creating typical types and quantities of data, resulting in quite a repetitive inventory and a consistent range of data issues faced. The sample tried to pick out more unusual data collections to explore any other data issues the unit faced that had not already been identified.

Decisions on the overall scope of the audit and level of granularity recorded are dependent on end user expectations. For large scale audits that attempt to cover whole schools, faculties or institutions it is more feasible to create inventories with a limited focus or that act as representative samples rather than comprehensive studies. Pockets of detailed information could feed into larger audits: if a unit is conducting innovative research its data assets may be comprehensively audited while those of other departments are just sampled. Conducting a very detailed audit across an entire institution is likely to demand too high a level of time and resources. As such, identifying the expected outcomes and scoping the audit appropriately is crucial.

There are three component forms used to complete the data audit as shown in Figure 4. These forms are provided in Appendix 1 and various guidance documents for completing the audit are available in Appendix 3.

Figure 4: Data Audit Components

b)   Support for those Using the Data Audit Framework

The Data Audit Framework will be iteratively developed through a series of pilot studies run at four universities throughout the UK and then by a range of audits by four pilot implementation projects. The institutions involved are suitably representative of the wider UK HE data landscape to offer a valuable degree of diversity in terms of scale, scope, materials and strategy.[5] The results of these pilot audits will facilitate the development of an online audit tool: DAF Interactive. DAF Interactive will assist the completion of audit forms and compilation of the final audit report. It will also provide a shared area where users of the tool can seek advice and share knowledge gained from their experiences. DAF Interactive will incorporate a central audit registry into which institutions and departments are encouraged to deposit their audit data so it can be federated at institutional and national level to assist funding councils and strategy makers to plan future work.

---

[5] Four pilot audits have been run in May-June. Details of these studies will be added to the methodology in July.

4)   Stage 1: Planning the audit

a)   Aim of this Stage

There are two key objectives of the planning stage: first to secure organisational buy-in by presenting a robust business case; and second to prepare as much as possible in advance of the audit so time spent on-site can be put to best use. For staff to commit time and effort to assisting the auditor they need to understand the benefits to be gained from auditing data holdings. Securing agreement from the top management and ensuring this commitment to the audit is filtered down is crucial. By conducting background research the auditor can minimise demands placed on staff, as s/he will be familiar with the departmental context and more able to navigate holdings, or at least aware of the best source of advice should challenges be encountered. Moreover organising interview times and locations in advance will ensure staff are prepared and ready to contribute to the audit.

b)   Tasks Associated with this Stage

Planning of the data audit should proceed through the following steps:

- Appoint an auditor;
- Establish a business case to approach the department;
- Conduct initial research to plan the audit;
- Set up the audit.

An example approach made to a department is available in Appendix 3.

It should be noted that the level of time required to complete these steps will depend on whether the auditor is internal or external to the department being audited. DAF has been developed with both scenarios in mind. The initial research stages in which the auditor learns about the departmental context may not be necessary for an internal auditor. An external auditor in contrast is likely to be unfamiliar with the department so will need to complete all tasks associated with this stage.

c)   Anticipated Results of this Stage

By the end of this stage the audit should be ready to go ahead. There are several outputs necessary for this to happen:

- an auditor must be appointed;
- the department in question must consent to the audit and inform staff;
- dates and times for the audit and, where possible, for interviews must be agreed;
- background research should be completed to familiarise the auditor with the department and its data assets.

d)   What Resources are Required to Complete this Stage?

Anticipated effort: 2-4 days, depending on the level of prior knowledge the auditor has of the department being audited. This effort will be spread across a period of around 3 weeks as the auditor waits on information to be provided and responses from staff with whom interviews are requested.

The main focus of effort in this stage will be on research, initially to prepare a business case with which to approach the department and secondly to plan the audit in full. In order to prepare the

business case the auditor will need to find information on current data management practices and where possible issues that are faced. This information will allow the approach and introduction to DAF to be tailored by profiling the benefits specific to the department's needs. Tailoring the approach and highlighting confidentiality is particularly important when the auditor is external as there are likely to be more concerns about granting access to departmental assets and systems. This is especially true in cases where research data is sensitive, so the benefits need to clearly outweigh the risks involved. In cases where an internal auditor is appointed it is likely the department has already recognised the benefits to be gained and decided to run an audit themselves, so the work of the auditor can begin with task 1.3: Conduct initial research to plan the audit.

Before starting this stage auditors should become familiar with the Data Audit Framework to be able to explain the process to the department. Once an agreement exists between the department and the auditor, more detailed research can begin to enable specific details of the audit to be fixed. If the auditor is internal many of the core details may already be known. The research conducted before the audit should attempt to establish:

- details of staff working in the department;
- who is mainly responsible for creating and curating data assets;
- what documents exist that provide information on data holdings;
- whether the department has previously analysed and documented its assets;
- the general context in which the department operates.

Once the auditor is aware of the key personnel to be contacted dates and times can be set for the audit and interviews to take place. Time spent administering these beforehand allows time on-site to be focused on collecting information about data assets. The departmental secretary is likely to be of assistance in allocating times and venues for interviews to take place.

In order to complete this stage, auditors:
- need to establish a relationship with the department;
- need to become familiar with the context of the department;
- need to contact staff to request documentation and set up interviews.

e)   Instructions for Completing the Stage
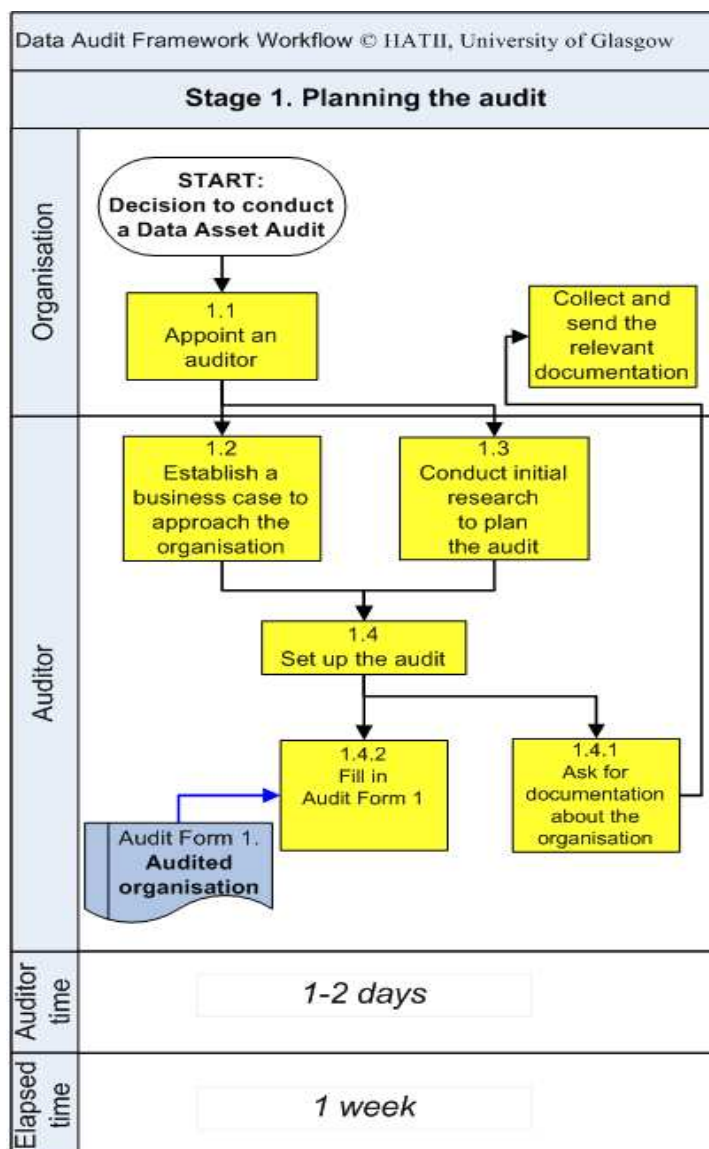
Figure 5: Workflow for Stage 1: Planning the audit

i)   Task 1.1: Appoint an auditor

The auditor can be internal or external. In cases where an entire institution is being audited it's likely a single auditor or audit team would be appointed to go into various departments and assess data holdings, however as the Framework has been developed to support self-audit it is likely that in many cases the auditor will be internal. PhD students have been suggested as ideal auditors as they understand the subject area, are familiar with staff, have access to internal documents and are likely to be able to focus effort on the audit. The results of internal audits may also be easier to maintain as the skills developed to create the inventory will reside within the department. Ideally the time invested to create a baseline inventory will feed into ongoing population and maintenance. Data policy should be adjusted to encourage researchers to add their assets to the inventory as they are created or updated, ensuring accurate information is captured early on to facilitate data management.

An important function of the auditor is to ensure the department takes ownership of the audit results, hence the consultation steps when classifying assets and reporting results. By taking ownership departments can ensure the audit brings long-term benefits with audit findings shaping ongoing data management policy and the inventory realising its full value as a regularly updated working document that supports effective data management and reuse.

The success of the audit will depend largely on the effectiveness of the auditor, so choosing the most appropriate person to perform this role is not an easy decision. Whether an auditor is internal or external to the department there are certain characteristics s/he should possess. Auditors need to be able to gather evidence, so an inquisitive nature and ability to quickly understand systems to determine where information is likely to be found are essential. People skills are also fundamental, especially in cases where the auditor is external as staff many fear the audit process and be unwilling to share information in case it reflects badly on their work. The auditor should be able to put people at ease and effectively communicate the benefits to be gained. Auditors also need a certain level of vision: time on-site is limited so s/he must be able to quickly determine the current state of play and identify what this may mean for the department in the future. In terms of conducting data audits it desirable that the auditor is familiar with data creation and curation practices. Ideally s/he would have a qualification in library, archive or information management, or significant experience working with data assets.

If an internal auditor is appointed the person chosen should be familiar with the organisation's data management. Ideally s/he would be involved in as many aspects of the department's work as possible to have a good breadth of understanding. The individual should be of a suitable level of seniority and trust to be able to encourage staff participation and be granted complete access to internal documents and systems. Time must be set aside for an internal auditor to conduct the data audit. The demands placed on the auditor could be lessened by establishing an audit team. A team would afford multiple perspectives so may result in a more objective outcome.

ii)   Task 1.2: Establish a business case to approach the department

In internal audits it is likely the department has already recognised the benefits of conducting a data audit so a formal business case may not need to be developed. When an external auditor is required to approach a department a business case should be provided to convince management of the benefits. By researching current holdings and collecting information of data management practices the auditor can identify relevant benefits to persuade management to take part. The business case should provide justification for undertaking the data audit, based on the expected costs, risks, benefits and savings. The business case should include:

- reasons why the outcome is needed;
- expected benefits;
- costs and timescale;
- and an investment appraisal.

An example business case for the Data Audit Framework is available in Appendix 3.

The espida model[6] is useful for constructing business cases where benefits are intangible. The approach centres on defining the objectives of an organisation from four perspectives (customers; business processes; innovation and learning; and the financial situation) so that benefits can be aligned with these to make explicit how the proposed action will be advantageous to the organisation.

---

[6] espida, espida Handbook: expressing project costs and benefits in a systematic way for investment in information and IT, (2007), available at: http://hdl.handle.net/1905/691

Once a business case is established the department could be approached in several ways. A common method would be to send a letter or email outlining the benefits, which could then be followed up with a phone call or meeting. Another approach would be to organise a meeting to convince management of the necessity to audit data. A good degree of research should be conducted in preparation for such a meeting to ensure queries could be answered by demonstrating specific benefits. In cases where an institution has agreed to an audit, the support of top-level management could be requested, for example in terms of writing a letter in support of the initiative that could be forwarded to departments. Examples of correspondence used to approach departments are provided in Appendix 3.

iii)   Task 1.3: Conduct initial research to plan the audit

This chapter should be read in conjunction with task 2.1 Analysis of documentary sources as much of the preliminary research needed to begin the inventory could be conducted before the on-site audit begins, particularly if documents are made available to the auditor in advance.

Research should be conducted to familiarise the auditor with the context in which the department operates. This will make it easier to understand the recordkeeping systems that are used. The auditor should also investigate who works within the department and identify the main individuals responsible for data assets. This information will enable the auditor to prepare an audit schedule and set up interview times in task 1.4. The initial research should also explore ways to gain access to data assets that the department holds. Details should be sought on the current location of the assets, who has access to them and whether they are confidential. Gaining consent to access sensitive or confidential datasets may be time consuming so the process of setting up approvals should be started at the earliest opportunity. The background information can first be collected from publicly available sources like:

- annual reports;
- strategic planning documents;
- statistical reports;
- website and intranet;
- scholarly publications;
- research reports.

These documents can often be found on the public websites of departments and may also include references to data assets that the organisation is using or is responsible for.

Further documentation should be requested from the department in advance. If documentation is of a sensitive nature a confidentiality agreement can be signed to assure the organisation of the auditor's commitment not to disclose such details. The confidentiality agreement used in the initial test audits is available in Appendix 2. This allowed the organisation to decide what information needed to be kept confidential and what could be incorporated into the Data Audit Framework registry. As audits involve surveys and interviews ethical approval may need to be sought by each institution to adopt the Framework. The forms used to obtain ethical approval by the DAF development project are available in Appendix 2 for reference. The internal documents that could be made available for the auditor in advance include:

- existing registers of data assets;
- RAE returns;
- funding council monitoring / final reports;
- strategic planning documents (e.g., business plans, departmental development plans);
- annual reports;
- research reports;
- a recent organisational chart;
- research publications;

- relevant legislation and regulations;
- organisation's risk register, disaster plan, contingency plan or exit strategy;
- procedural manuals and system workflows;

- lists and registers of IT systems and information systems;
- documents describing the technical architecture in place;
- system documentation;
- disaster recovery plans for IT systems;

- results of any other relevant audits, departmental assessments or self-assessments;
- vital records lists;
- risk registers.

Information in these sources will help the auditor to navigate the landscape of data assets that the department is likely to have. It should be noted that internal documents must often be treated with caution, because they may be out of date, incomplete or representative of an official viewpoint that has little relation with actual practice. However, these inconsistencies can be rectified in the next tasks.

iv)   Task 1.4: Set up the audit

As many specific details as possible should be set up in advance of the audit. The research conducted in task 1.3 should have identified key personnel the auditor will need to work with. The schedule for the audit should be planned around the availability of these key staff. An audit schedule should be established in advance so all staff in the department are aware of the required contribution and have agreed to spend time sharing information with the auditor. Questionnaires could be prepared during this time and sent through to the department in advance. As many interview times and venues should be arranged in advance to make sure staff are available and have information ready to hand. It should be possible to make an assessment of the time needed to complete the audit at this stage.

It may be helpful for the auditor to establish a main point of contact within the department. The main administrator could be useful in this role and s/he should be familiar with all the staff, perhaps also having an awareness of their work schedules and a certain level of influence to encourage their participation. The main contact should have a good overall awareness of the work of the department to be able to point the auditor in the right direction or suggest the most relevant person to contact.

f) Checklist

Before proceeding to the next stage, the auditor must ensure that s/he has successfully completed the following items:

- established who will be responsible for conducting the audit;
- obtained approval from the departmental management for the audit to take place;
- agreed for staff time to be set aside for the audit and for the auditor to be granted access to the departments records and systems;
- sign a confidentiality agreement if necessary – a template is available in Appendix 2;
- conducted initial research into the department and its data creation / curation practices;
- agreed dates for the audit to take place and if possible set-up interviews with key personnel;
- ensured staff are aware of what the audit process entails and are committed to taking part;
- have attempted to minimise demands on staff time by sending interview questions and requesting documentation in advance.

5)   Stage 2: Identifying and classifying data assets

a)   Aim of this Stage

The purpose of this stage in the audit process is to establish what type of data assets exist in the department and classify them according to their value for the organisation. Essentially, an inventory of data assets is compiled through a mapping exercise. The mapping exercise should produce as complete a picture of the existing data assets as possible. Quality of the entire audit depends on this first knowledge-gathering exercise. Vital assets should not be missed or they will be excluded from further scrutiny and evaluation. Therefore, the first stage should identify all data assets and resources that members of staff use in their work, in order to paint a complete picture of what resources are necessary for the organisation to conduct its work. The classification step will determine the scope of further audit activities. The next stage of the audit – assessing the management of data assets – will only be applied to these assets that have been appraised as vital or important for the department. The scope of further tasks in the audit should only concern the assets that really are the responsibility of the department itself, leaving aside shared data assets that cannot be controlled by it.

b)   Tasks Associated with this Stage

Throughout this stage the auditor will develop a structured list of data assets and their managers that help the department to achieve its goals and objectives and document the way it works.

The identification of data assets should proceed through the following steps:

- Analysis of documentary sources;
- Conducting a written survey;
- Following it up with interviews;
- Preparing the data asset inventory (including classifying identified assets)
- Approving and finalising the asset classification.

Each stage is documented using the audit forms (later the electronic audit tool). See Audit Form 2 in Appendix 1.

c)   Anticipated Results of this Stage

This stage will result in an inventory of data assets and their managers, divided into groups according to their appraised value for the organisation. The appraisal decisions will be approved by the management of the organisation. The inventory will form the main basis for the next stages of the audit where the data assets are assessed and described in detail. The comprehensiveness of the list prepared in this stage plays a crucial role in the usefulness of later recommendations.

d)   What Resources are Required to Complete this Stage?

Anticipated effort required: 4-6 days, depending on the size and type of department or organisation being audited. If interviews have been planned in advance during Stage 1, elapsed time should only be a couple of weeks. This could increase if staff are unavailable to participate.

Analysing a department's data assets is a rigorous and resource-intensive process. The main time investment is expected to go into identifying all the data assets and into considering activities, assets and staff as an interlinked organism. Unless existing and up-to-date inventories or lists of assets and

technology are available, these will have to be created as part of this stage. Classification can be done in parallel to identification, though the results of this process should be verified with management and if necessary reviewed before proceeding to the next step.

Auditors can return to the list developed during this stage at a later point during the assessment process in order to add to it when new information about further assets comes to light.

Before starting this stage, auditors should:

- have a general understanding of the department and the contexts within which it operates;
- obtain managerial support to undertake the analysis of its data assets;
- acquire a list of members of staff and their responsibilities;
- determine whether the department has previously analysed and documented its data assets.

If the department has recently been analysed for other purposes it may be possible to draw on the results of such work, rather than starting from scratch. Projects that may involve an analysis of assets and activities include:

- compiling inventory of IT systems and resources;
- systems implementation;
- activity-based costing or management;
- business classification development;
- quality accreditation;
- business process re-engineering
- workflow automation.

Lists, registers or inventories of assets and technology may have been compiled for various purposes during analyses of business, compliance studies and audits, contingency planning exercises, etc. Most departments maintain an inventory of IT hard- and software and their licenses, which can also point to data assets that are managed with the help of the inventoried technology. If the analysis arising from such projects is available, auditors will need to consider how, why and when the projects were undertaken to determine whether their findings are applicable for the purposes of this audit.

In order to complete this stage, auditors:

- must have access to internal documents, such as strategic planning documents, departmental structure charts, lists of assets, technology and systems, operational manuals, procedural guides, task and target lists to identify assets and records that identify vital assets such as risk registers, business contingency plans and disaster plans for classification;
- may need to have the right to circulate a written survey questionnaire to all members of staff;
- will preferably need to have access for interviews to all members of staff who are using or are responsible for department's data assets;
- will need to have access for interviews to at least managerial and IT personnel with further knowledge of the data assets, technology and systems, as well as those responsible for their long-term curation.

e)   Instructions for Completing the Stage
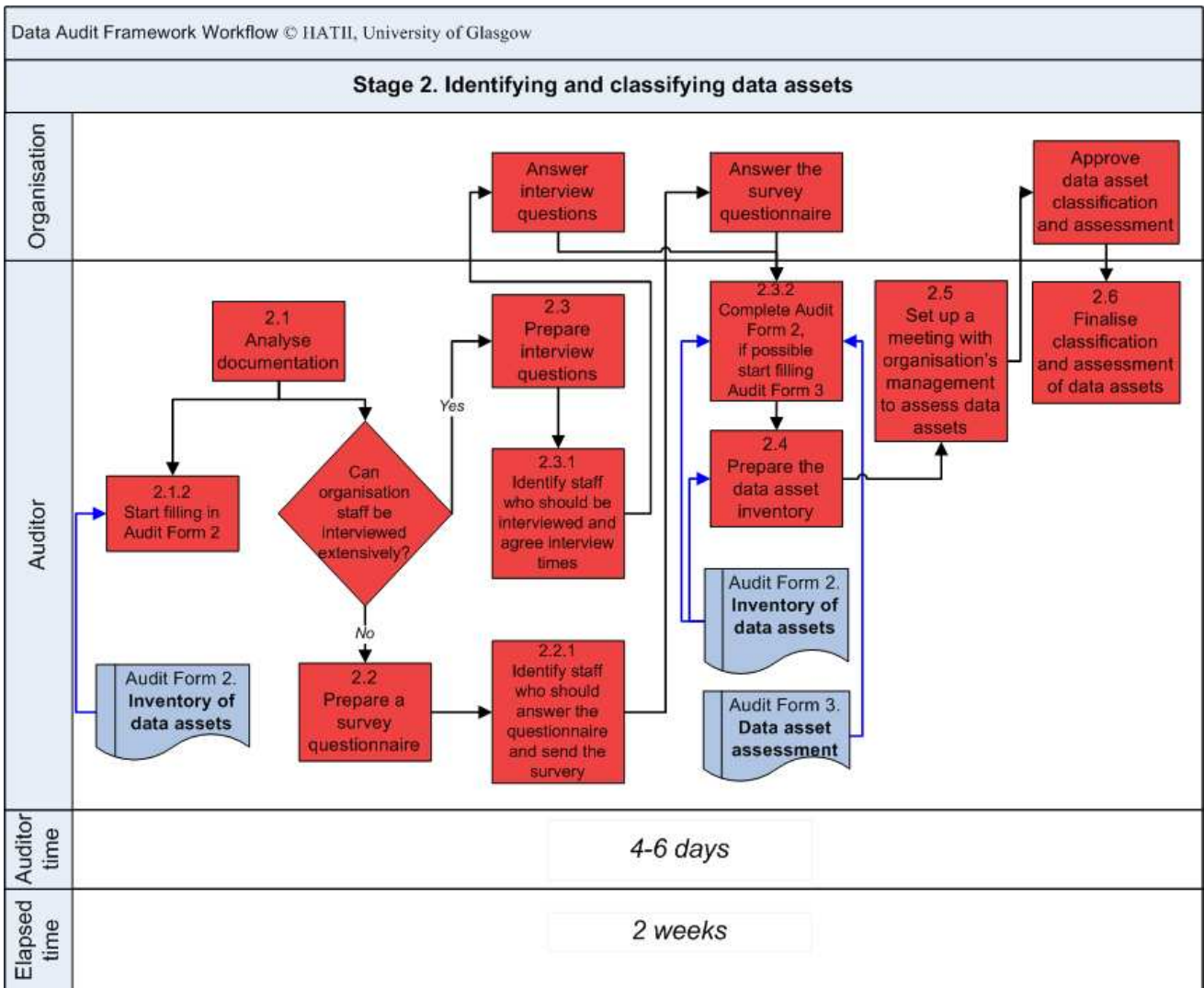
Figure 6: Workflow for Stage 2: Identifying and classifying data assets

i)   Task 2.1: Analysis of documentary sources

The audit process will begin some time prior to a visit to the department taking place, with initial research being undertaken into the context of the department and its activities (see also Chapter 4.e.iii Task 1.3: Conduct initial research to plan the audit). Understanding the context where the data assets are being created, used and managed will help with identification and facilitate the interview process described in Task 2.3. Supporting documents will be requested from the department and gathered in advance, in order to ensure that time spent on-site is optimised. Data assets that are mentioned in department's documents should be documented using the audit forms as in the following example:

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**
Bach bibliography database	A database listing books, articles, thesis, papers and facsimile editions on the works of Johann Sebastian Bach	Senior lecturer	RAE return for 2007, http://www....ac.uk/...	Vital	Part of ongoing research project so still being populated. High usage	An MS Access database in H:\Research\Bach\ Bach_Bibliography.mdb. NB! Must ask what version of MS Access is used!

Figure 7: Data Asset Audit Form 2

The first column of the form should include the name of the data asset as found in the documents. An optional column could also be used to record the variant or name commonly used for the asset in the organisation. The second column should provide a brief description of what the data asset includes and how it is used, based on the information that can be drawn from the documentation. The third column identifies the person responsible for managing and updating the data asset, as can be determined from documentation. In many cases the responsibility for data assets is shared by members of a research team, several teams or even several organisations. It is advisable to identify at this stage at least one person as the asset manager (e.g., leader of the research team) who could be contacted with further questions about the asset. The fourth column includes a reference to the document where the information about the data asset was discovered and a reference where the document itself can be found. The following two columns deal with classification: the first attributing a classification to the asset and the second justifying this classification. The final column is reserved for comments and reminders for the auditor. Any additional information that is readily to hand such as details of the creation date, original purpose, type of data, or file format may also be included here.

Having become acquainted with the organisation, its context and activities the auditor should be prepared to start a more detailed information collection exercise.

ii)   Task 2.2: Conduct survey questionnaire

Depending on the availability of staff at the audited organisation, the auditor will have to choose the appropriate method for collecting detailed information about data assets. If the members of staff can be interviewed in person, it would be preferable to use the interviews described in Task 2.3. If access to members of staff is limited during the period when the audit is conducted, a written survey questionnaire should be used to collect the information. In cases where the majority of information is collected by questionnaire due consideration must be given to classification early on so guidance can be provided with the questions to ensure all respondents are working to the same criteria. Suggested classification schemes are discussed in more detail in chapter 5.e.iii Conduct follow-up interviews.

Questionnaires provide a simple means of gathering information from a large number of respondents. The survey questionnaire should be based on the Audit Form 2, and could be disseminated in electronic form to all members of staff or made available as a web-based form on the organisation's intranet, wiki or a separate password-protected website. When the electronic data asset audit tool becomes available, members of staff can be advised how to access the audit tool and how to fill in the forms about each data asset that they use and/or are responsible for.

Questionnaires have the disadvantage of usually achieving low rates of return. Personal contact with staff is lacking and the quality of the information is depending on respondents understanding the task and questions, and being willing to reply thoroughly and accurately. Usually follow-up enquiries are needed to clarify or complete the forms. Thus, information collected through the survey can be used to fill in Audit Form 2 to prepare follow-up interviews to fill any gaps remaining in Audit Form 2. Analysis of the survey results will help the auditor to decide:

- what other documents can be consulted, based on the references column, that could reveal further information about data assets;
- where are the biggest gaps in the knowledge about the data assets (for example, when description of assets has been left incomplete; when assets were found in the documentation, but the survey did not reveal any further information about these);
- who are the asset managers who need to be interviewed further to complete the data asset identification exercise;
- what questions need to be asked during follow-up interviews.

It is also possible that the initial survey produces more information than is asked in Audit Form 2, for example technical information can be added about an asset, or a problem with it reported. It would be useful to include a comments or a free text field in the survey questionnaire. This information can already be used to begin filling in Audit Form 3 (cf. Chapter 7.e.i below).

iii)   Task 2.3: Conduct follow-up interviews

Follow-up interviews should be set up to clarify unclear information or gaps in information gathered from documentation and written questionnaire. Interviews can be organised on an individual basis or as group interviews that involve all managers and users of the same data asset. The selection of staff to be interviewed should be based on their knowledge of the data assets, not their seniority or position. At this stage of the audit, gathering a number of interviewees into a focus group for a group interview may be the fastest way to collect the required information and to agree classifications by consensus.

Interviews can be set up as:
- structured interviews using a questionnaire distributed to the interviewees prior to the interview;
- unstructured interviews as a free-flowing discussion;
- group interviews as discussions or brainstorming exercises;
- auditor participating at a department or unit meeting where this can be discussed.

Interviews can potentially be time-consuming, but they have the benefit of providing personal contact with members of staff and it is often possible to collect more detailed information than from questionnaires. Interviews allow supplementary questions to be asked on the spot and thus offer greater flexibility in following up initial answers. Informal information can also be collected, such as opinions on data assets and their users, or suggestions for improvement.

The interviews should be documented using the Audit Form 2. It would be preferable to have any previously collected information already in the audit tool available and visible during the interview, for example by using a data projector.

Interviews are arguably the best method of agreeing classifications as staff can engage in discussion as to which criteria are most important to the organisation being audited. The DAF audit tool proposes three categories for classifying the organisation's data assets, though obviously the most appropriate classification will vary on a case by case basis:

Category	Description
Vital	Vital data assets are crucial for the functioning of the organisation, their efficient management and protection should be the first concern of the organisation. Vital data assets will include datasets and information systems that: • are still being created or added to; • are used on frequent basis in the course of organisation's work; • underpin scientific replication e.g. revalidation; • play a pivotal role in ongoing research; • or are being using to provide services to external clients and partners.
Important	Important data assets include the ones that: • the organisation is responsible for, but that are completed (i.e. no data is added to them); • the organisation is using in its work, but less frequently; • the organisation may use in the future to provide services to external clients.
Minor	Minor data assets include those that the organisation has, but has no explicit need for, or it no longer wants to have the responsibility for (e.g., data assets that could be handed over to specialised preservation and curation service providers). Shared data assets or information systems which the organisation cannot control may also be classified as minor as somebody else is responsible for long-term curation.

Other classifications are possible for categorising data assets and in discussion with the audited organisation the proposed three-tier classification can be extended to include further categories, or to include sub-classification. Examples of sub-classification would be to show whether important data assets need to be retained and managed on-site or to determine if minor data assets need to be actively managed or if they could be disposed of.

It is feasible to have a numerical rating aligned to the three categories to help to show where on the scale each asset falls. For example, the vital assets could be characterised with scores from 8 to 10, important assets with scores from 4 to 7, and minor assets with scores from 0 to 3.

The determination of the value of a data asset is important for managing risk and reliability for the organisation's business. Value helps to decide the level of security and control the asset requires, as well as the investment in maintenance that should be made.

Other examples of classification exist, for example the United States National Science Foundation is using the following three functional categories of data collections:[7]

- research database collections, which are specific to a single investigator or research project;
- resource or community database collections, which are intermediate in duration, standardisation, and community of users;
- and reference collections, which are managed for long-term use by many users.

The classification used at this stage of the audit does not need to be final. Indeed, based only on the information collected in this stage, it would be very difficult to offer a final classification of data assets. The main purpose of the classification here is to help scope the next stage of the audit and to exclude data assets that do not have significant value for the department being audited from further analysis. A more elaborated version of the data asset classification can be offered in the final audit report.

iv)   Task 2.4: Preparing the data asset inventory

Following the information collecting exercise the data asset inventory (Audit Form 2) should be cleaned and prepared for the next stage by the auditor, filling in gaps and highlighting the issues that have come to light about some assets (e.g., technical difficulties of using the asset, inadequately described assets, orphaned assets with nobody assigned responsibility for maintaining them, etc.). If possible, the information provided for the inventory in questionnaires should be edited using the same editorial style. The first draft of the complete inventory will be presented to the departmental management for approval and further comments.

v)   Task 2.5: Approving classification with departmental management

The classification of data assets prepared by the auditor should be discussed with management, or, if possible, with all managers identified as responsible for the data assets. The best format for the discussion would be a physical meeting with everyone involved participating and being able to voice their concerns and requirements. As a result of the discussion a decision should be made which data assets should be analysed further. It is likely that amendments and changes will need to be made to the inventory to move data assets from one category to another, or to add further sub-classifications if the organisation has requested these. The auditor will make any necessary changes in the inventory (Audit Form 2) based on the feedback received during the discussion.

f)   What to do in the Event of Required Information Being Unavailable

The information on assets of the organisation should be made available to the auditor in a complete and unabridged form. If this proves difficult, senior management should be contacted and the necessary authorisations acquired to gain access to the required information. Further usage restrictions can be agreed for the list of assets that is created in this stage in order to protect the

---

[7] NSF, *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century*, Appendix D. *Digital Data Collections by Categories*. http://www.nsf.gov/pubs/2005/nsb0540/nsb0540_11.pdf

sensitive information it may contain. Not being able to access complete information on what the organisation is doing will be detrimental to the audit.

g)   Discussion

Determining responsibility for maintaining data assets and assuming an adequate level of granularity for this information can be challenging. Depending on the source of information, it could be recorded as an individual person, a project team, an office or a unit within the organisation (e.g. literature, legal). It is difficult to prescribe a fixed level for this information at this stage of the audit – further details on the responsibility for the data asset can be documented at stage 3 of the audit. Sufficient information should be collected about asset managers at stage 2 to be able to approach the correct people with further questions later on.

The responsibility for data assets that an organisation is using in its work can sometimes be difficult to determine. In addition to data assets that have been created internally by the organisation itself, the work usually relies on other, shared data assets and information systems. In the case of a university department or institute this could be for example:

1) systems and data assets owned and managed centrally by the university:
   - student research databases;
   - teaching course information databases;
   - library;
   - institutional repository (e-prints, theses);
   - electronic records and document management systems;
   - archive of the university;
2) external resources
   - international and national level data collections archives;
   - subject-based data collections, libraries, on-line resources;
3) collaborative environments;
   - wiki;
   - intranet and other group-work tools.

Although the focus of this audit will be the data assets managed by the department itself, it is useful to chart all the assets that are being used by the institution, in order to:

1) be able to better ascertain the value of each asset held by the organisation and its relationship with data assets curated elsewhere;
2) build a network of data assets used by the organisation and to determine the vital systems and data assets among them that the institution at large should support and preserve (because units within the institution are relying on these assets).

Many institutions are engaged in collaborative research projects, creating data assets that are maintained by a whole community. This naturally raises the question of who is responsible for long-term maintenance. Editorial rights for such collaborative data assets as well as parties responsible for long-term curation and preservation must be defined so collaborators recognise and fulfil their role.

The decision of what criteria to apply when appraising data assets can be difficult to make. Should it include only the current business needs of the department or should the auditor also look to potential future uses of assets and try to assess their future or even archival value? Trying to do both may result in a decision to classify everything as vital assets. It is often difficult to know who may want to make

use of a dataset in the future, but the creators of the dataset can see many potential uses for it, even though it is unlikely that all of these will be realized.

The classification applied at this stage of the audit can be quite crude at first, to encourage further discussion of how the organisation is making use of its data assets. As a result of the discussion, some minor or important data assets can be moved to a higher category, because not enough is known about them and it is expected the analysis stage will reveal further details about the asset to allow for better classification. Although a few methods of classification have been suggested here, audits may require different approaches. One of the preliminary test audits altered the classification scheme as the unit in question was a commercial centre with many active projects at any one time. To restrict the number of vital assets the classification was amended to focus more heavily on reuse and revenue.

h)   Checklist

Before proceeding to the next stage, the auditor must ensure that s/he has successfully completed the following items:

- contacted all key personnel either via a questionnaire or in an interview;
- identified all or most data assets;
- assigned all data assets to a category to define the scope of the next stage;
- documented the audit process and findings.

6) Stage 3: Assessing the management of data assets

a) Aim of this Stage

The aim of Stage 3 of the audit is to collect further information about vital and important data assets and define their place in the organisation's knowledge base. A recommendation is also made for collecting further information about data assets that would allow us to assess whether the current level of resources provided for management, control and curation is sufficient to maintain the value of data asset.

Based on the information collected at this stage, the auditor may need to re-classify data assets further into selected categories, and highlight problem areas and issues with the way department is managing its data, or point out significant threats or risks that data assets are facing. The final stage of the audit will provided an audit report that is presented to the organisation.

b) Tasks Associated with this Stage

In this stage the auditor will:

- begin filling in Audit Form 3 for each data asset classified as within the scope of this stage;
- collect further information on each data asset from members of staff and complete Form 3;
- finalise the documentation created during the audit.

Information required to complete this stage of the audit will have already begun to accrue in the previous stages of the audit and the auditor may have started to use Audit Form 3 already. In this case, the forms just need to be completed here.

c) Anticipated Results of this Stage

The primary result of this stage of the audit is a complete set of documentation on each vital data asset that makes it possible to complete the charting exercise of data assets.

d) What Resources are Required to Complete this Stage?

Anticipated effort required: 3-4 days, depending on the number and nature of data assets identified. Depending on how many interviews have been planned in advance and staff availability, elapsed time is expected to be in the region of 2 weeks.

A significant proportion of information to be collected will be readily available and may have been copied to the audit forms when documentation was analysed in Stage 2 of the audit. However, some information, for example dates, figures on funding and access restrictions, can be more difficult to trace or to get hold of, extending the effort required for this stage.

Before starting this stage, auditors should:

- have a complete inventory of classified data assets.

Documentation identified for Stage 2 of the audit will provide a lot of the information required to fill in Audit Form 3. This allows the auditor to begin working on Stage 3 off-site and only approach members of staff for further information once the existing information has been documented in the forms.

In order to complete this stage, auditors:

- will need to contact members of staff for further information, either by setting up interviews or sending questions via e-mail;
- will potentially need authorisation to contact members of staff of other units or organisations that provide services to the audited organisation, for example IT managers, system administrators or the financial department.

e)   Instructions for Completing the Stage
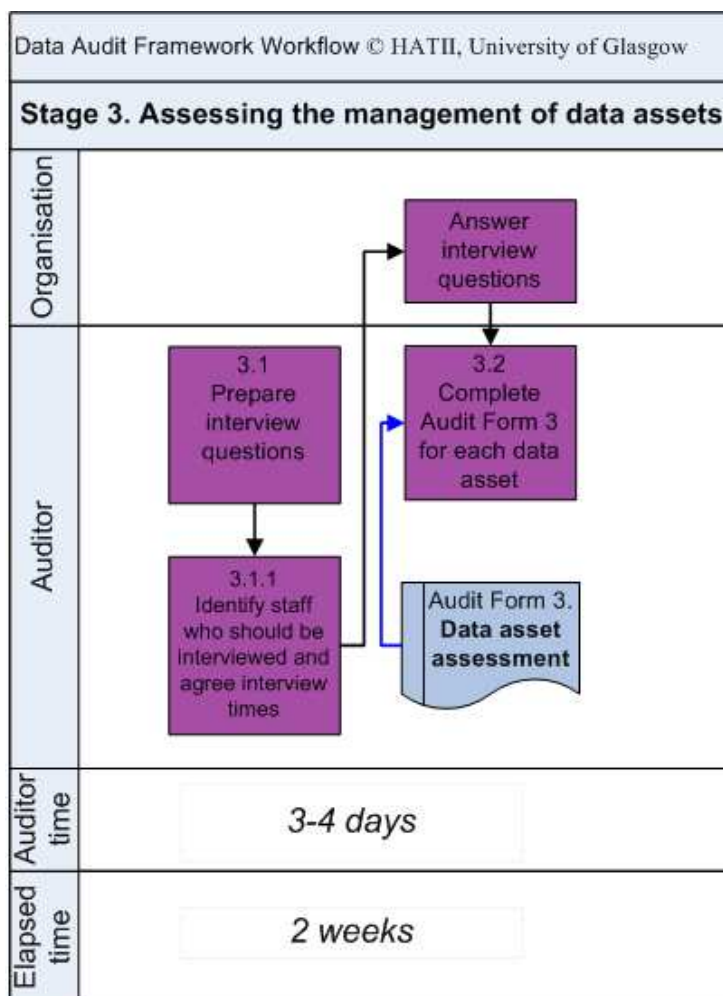
Figure 8: Workflow for Stage 3: Assessing the management of data assets

i)   Task 3.1: Prepare interviews with members of staff

The auditor will probably have started filling in the Audit Form 3 for each vital and important data asset already in the previous audit stages. It is advisable to complete as much of the forms as possible off-site and then agree interviews with the individual asset managers or members of staff with knowledge about the particular information that is still missing from the forms.

It is likely that some information will be easier to collect for all or most data assets from representatives of shared service departments that the organisation is relying on, for example, IT services, legal department or even finance.

In collecting information on data assets, the auditors have been provided with two options:

1) collect a minimal set of information to expand the asset inventory and gain an overview of existing data management;
2) collect an extended set of information to look in fine detail at existing data asset management.

The minimum set of information allows the organisation to discover synergies and duplication of effort between data assets, as well as to identify orphaned or disused data assets that could be resuscitated to actively take part in the organisation's knowledge management again. It will also support data asset management decision making. The extended set will provide departments with a wealth of information on how assets are currently being curated, used and preserved, enabling them to drastically improve existing management. In the online tool an additional field numbered 0 will be provided in both element sets for an automatically generated system ID.

The minimum required information about each data asset is the following:

No	Parameter	Comment
Audit Form 3A: Data asset management (Core element set)		
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 the 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	Back-up and 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 or preservation activities performed*

Figure 9: Data Asset Audit Form 3A (core element set)

The first column in this form provides the number of the parameter that is to be collected, the second includes the name of the parameter (based on the ISO 15836 *Dublin Core metadata element set*), the third is the space where the auditor is expected to write information about each data asset.

The extended set of information that could be collected by the auditor, should the organisation see the need for it and be prepared to invest staff time in providing all the information, is divided into six interrogatories: what, why, who, where, when, and how. The parameters are divided into mandatory (indicated with M in the last column) and optional (indicated with O in the last column).

No	Parameter	Comment	M/O*
Audit Form 3B: Data asset management (Optional extended element set)		
		* M – mandatory, O - optional	
Description		
1	ID	*Assigned by the audior or organisation*	M
2	Title	*Official name of the data asset*	M
3	Type	*Description of the technical type of the data asset (e.g., database, photo collection, text corpus, etc.)*	M
4	Owner(s)	*Formal owner(s) of the data in terms of intellectual rights*	M
5	Subject	*Information about the subject coverage of the data*	M
6	Language	*The language(s) of the data asset content*	M
7	Variant name	*Alternative or commonly used name, if available*	O
8	Level	*What level is the current description being applied (e.g., and entire collection of data objects, an individual database, a coding table used in conjunction with the main database)*	O
9	Abstract	*Text describing the data asset*	O
10	Keywords	*Relevant keywords that describe the data asset*	O
Provenance		
11	Original purpose	*Description of what was the main reason for the data asset's creation*	M
12	Description	*A description of the information contained the data asset*	M
13	Start date	*Date when the data asset was created / started*	M
14	Updating frequency	*The frequency of updates to this dataset to indicate currency*	M
15	Description of context	*Description of the original use and context of the data asset*	M
16	Source	*The source(s) of the information found in the data asset, description of data collection methods or third party datasets that the data asset is using*	M
17	Completion date	*Date when the data asset was completed / data collection ceased (if data is no longer being added)*	O
18	Date last modified	*Date when the data asset was last updated or changed*	O
19	Management to date	*History of maintenance and integrity of the data asset*	O
20	Curation to date	*History of preservation and curation activities*	O
Ownership		
21	Data creator(s)	*Person, group or organisation responsible for the intellectual*	M

No	Parameter	Comment	M/O*
Audit Form 3B: Data asset management (Optional extended element set)		
* M – mandatory, O - optional		
		content of the data asset	
22	Asset manager(s)	*Name and contact details of the person responsible for the management of the data asset*	M
23	Rights	*Indication of the user's rights to view, copy, redistribute or republish all or part of the information held in the data asset*	M
24	Usage constraints	*Access restrictions applied to the data asset*	M
25	Former asset managers(s)	*Curation history / chain of custody for the dataset*	O
26	Other acknowledgments	*Contact names of the other researchers and co-authors who have worked on the data asset*	O
27	Usage frequency	*Estimated frequency of use and if known required speed of retrieval to determine IT infrastructure and storage needs*	O
28	FoI, DP, personal privacy issues	*Description of any potential data protection or ethical issues related to content of the data asset and if any restrictions based on these are currently applied*	O
29	Potential re-uses	*Description of the potential re-uses of the data asset that its current manager(s) can envisage*	O
Location		
30	Current location	*Path or www address where the data asset can be found*	M
31	Coverage	*Intellectual domain or subject area covered by the information in the data asset. Geographisc area and time period covered*	M
32	Relation	*Description of relations the data asset has with other data assets*	M
33	Version	*Current version of the dataset*	O
34	Responsibility for the asset in the long term	*Description of the retention policy and management of the data asset for the longer term*	O
35	Can / should it be handed to a service provider for curation?	*Can the organisation hand over curation of the data asset in the long term or will it need to seek expert advice or services for this?*	O
Retention		
36	Long term value	*Description of the value the data asset could have in the long term*	M
37	Back-up and archiving policy	*Number of copies of the data asset that are currently stored, frequency of back-up and archiving procedures*	M
38	Disaster recovery measures	*Description of recovery process in case the data asset has been damaged*	M
39	Retention period	*Planned end date or a retention period for the data asset (if exists)*	O
40	Preservation policy	*Description of any digital preservation or curation activities planned or applied to the data asset*	O
Management		
41	File format(s)	*File format(s) and their version(s) the data asset is using*	M
42	Structure of the data asset	*Details of the structure of the data asset (e.g. table structure of a database, components of a text or photo collection)*	M

Audit Form 3B: Data asset management (Optional extended element set)			
No	Parameter	Comment	M/O*
			* M – mandatory, O - optional
43	Documentation available	Documentation that is available (e.g. user manuals, coding tables), including references to its location	M
44	Audit trail and fixity	Description of any measures in use for ensuring the authenticity of the data asset (e.g., checksums, audit trail)	M
45	Current cost	Current maintenance costs (annually) of the data asset	M
46	Funding basis	Source of funding available for the data asset now and likelihood of its continuance in the future	M
47	Original cost of creating the asset	The original cost of creating the data asset	O
48	Planned costs for maintenance	Existing plans and costings for improving the data asset or its maintenance (e.g., for the next budget period)	O
49	Size	Size of the data asset in Mb/Gb	O
50	Hard- and software requirements	Description of any specialised hard- or software requirements the data asset has	O

Figure 10: Data Asset Audit Form 3B (optional extended element set)

ii)   Task 3.2: Finalise the information on each data asset

Once information on each data asset has been collected, the forms (multiple copies of Audit Form 3) should be completed, checked for any remaining gaps, and compiled into audit report for assessment of data assets.

If any data assets were identified as needing re-classification in Stage 3, then these assets should now be assigned to their categories, based on the detailed information available about them.

f)   Checklist

Before proceeding, the auditors must ensure that they have:

- Completed Audit Form 3 for each data vital and important asset as determined by the scope of the audit;
- Finished classifying all data assets (re-classifying where necessary).

7)   Stage 4: Reporting results and making recommendations

a)   Aim of this Stage

The final stage is concerned with drawing together the results of the data audit to produce a final report that recommends actions to improve ongoing data management. Once the audit is over the organisation can apply complementary methods and tools to improve its data management. Relevant details of such tools are provided in chapter 7.e.i.

b)   Tasks Associated with this Stage

Finalising the data audit should proceed through the following two steps:

- produce an audit report;
- meet with management to finalise audit report.

c)   Anticipated Results of this Stage

The output of this stage will be a final report. This should provide information on the data collected during the audit and make recommendations for change. The report should provide management with the necessary information to prepare a business case to justify any investment needed to remedy existing data issues.

d)   What Resources are Required to Complete this Stage?

Anticipated effort: 2-3 days. Depending on the time taken to convene a meeting with departmental management and to amend the final report in light of their feedback, elapsed time for this stage may be in the region of 1 week.

Before starting this stage auditors should have:
- created an inventory of organisational data assets;
- classified data assets into vital, important and minor records;
- assessed vital and important records to be able to comment on existing data management.

In order to complete this stage, auditors:
- will collate and analyse information from the audit;
- identify weaknesses in data management and provide recommendations for improvement.

e)   Instructions for Completing the Stage
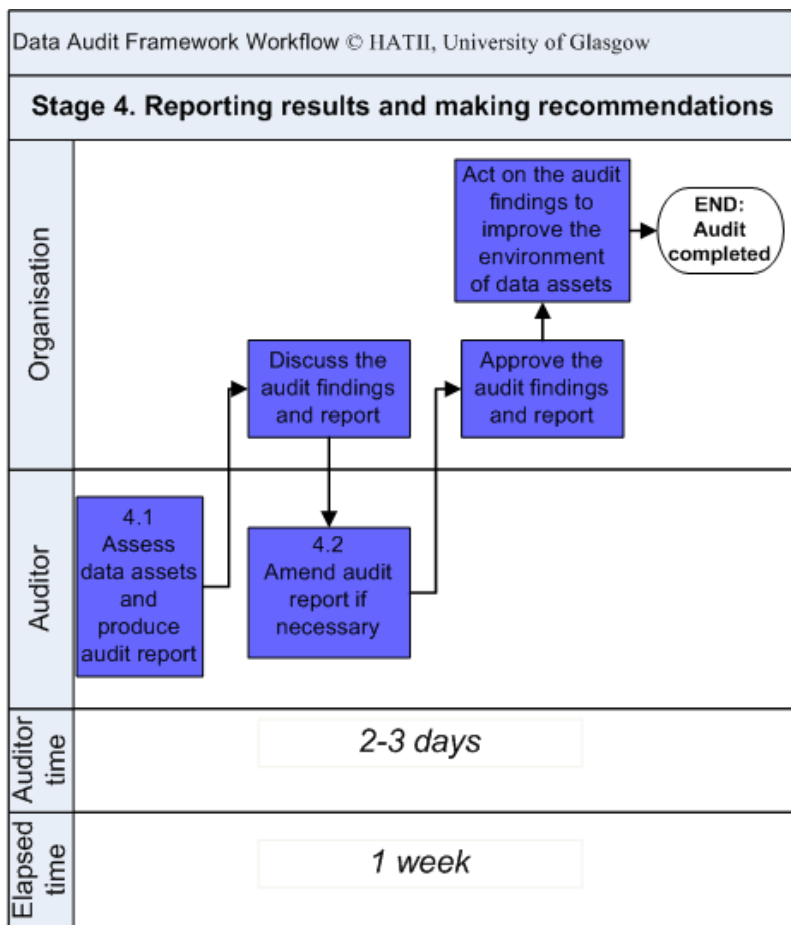
Figure 11: Workflow for Stage 4: Reporting results and making recommendations

i)   Task 4.1: Produce audit report

This initial version of the Data Audit Framework is a paper-based tool. Over the course of the next few months an online tool will be developed in line with recommendations made by the series of pilot audits. The online tool will enable a large part of the final report to be automatically generated from details entered into the audit forms. The information provided will be enhanced with analytical material to assist the organisation's senior management to initiate change in light of identified weaknesses. Over time comparative statistical information can be provided from the DAF registry.

For the initial pilot audits the final report will need to be compiled manually. Suggested chapters for the report are:

1. Brief description of the organisation being audited;
2. Profile of the data holdings based on the asset inventory and classification;
3. Recommendations for improved data asset management.

Details of the organisation being audited will have been compiled during Stage 1 using Audit Form 1. A summary of present data management systems and practices, as well as details of the budget available for such work should be provided in the final report. The data asset inventory will provide all

of the information required for the next section of the report. An overview of data holdings should be provided that profiles the main data types and asset managers and identifies any weaknesses in current curation management, for example inadequate metadata or a lack of information on the location or condition of data assets. The inventory also provides information on the classification of data assets. The final report should profile these findings to inform management of the breakdown of their assets into vital, important and minor holdings. Appropriate recommendations should be made to dispose of minor assets or move little-used holdings to more cost-effective storage. The final section of the report should provide general recommendations and point the organisation to complementary tools and sources of advice. Some complementary tools and sources of advice are provided below.

- The DRAMBORA self-audit tool could be used by organisations to review their capability to preserve data assets and to develop a risk register to help manage risks related to preserving the data holdings.[8]
- The ISO-BS 27001 standard method could be followed to conduct an information security audit to ensure data assets were not being corrupted.
- JISC's Information Management infokit can be applied to ensure best practice is followed during the creation, active use, semi-active use and final stages of a data asset's lifecycle.[9]
- Tools available through the Digital Curation Centre can be used to identify, describe, manage and preserve data assets.[10]
- Public bodies can apply The National Archives' Information Management Assessment Tool to ensure compliance with data protection and Freedom of Information legislation.[11]
- The International Records Management Trust's Records Management Capacity Assessment System (RMCAS) can be applied to map capacity levels to the infrastructure and systems needed to deliver effective records management.[12]
- Procedures can be implemented that draw on best practice guides for the creation and curation of digital resources. Guides are available from various sources including: Arts and Humanities Data Service; Digital Curation Centre; Digital Preservation Coalition and UKOLN.[13]

ii)   Task 4.2: Meet with management to finalise report

After an initial draft of the audit report has been completed it should be presented to management to discuss the findings and recommendations. This provides the organisation with an opportunity to seek further guidance from the auditor on the best way forward. It also allows the auditor to elicit management feedback and amend the final report in view of any suggested changes. Any necessary changes should be made after this meeting so the final version of the audit report can be passed to organisational management for formal approval. Inviting management comments at this stage will encourage ownership of the audit results and should help ensure the asset inventory becomes a working document that is maintained by the department and used as a tool to support ongoing data management.

---

[8] The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) tool is available to download at: http://www.repositoryaudit.eu/

[9] Details of the JISC Information Management Infokit are available at: http://www.jiscinfonet.ac.uk/information-management

[10] For details of DCC tools see: http://www.dcc.ac.uk/tools/digital-curation-tools/

[11] To download the Information Management Assessment Tool so to: http://www.nationalarchives.gov.uk/recordsmanagement/section46.htm

[12] To download RMCAS go to: http://www.nationalarchives.gov.uk/rmcas/downloads.asp

[13] Best practice guides can be found at the following sites: http://ahds.ac.uk/creating/guides/index.htm, http://www.dcc.ac.uk/resource/, http://www.dpconline.org/graphics/handbook/, http://www.ukoln.ac.uk/interop-focus/gpg/

f)   Checklist

By the end of this final stage the auditor should have:

- collated and analysed audit findings for inclusion in the final report;
- recommended changes to curation practices in light of data management inefficiencies that became apparent during the audit;
- met with organisational management to discuss the draft audit report;
- prepared the final audit report and obtained formal approval of the findings.

APPENDICES

8) Appendix 1: Template audit forms and worked examples

a) Audit Form 1: Audited organisation

Audit Form 1: Audited organisation	
Organisational details	
Organisation identifier	*assigned by auditor when auditing, for example, more than one department within an university;* *does not need to be used if not needed*
Organisation name	
Organisation address	
Organisation contact details	
Organisation contact name	*the main contact for this audit*
Organisation type	*University, Department, Institute, Archive, Data Centre, Computer Service, Research Team etc*
Description of the organisation	*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. The answer should include the name and a reference where this document can be found (if known).*
Responsibilities	*Details of staff who have clear responsibility for managing the data assets*
Budget	*Level of resources available and details on whether it is sufficient to manage data at the present time*
Audit details	
Dates of the audit	
Name of the auditor	
Auditor contact details	
Date of completing the form	

Audit Form 1: Audited organisation

Organisational details

Organisation identifier	GU_TFTS_001
Organisation name	University of Glasgow Theatre Film and Television Studies Department
Organisation address	Theatre Film and Television Studies Department, University of Glasgow, The Gilmorehill Centre, University Avenue, Glasgow G12 8QQ
Organisation contact details	tel: +44 (0)141-330 2678 fax: +44 (0)141-330 1084 email: info@tfts.gla.ac.uk
Organisation contact name	Jayne Birch, Departmental Secretary J.Birch@tfts.gla.ac.uk
Organisation type	University department
Description of the organisation	The Theatre, Film and Television Studies Department is one of most well-established departments for these subjects in the UK. We offer courses which build up critical knowledge, creative approaches and confidence in your own skills. We have a strong emphasis on research but we also encourage students who bring together academic and practical skills and activities.

Data management details

Strategies	Data assets are managed in line with the departmental data management policy, wich can be seen at: T:policies/data_management_policy.doc
Responsibilities	The departmental secretary is responsible for updating the list of research assets and ensuring they are managed according to departmental policy. Individual researchers are responsible for specific data assets they create.
Budget	1% of the current departmental budget is set aside for managing its data. This appears sufficient at present.

Audit details

Dates of the audit	19th – 28th May (23rd, 26th & 28th May on-site)
Name of the auditor	Richard Kellerman
Auditor contact details	R.Kellerman@arts.gla.ac.uk
Date of completing the form	29th April 2008

b)   Audit Form 2: Inventory of data assets

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 or position*	*Where reference to asset was found*	*Choice of: Vital Important Minor*	*Reason for classification or comment on the classification chosen (should be based on the discussion with organisation's management) Could also include action suggested for each asset*	*Could include creation date, original purpose, type of data asset, or file format.*

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
Shakespeare performance database	*A database listing performances of Shakespeare's works, detailing location, dates, production company and often including images.*	*Christine Ryan, Senior theatre lecturer*	*RAE return 2007 http://www....ac.uk/*	*Vital*	*Still being added to and plays pivotal role in several ongoing research projects.* *Need to be preserved to ensure ongoing accessibility for a minimum of 5 years.*	*Originally created with AHRC grant between 2004-2006. Database continually updated and used by HE community.*

Student productions image collection	Collection of images of student productions staged between 1965 and 1996	Jayne Birch, departmental secretary	Mentioned in interview with Jayne Birch on 23rd April 2008. Resource can be seen at: T:teaching/resources/studprods/img.jpg	Important	Resource completed but still used within the department for teaching and creative purposes	Collection created by head of theatre until she left the department. Collection placed under care of secretary. Images used as student resource for new productions.
Glasgow cinema-goers dataset	Collation of data on Glasgow cinema-goers drawn from government statistics for use in journal article	Dave Berry, Head of Film	'Sitting in the back row: the habits of Glasgow cinema-goers' in Film Monthly, available at http://www.....co.uk	Minor	Data may be needed to validate claims but as it stems from government statistics it is already available elsewhere	Created in course of research into Glasgow cinema-goers. Dataset only used by Dave Berry.
Shakespeare performance database	A database listing performances of Shakespeare's works, detailing location, dates, production company and often including images.	Christine Ryan, Senior theatre lecturer	RAE return 2007 http://www....ac.uk/	Vital	Still being added to and plays pivotal role in several ongoing research projects. Need to be preserved to ensure ongoing accessibility for a minimum of 5 years.	Originally created with AHRC grant between 2004-2006. Database continually updated and used by HE community.

c)   Audit Form 3: Data asset management

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 the 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	Back-up and 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 or preservation activities performed*

Audit Form 3B: Data asset management (Optional extended element set)			
No	**Parameter**	**Comment**	**M/O***
		* M – mandatory, O - optional	
		Description	
1	ID	*Assigned by the audior or organisation*	M
2	Title	*Official name of the data asset*	M
3	Type	*Description of the technical type of the data asset (e.g., database, photo collection, text corpus, etc.)*	M
4	Owner(s)	*Formal owner(s) of the data in terms of intellectual rights*	M
5	Subject	*Information about the subject coverage of the data*	M
6	Language	*The language(s) of the data asset content*	M
7	Variant name	*Alternative or commonly used name, if available*	O
8	Level	*What level is the current description being applied (e.g., and*	O

No	Parameter	Comment	M/O*
Audit Form 3B: Data asset management (Optional extended element set)			
		M – mandatory, O - optional	
		entire collection of data objects, an individual database, a coding table used in conjunction with the main database)	
9	Abstract	Text describing the data asset	O
10	Keywords	Relevant keywords that describe the data asset	O
Provenance			
11	Original purpose	Description of what was the main reason for the data asset's creation	M
12	Description	A description of the information contained the data asset	M
13	Start date	Date when the data asset was created / started	M
14	Usage frequency	Estimated frequency of use and if known required speed of retrieval to determine IT infrastructure and storage needs	M
15	Description of context	Description of the original use and context of the data asset	M
16	Source	The source(s) of the information found in the data asset, description of data collection methods or third party datasets that the data asset is using	M
17	Completion date	Date when the data asset was completed / data collection ceased (if data is no longer being added)	O
18	Date last modified	Date when the data asset was last updated or changed	O
19	Management to date	History of maintenance and integrity of the data asset	O
20	Curation to date	History of preservation and curation activities	O
Ownership			
21	Data creator(s)	Person, group or organisation responsible for the intellectual content of the data asset	M
22	Asset Manager(s)	Name and contact details of the person responsible for the management of the data asset	M
23	Rights	Indication of the user's rights to view, copy, redistribute or republish all or part of the information held in the data asset	M
24	Usage constraints	Access restrictions applied to the data asset	M
25	Former asset manager(s)	Curation history / chain of custody for the dataset	O
26	Other acknowledgments	Contact names of the other researchers and co-authors who have worked on the data asset	O
27	Usage frequency	Estimated frequency of use	O
28	FoI, DP, personal privacy issues	Description of any potential data protection or ethical issues related to content of the data asset and if any restrictions based on these are currently applied	O
29	Potential re-uses	Description of the potential re-uses of the data asset that its current manager(s) can envisage	O
Location			
30	Current location	Path or www address where the data asset can be found	M
31	Coverage	Intellectual domain or subject area covered by the information in the data asset. Geographisc area and time period covered	M
32	Relation	Description of relations the data asset has with other data assets	M

No	Parameter	Comment	M/O*
**Audit Form 3B: Data asset management (Optional extended element set)**			
		M – mandatory, O - optional	
33	Version	*Current version of the dataset*	O
34	Responsibility for the asset in the long term	*Description of the retention policy and management of the data asset for the longer term*	O
35	Can / should it be handed to a service provider for curation?	*Can the organisation hand over curation of the data asset in the long term or will it need to seek expert advice or services for this?*	O
		Retention	
36	Long term value	*Description of the value the data asset could have in the long term*	M
37	Back-up and archiving policy	*Number of copies of the data asset that are currently stored, frequency of back-up and archiving procedures*	M
38	Disaster recovery measures	*Description of recovery process in case the data asset has been damaged*	M
39	Retention period	*Planned end date or a retention period for the data asset (if exists)*	O
40	Preservation policy	*Description of any digital preservation or curation activities planned or applied to the data asset*	O
		Management	
41	File format(s)	*File format(s) and their version(s) the data asset is using*	M
42	Structure of the data asset	*Details of the structure of the data asset (e.g. table structure of a database, components of a text or photo collection)*	M
43	Documentation available	*Documentation that is available (e.g. user manuals, coding tables), including references to its location*	M
44	Audit trail and fixity	*Description of any measures in use for ensuring the authenticity of the data asset (e.g., checksums, audit trail)*	M
45	Current cost	*Current maintenance costs (annually) of the data asset*	M
46	Funding basis	*Source of funding available for the data asset now and likelihood of its continuance in the future*	M
47	Original cost of creating the asset	*The original cost of creating the data asset*	O
48	Planned costs for maintenance	*Existing plans and costings for improving the data asset or its maintenance (e.g., for the next budget period)*	O
49	Size	*Size of the data asset in Mb/Gb*	O
50	Hard- and software requirements	*Description of any specialised hard- or software requirements the data asset has*	O

Audit Form 3A: Data asset management (Core element set)		
No	**Parameter**	**Comment**
1	ID	TFTS_0001
2	Data ceator(s)	Christine Ryan, Senior theatre lecturer, Theatre Film and TV Studies, University of Glasgow
3	Title	Shakespeare performance database
4	Description	Database listing performances of Shakespeare's works, detailing location, dates, production company and often including images. Covers all Shakespeare's plays that have been performed by RSC between 1950-1990 in Stratford and London as well as a range of stagings by other production companies.
5	Subject	William Shakespeare; performance; theatre;
6	Creation date	2004-2006
7	Purpose	Was conceived and has always been used as a research tool for theatre scholars investigating the staging of Shakespeare's works and social historians investigating the theatre
8	Source	Theatre Museum, production company and local theatre holdings, newspapers
9	Updating frequency	Monthly
10	Type	Database and image collection
11	Format	MS Access 2003
12	Rights and restrictions	Rights held by University of Glasgow and photographers as detailed in metadata. Content has been cleared for educational use.
13	Usage frequency	c.25 unique users per day – resources available within minutes
14	Relation	Referenced in *Shakespeare's Works on Stage* C Ryan (2007) ISBN 1-85604-330-3
15	Back-up and archiving policy	Three copies stored: one in department; one back-up copy at archives; and one back-up copy offsite. The database will be offered to the Instiuional Repository after the end of the research project for long-term preservation.
16	Management to date	Managed in accordance to the University of Glasgow theatre department's data management policy. Back-up copies created weekly and held within the University Archives and at a location off-site and limited people have editorial permissions. No preservation actions have been performed

No	Parameter	Comment	M/O*
	Audit Form 3B: Data asset management (Optional extended element set)		
		* M – mandatory, O - optional	
		Description	
1	**ID**	TFTS_0001	M
2	**Title**	Shakespeare performance database	M
3	**Type**	Database and image collection	M
4	**Owner(s)**	Christine Ryan, Senior theatre lecturer, Theatre Film and TV Studies, University of Glasgow	M
5	**Subject**	William Shakespeare; performance; theatre;	M
6	**Language**	English (British)	M
7	**Variant name**	n/a	O
8	**Level**	Item level – individual database	O
9	**Abstract**	The database was conceived as a research tool for theatre scholars investigating the staging of Shakespeare's works. It lists UK performances of Shakespeare's works, detailing location, dates, production company, cast, and director. Many records also include images. The database covers all of Shakespeare's works and has complete entries for the Royal Shakespeare Company and productions at the Globe.	O
10	**Keywords**	William Shakespeare; performance; theatre;	O
		Provenance	
11	**Original purpose**	The database results from an AHRC award and as such was conceived as a research tool for theatre scholars investigating the staging of Shakespeare's works and social historiansd investigating the theatre.	M
12	**Description**	Database listing performances of Shakespeare's works, detailing location, dates, production company and often including images	M
13	**Start date**	February 2004	M
14	**Updating frequency**	Monthly	M
15	**Description of context**	Was conceived and has always been used as a research tool for academic sholars	M
16	**Source**	Theatre Museum, production company and local theatre holdings, newspapers	M
17	**Completion date**	Ongoing	O
18	**Date last modified**	15/05/2008	O
19	**Management to date**	Managed in accordance to the University of Glasgow theatre department's data management policy. Back-up copies created weekly and held within the University Archives and at a location off-site and limited people have editorial permissions.	O
20	**Curation to date**	No preservation actions have been performed to date but the database would be migrated through software changes	O
		Ownership	
21	**Data creator(s)**	Christine Ryan, Senior theatre lecturer, Theatre Film and TV Studies, University of Glasgow	M
22	**Asset Manager(s)**	Christine Ryan, Senior theatre lecturer, Theatre Film and TV	M

No	Parameter	Comment	M/O*
Audit Form 3B: Data asset management (Optional extended element set)			
			*M – mandatory, O - optional
		Studies, University of Glasgow	
23	Rights	Rights held by University of Glasgow and photographers as detailed in metadata. Content has been cleared for educational use	M
24	Usage constraints	Freely accessible for educational purposes	M
25	Former asset manager(s)	Created and maintained by Christine Ryan	O
26	Other acknowledgments	*Co-authors:* James Boyle; Matthew Leigh; Annabel Green.	O
27	Usage frequency	c.25 unique users per day – resources available within minutes	O
28	FoI, DP, personal privacy issues	Content has been cleared for educational use	O
29	Potential re-uses	Could be reused by the general public to find details of main production companies / theatre locations, or perhaps by commercial companies interested in staging works	O
Location			
30	Current location	Available online at: http://www....ac.uk/ and internally at: T:research/funded_research/outputs/shakespeare.mdb	M
31	Coverage	Theatre studies / performance	M
32	Relation	Referenced in *Shakespeare's Works on Stage* C Ryan (2007) ISBN 1-85604-330-3	M
33	Version	Version 2.1	O
34	Responsibility for the asset in the long term	Maintenance agreed for 5 years in first instance though asset likely to be maintained after this.	O
35	Can / should it be handed to a service provider for curation?	Decision made to managed internally for short-term but may be handed over to a data centre thereafter.	O
Retention			
36	Long term value	Resource already very well used so raising profile and reputation of the department. Continued success could help leverage more research council funding.	M
37	Back-up and archiving policy	Three copies stored: one in department; one back-up copy at archives; and one back-up copy offsite.	M
38	Disaster recovery measures	Regular back-up copies created and three copies stored to ensure one will be recoverable in case of disaster.	M
39	Retention period	Planned to be maintained within the department for a minimum of 5 years.	O
40	Preservation policy	No preservation actions have been performed to date but the database would be migrated through software changes.	O
Management			
41	File format(s)	Microsoft Access 2003	M
42	Structure of the data asset	Table structure: play title | staging date(s) | location | production company | cast names | director | image files	M

Audit Form 3B: Data asset management (Optional extended element set)			
No	**Parameter**	**Comment**	**M/O***
		* M – mandatory, O - optional	
43	**Documentation available**	A user guide was produced to help people search the entries. This is available online at: http://.....ac.uk and within the department at T:research/funded_research/outputs/shakespeare_guide.pdf	M
44	**Audit trail and fixity**	Checksums performed bi-monthly and edits only made by approved users. Additions periodically checked by asset manager.	M
45	**Current cost**	c.£3,000 per annum	M
46	**Funding basis**	AHRC funding has come to an end so maintenance costs and additions to database absorbed by department	M
47	**Original cost of creating the asset**	c.£300,000 over 2 years	O
48	**Planned costs for maintenance**	Ongoing cost at level of £3,000 agreed for next 5 years	O
49	**Size**	4Gb	O
50	**Hard- and software requirements**	Currently using MS Access	O

9)   Appendix 2: Forms

a)   Ethical approval forms

Ethical approval was gained from University of Glasgow to cover all project partners in the Data Audit Framework Development project. The forms used are provided here for reference. Three consent forms were used:

- confidentiality and non-disclosure agreement (1 copy per organisation)
- human consent form (1 copy per interviewee)
- interview release form (1 copy per interviewee)

At the start of the audit process the confidentiality and non-disclosure agreement was signed with the management of the organisation being audited. Before any surveys or interviews took place an agreement was made as to which staff within their organisation would be available as potential interview candidates. Each of these individuals were approached and asked if they would be willing to participate. If they initially indicated that they would be willing to take part we provided them with a human subject's consent form outlining the scope of the research and what would be expected of them if they participated in the process. Once the signed human consent form was returned we conducted the interviews. After an interview was completed the interviewer wrote up the discussion. Depending on the time available and interviewees preference, this report could be returned for approval and to agree the level of anonymity.

CONFIDENTIALITY AND NONDISCLOSURE AGREEMENT

between the parties

(**[INTERVIEWER NAME])**

member of the Consortium carrying out the Project called DAFD under JISC Digital Repositories Programme (hereinafter "**DAFD Partner**")

and

[**COMPANY NAME**]
[COMPANY DESC] (hereinafter "**Company**")

Generally referred to as "**Parties" or "Party"**

WHEREAS, [COMPANY NAME] agrees to furnish DAFD Partner certain confidential information relating to the digital preservation of digital material including, but not limited to, information about strategies, processes and digital objects for the purpose of conducting research to identify existing research data holdings for the Data Audit Framework Development Project (hereinafter "**Purpose**").

WHEREAS, DAFD Partner agrees to review, examine, inspect or obtain such confidential information (including interview data) only for the purposes described above, and to otherwise hold such information confidential pursuant to the terms of this Agreement.

BE IT KNOWN, that [COMPANY NAME] has or shall furnish to DAFD Partner certain confidential information to be used for the Purpose and the right to discuss or interview representatives of [COMPANY NAME], the Parties hereby agree as follows:

1.   Each Party (hereinafter referred to as the "Receiving Party") agrees (i) to use the other Party's (hereinafter referred to as the "Disclosing Party") Confidential Information solely for the Purpose of this Agreement and (ii) to keep and not disclose it to any third party, except for Confidential Information which:

   a) was lawfully in the Receiving Party's possession before receipt,

   b) was public knowledge at the time of disclosure hereunder or which thereafter becomes public knowledge through no omission of the Receiving Party regarding its obligations pursuant to the present Agreement; or

   c) becomes available to the Receiving Party from an independent third source not bound by an obligation of secrecy to the Disclosing Party with respect to such Confidential Information, or

   d) is, at any time, developed by employees of the Receiving Party independently from the Confidential Information and such independent development can be substantiated, or

   e) is required to be disclosed by law (including but not limited to the Freedom of Information legislation) or by a requirement of a regulatory or government body or court of competent jurisdiction with power to compel the disclosure

   f) is cleared for inclusion in the Data Audit Framework registry

2.   The Parties agree that the Confidential Information received hereunder will be treated with the same care as if it were their own proprietary confidential information and will be disclosed only to those of their employees needing to use the Confidential Information for the Purpose of this Agreement. The Parties will advise each such employee of the confidential nature of the Confidential Information and of the existence as well as of the content of this Agreement.

3.   DAFD Partner may disclose the Confidential Information received from the Company to other members of DAFD consortium for the Purpose specified above. Such a disclosure is not considered to be an unauthorised disclosure or breach of this agreement. DAFD Partner shall limit disclosure of confidential information to those within the DAFD consortium who need to access the Confidential Information in order to use it for the Purpose. The disclosing DAFD

Partner shall request the receiving member of DAFD consortium to treat the Confidential Information as confidential under the terms of the DAFD agreements.

4. Upon demand by the Party that disclosed the Confidential Information, all confidential information, including notes taken, documentation, models, guidelines, pictures or prototypes shall be returned to this Party.

5. No Party shall disclose the existence and character of this Agreement, nor the fact that the Confidential Information is evaluated. No Party shall use the name of the any Party in any publications or advertising materials without prior written consent of the other Party.

6. This Agreement shall come into effect on the day when the last Party signs (Effective Date) and shall remain in force for 5 years. The obligations contained in clauses 1, 2, 3 and 4 of this Agreement shall be continuing for five years from the date of disclosure or until the Confidential Information disclosed to receiving Party is no longer confidential, whichever period is shorter. The provisions of the clauses 6, 7, 8 and 9 shall survive the expiration of the term of the agreement.

7. This Agreement shall be binding upon, inure to the benefit of, and be enforceable by (a) DAFD Partner, its successors, and assigns; and (b) [COMPANY NAME], its successors and assigns.

8. If any of the provisions of this Agreement are found to be unenforceable, the remainder shall be enforced as fully as possible and the unenforceable provision(s) shall be deemed modified to the limited extent required to permit enforcement of the Agreement as a whole.

9. This Agreement states the entire agreement between the parties concerning the disclosure of Confidential Information and supersedes any prior agreements, understandings, or representations with respect thereto. Any addition or modification to this Agreement must be made in writing and signed by authorized representatives of both parties. Both Parties acknowledge that they have read and understand this Agreement and voluntarily accept the duties and obligations set forth herein

IN WITNESS WHEREOF, the parties have caused this Agreement to be duly signed by their respective duly authorized officers, on the dates and in the places indicated below.

Executed in two (2) original counterparts

AGREED AND ACCEPTED BY:

DAFD Partner   [COMPANY NAME]

Signed: _____   Signed: _____

Print name: _____   Print name: _____

Title: _____   Title: _____

Date: _____   Date: _____

Human Subjects: Consent to Participate in DAFD Research

CONSENT TO PARTICIPATE IN DAFD RESEARCH
The aim of the Data Audit Framework Development (DAFD) Project is to develop an online audit tool to enable higher and further education institutions to establish what research data they hold, where it is located and how it is being managed. To define the scope of the audit tool and identify system requirements, four pilot audits will be carried out. These pilot audits will take place at Glasgow University, Edinburgh University, King's College London and UKOLN at the University of Bath. You were selected as a possible participant in this study because you can provide an interesting viewpoint on how research data is being created and managed. Therefore you are asked to participate in this research activity.

PRIMARY CONTACTS
Prof Seamus Ross, HATII, George Service House, 11 University Gardens, Glasgow, G12 8QQ, Scotland, s.ross@hatii.arts.gla.ac.uk, +44-141-330-5512.

Sarah Jones, George Service House, 11 University Gardens, Glasgow, G12 8QQ, Scotland s.jones@hatii.arts.gla.ac.uk, +44-141-330-3549.

INDEPENDENT CONTACT
If you have any questions that you feel the interviewer has not or can not answer satisfactorily about the conduct of this research please contact Robin Rice, Edinburgh University Data Library, 2nd Floor, Main Library Building, George Square, Edinburgh EH8 9LJ, R.Rice@ed.ac.uk, +44-131-651-1431

PURPOSE OF THE RESEARCH
The research will involve conducting audits that identify research data assets, their current location and management. This will enable institutions to identify weaknesses in current data management and potential for exploitation of research data. The results of the initial audits will feed into the development of an online audit tool.

PROCEDURES
As you have tenatively agreed to participate in interviews, we are asking you to sign this document to give your formal assent to your participation.

Once we receive your form we shall arrange an interview date and time with you. This interview will last approximately [1] hour and will be conducted in person or by telephone. During the interview, the interviewer will ask you a series of questions relating to [research data you have created or are responsible for managing].

POTENTIAL RISKS AND DISCOMFORTS
Participation in this research does not pose any foreseeable risks or discomforts.

METHOD OF INTERVIEW CAPTURE
The interviewer will make notes during the interview (whether it is being conduct in person or by telephone) which will be used for research purposes only. This will form the basis of our record of the interview and be subject to analysis as part of the DAFD programme of research. It will only be accessed by the partners of the work package or previously identified project members whose research outputs depend on the work package's findings. Information from the interviews may be included in the Data Audit Framework registry is agreed by the researchers and department management.

POTENTIAL BENEFITS TO SUBJECTS
The primary benefits to participants will be the influence the results will provide to the DAFD project and their development of an online audit tool for research data.

POTENTIAL BENEFITS TO SOCIETY
There is general agreement that the cultural and scientific memory of our society should be managed in ways that will ensure it is accessible and understandable by future generations. For this to happen we require further knowledge of the way digital information is created and used in order to influence and engineer the technological methods being developed.

PAYMENT FOR PARTICIPATION
You will not receive any payment for participation.

PARTICIPATION AND WITHDRAWAL
Your participation is voluntary. If you decide to participate, you are free to withdraw your consent and discontinue participation at any time without penalty. You may also refuse to answer any questions during the interview which you do not want to answer. The investigator may terminate your participation in the study if circumstances arise which warrant doing so.

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law.

DETAILS OF INTERVIEWER:
If you have any questions or concerns about the research, please feel free to contact:

[INSERT NAME, TITLE, POSTAL ADDRESS, TELEPHONE NUMBER, AND EMAIL ADDRESS OF THE INTERVIEWER]

NAME, DATE, AND SIGNATURE OF INTERVIEW SUBJECT
I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

_____   _____
Name of Research Subject   Date

_____
Signature of Research Subject

Human Subjects: Consent for Interview Release

You recently participated in an Interview in support of the Data Audit Framework Development (DAFD) Project. *Attached to this form you will find a copy of the interview report that we have prepared for your approval.* (optional) We would be grateful if you could agree the level of attribution that we may ascribe to you as interviewee. This ranges from full anonymity to full acknowledgement.

PRIMARY CONTACTS

Prof Seamus Ross, HATII, George Service House, 11 University Gardens, Glasgow, G12 8QQ, Scotland, s.ross@hatii.arts.gla.ac.uk, +44-141-330-5512.

Sarah Jones, George Service House, 11 University Gardens, Glasgow, G12 8QQ, Scotland s.jones@hatii.arts.gla.ac.uk, +44-141-330-3549.

INDEPENDENT CONTACT

If you have any questions that you feel the interviewer has not or can not answer satisfactorily about the conduct of this research please contact Robin Rice, Edinburgh University Data Library**,** 2nd Floor, Main Library Building, George Square, Edinburgh EH8 9LJ, R.Rice@ed.ac.uk, +44-131-651-1431

DETAILS OF INTERVIEWER:

If you have any questions or concerns about the research, please feel free to contact:

[INSERT NAME, TITLE, POSTAL ADDRESS, TELEPHONE NUMBER, AND EMAIL ADDRESS OF THE INTERVIEWER]

AGREEMENT OF INTERVIEW DATA USE AND RELEASE

Place your signature next to the data release level that you are prepared to agree to and score through other levels:

Agreement Data Use and Release Level	**Research Subject Signature & Date**
I agree to the use of data collected through participation in this study if the data are fully anonymised. Any quotations of statements made by me must be both anonymised and paraphrased.	
I agree to the use of data collected through participation in this study if the data are fully anonymised. Any quotations of statements made by me must be anonymised.	
I agree to the use of data collected through participation in this study if the data are fully anonymised. Use of direct quotations of statements, however, is permissible provided I am given the opportunity to approve the text of any quotation after the interview and I am only identified by job title.	
I agree to the use of data collected through participation in this study and it need not be anonymised. Use of direct quotations of statements is permissible.	

NAME, DATE, AND SIGNATURE OF INTERVIEW SUBJECT

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

_____   _____
Name of Research Subject   Date

_____
Signature of Research Subject

b)   Proposed audit timetable for on-site activities

Data Audit Schedule: [*name of organisation being audited*], [*date*].

Data audit schedule for on-site activities					
Date	**Time**	**Activity**	**Location**	**Staff**	**Broad Issues Covered**
	c.4hrs	Review of what has already been achieved, discussion of the scope and purpose of the audit			■ Taking stock of the information already gained and discussing how to proceed ■ Defining the expected benefits of the audit
	c.3hr	Group discussion on how to conduct the audit			■ Discussion of staff availability, sources of information and how to communicate details of the audit to all staff members
	c.2hrs	Informing staff of the data audit			■ Communicating objectives of the data audit and level of assistance required from staff
	c.4hrs	Distributing questionnaires			■ Collating information to input into asset inventory
	c.5hrs	Identify and classify data assets			■ Adding to the current list of assets to supplement missing information, classifying as you go
	c.5hrs	Identify and classify data assets			■ Adding to the current list of assets to supplement missing information, classifying as you go
	c.5hrs	Identify and classify data assets			■ Adding to the current list of assets to supplement missing information, classifying as you go
	c.4hrs	Assess management of assets			■ Identify condition and current weaknesses in the curation of vital and important assets
	c.4hrs	Assess management of assets			■ Identify condition and current weaknesses in the curation of vital and important assets
	c.2hr	Meet with management to discuss audit findings			■ Discuss audit process, findings and suggested next steps to get management comments
	c.1hr	Release final report and recommendations for approval			■ Submit final report to organisational management and feedback audit findings to staff

10) Appendix 3: Guidance documents

a)   Example email approach to target organisation

Dear ,

Seamus Ross suggested contacting you to invite participation in a JISC-funded project we're running in HATII. The Data Audit Framework Development project will create a fully-validated online tool that enables institutions to identify data holdings, ascertain their location and condition, and improve ongoing data management. As part of the development process we hope to validate our methodology through case studies in a range of University departments and wondered whether the archaeology department, perhaps through GUARD, would be willing to take part in this work.

The case study would involve a member of HATII staff coming in to the department to audit data holdings. We anticipate this would take between 3-5 days and would hope to find a convenient time in May for the audit to take place. In order to complete the study the auditor would need to be permitted access to internal documents, technologies and systems (a confidentiality agreement would be signed beforehand) and an agreed level of archaeology / GUARD staff time should be set aside for completion of questionnaires and interviews on departmental data holdings.

Being involved in the Data Audit Framework Development project would assist the archaeology department to recognise the wealth and quality of data being created by its research staff. The audit will also identify critical assets and those at risk of loss, providing professional recommendations for more efficient data management. Data assets have immense potential both financially and in terms of boosting research reputation; departments in control of their data assets are empowered to exploit this potential.

We are keen to work with you and believe involvement in the DAFD project would be of mutual benefit. Seamus and I are very happy to accept any questions you have regarding the proposal and can be reached at the details below.

We look forward to hearing your response

Kind regards

Seamus Ross and Sarah Jones

b)   Formula for establishing how much time is needed to complete a data audit

In order to help institutions establish how much time needs to be allocated for conducting a data audit within departments and across the entire institution, a basic formula has been created that assesses the level of effort contributed by the auditor and the organisation's staff. The formula is based on DAF audits taking place at departmental level, but a coefficient of complexity is suggested to get a clearer picture of how long data audits would take across departments or in units with large volumes of data or complex data types.

The formula has four parts: evaluating the auditor's effort; evaluating the departmental effort; calculating the total effort per department; and evaluating the total institutional effort.

1. Evaluation of the auditor effort

The auditor's effort is based on the number of days needed to complete the audit. The DAF development team expect this will be betweeen 11-17 days for a departmental or unit audit with around half of this time sepnt on-site. Five hours is taken as the average number worked daily.

a – estimation of auditor's time contribution based on 5 hours per day
d – number of days expected to complete the audit

$$a = d \times 5$$

2. Evaluation of the department effort

The evaluation of departmental effort is based on the number of staff. Around ¼ of the staff members are expected to be actively involved in data management with the remaining staff focusing on other activities. The majority of staff will provide a nominal amount of time for the audit, most likely by completing questionnaires to assist with identifying data assets. In the case of large departments or schools it may not be feasible to speak to ¼ of all staff so the scope could be reduced.

h – estimation of the staff time for a department which has n (number of staff)
n – number of staff of which n_d is the number of staff involved in data management who will provide roughly 5 hours each and n_g is the remaining number of staff involved for roughly 1 hour each.

Our practical experience shows that usually ¼ of the staff members are involved in data management, either creating or curating assets, and ¾ in other activities, or

$$n_d = \frac{1}{4} \times n,$$
$$n_g = \frac{3}{4} \times n,$$
$$n_d + n_g = n$$

The estimation of the number of hours h which the department staff will contribute to the audit based on a standard departmental audit is

$$h = n_d \times 5 + n_g$$

In the real world, departments and their data collections are different. In order to provide a more accurate estimation we should take into account a coefficient of complexity c which is based on the size of the department, the extent of data holdings and their composition in terms of file formats, structure and sizes.

$$h = (n_d \times 5 + n_g) \times c$$

Currently we expect large departments or units with more complex data holdings to use a coefficient of around 3. Data collected from initial audits will help to establish a more accurate coefficient measurement.

3. Total effort per department

The total effort required for the audit v is the sum of the auditor's and departmental effort

$$v = a + h$$

4. Evaluation of the total institutional effort

The sum of efforts of all departments $v_1, \dots v_k$ where k is the total number of departments will provide an evaluational of the total effort required to complete an institution-wide audit t. In the case of instituion wide audits a decision may be taken to only audit a representative sample of departments rather than the entire institution, or to vary the level of granularity (hence also effort) used in each departmental audit. These decisions should be refelected in the figures input.

$$t = \sum_{i=1}^{k} v_i$$

A practical example

Let us consider the case of an standard departmental audit. If we expect the auditor to spend 14 days the effort required will be:

$$a = 14 \times 5 = 70 \text{ hours}$$

The departmental effort based on an average sized university department with 20 staff will be:

$$n_d = \frac{1}{4} \times 20 = 5, \text{ and}$$
$$n_g = \frac{3}{4} \times 20 = 15$$

$$h = 5 \times 5 + 15 = 40 \text{ hours}$$

The total effort for the departmental audit, assuming the department does not have particularly large or complex data holdings will be:

$$v = 70 + 40 = 110 \text{ hours}$$

c)   Example business case

The UK Archaeological Service (UKAS) provides a full range of archaeological services for commercial developers, central and local government, public utilities and private individuals. In the course of this work large and diverse datasets are created that the Service is required to curate. UKAS currently has a significant digital data backlog so needs to urgently review and improve on its data management systems. Information on the extent, location and condition of data assets is lacking so the Service is not yet in a position to make recommendations for change.

Using the Data Audit Framework would provide UKAS with an established method of identifying data assets, evaluating data quality, and recognising pitfalls in current curation practices. Producing an inventory of assets and gathering information on data management practices will allow the Service to take action to reduce the risks associated with data loss. This is crucial as UKAS is a commercial enterprise so failure to adequately curate data holdings may result in litigation or a loss of revenue. Improvements to the present data management systems are required as a backlog has developed. Reviewing processes will identify areas where a more efficient workflow could be introduced and indicate areas where investment in new systems would be beneficial.

A data audit should take between 11-17 days to complete. An internal auditor or audit team could be appointed to avoid consultant fees. Staff time will also need to be set aside for the completion of surveys and interviews and to collect together relevant documentation to pass to the auditor. Using the formula in Appendix 11b an estimation of the overall cost in man hours is 110 hours, of which around 70 hours will be spent by the auditor and the remaining 40 hours by staff within the organisation. As UKAS has 20 staff, this equates to most staff setting aside 1 hour for the audit and the 5 main staff responsible for data collections providing about 5 hours each.

If we take an average hourly rate to be £25, the departmental data audit would cost in the region of £2,750 to complete. Additional costs are likely to be required to implement change, however the knowledge needed to be in a position to make such change requires a minimal investment of time and effort. The risks UKAS face by continuing with the present inadequate data management system could easily amount to more than £2,750. The cost of litigation would undoubtedly be higher. Potential damage to the Service's reputation and the likely loss of revenue that would follow are also likely to outweigh this initial cost. It is therefore recommended that UKAS adopt the DAF methodology to conduct an audit of their data holdings and curation practices.

d)   Lessons learned reports

These lessons learned reports provide feedback on the initial three implementations of the DAF methodology. The three audits were conducted within three different HE institutions in schools and departments of varying size and research area. Each took place between May-July 2008.

Engineering at University of Bath by Alex Ball

1 Background

In June and July 2008, an audit of research data assets held by the IdMRC at Bath University was conducted using the Data Audit Framework (DAF) methodology. This report briefly describes how the audit was carried out and the lessons learned from it.

The Innovative Design and Manufacturing Research Centre (IdMRC) is a research group within the Department of Mechanical Engineering at the University of Bath. It was set up in October 2001 with funding from the EPSRC's IMRC programme, and is one of sixteen such centres in the UK. It has four research themes: Advanced Machining Processes and Systems (AMPS), Constraint-Based Design and Optimization (CBDO), Design Information and Knowledge (DIAK), and Metrology and Assembly Systems and Technologies (MAST). The IdMRC's work is widely supported by industry, especially from the aerospace and packaging sectors and with emerging strengths in shoe and electronics manufacture.

2 Performing the audit
Stage 1: Planning the audit
An initial phone interview was held with the Director of the IdMRC to establish the scope, purpose and requirements for the audit. The IdMRC website was used to compile a list of staff and to clarify points about the history, structure and academic interests of the Centre; interviews were scheduled with the representatives (lead researchers) for the four research 'themes'. The IT Administrator was contacted about accessing shared drives.

Stages 2 & 3: Identifying and classifying data assets
A snowball sampling technique was used to choose interviewees, starting with the four theme representatives. In all, ten face-to-face interviews were conducted in this pattern. The interviews consisted of:1. going through the interviewee's personal drives (and, where appropriate, shared drives) and determining which collections of data constituted data assets; 2. recording names, descriptions, statements of responsibility and locations; 3. discussing the importance of the asset in terms of current and future research; 4. recording additional information about file formats, software requirements, derived reports/ papers, dates of creation and update, etc.; 5. discussing how the interviewee managed the data. The remaining 17 members of the Centre were contacted by e-mail with a questionnaire similar to Audit Form2. This resulted in three completed questionnaires, two e-mail responses and one telephone interview. The resulting inventory consisted of 63 data sets, of which 18 are vital, 15 are important and 30 are minor. This was not comprehensive but was representative of the data assets of the Centre.

Stage 4: Assessing the management of data assets
Of the data assets in the inventory, 30 were chosen for analysis: the vital and important assets, less the three held by an external organization. Much of the information required for this stage had already been collected, with just a few gaps which were filled by e-mail queries. The basic metadata set was used in all cases.

3 Lessons learned

3.1 Time
Be prepared to 'badger' senior management. At various points the audit methodology calls for contact with the organization's management, who are among the busiest people in the organization. A certain amount of persistence is required to keep the actual elapsed times in line with the estimates in the methodology. I found that e-mails to arrange phone calls worked better than long e-mails, and that phone calls were easier to arrange than face-to-face meetings.

Picking your moment. I found that attempting to perform the audit in June meant that academics were rather hard to get hold of (due to exams boards), but I had no trouble arranging meetings with research staff.

Choosing a sampling technique. I tried snowball sampling, starting with research 'theme' leaders. While I found it an effective way to get a good range of results, it did have a cost in terms of elapsed time, due to the fact that interviews were arranged on a rolling basis, instead of all at once, well in advance. I also tried filling in gaps with a questionnaire, but as predicted these had a low response rate.

3.2 Gathering information
Deciding on scope and granularity. The thing I had most trouble with was determining and communicating the scope of the audit. Very few of the research data sets held by the IdMRC were straightforward data bases or homogeneous data sets; in the main they tended to be ad hoc collections of resources supporting particular pieces of work. This made it hard to communicate the scope of the study in a way that would include, say, a set of a company's internal communications (used to analyse information flow within that company), while excluding collections of relevant literature accumulated by researchers in the course of their work. It also made it hard determining a suitable granularity at which to record the data assets. For example, two of the 'themes' are engaged in consultancy work, with each consultancy generating a small set of documents and data. On the one hand, each of these sets has its own character, life cycle and confidentiality conditions, and so could be treated individually, but on the other hand, taking them all together as one asset enables one to see better the overall character of the data and how the asset relates to other assets, and makes it considerably easier to make a comprehensive statement of holdings.

What information to ask for. I found that it saved more time than it wasted, to try and do as much information gathering in one go as possible. Instead of just getting a name, description and owner of a data asset in the initial interview, I also discussed the value of the data asset, and collected information such as creation dates, updating frequency, locations, formats and related publications. This meant I could just fill in any gaps later by means of a quick e-mail rather than having to reschedule an interview.

GeoSciences at University of Edinburgh by Cuna Ekmekcioglu

Background and context
In May 2008 an audit of data assets was conducted in the School of GeoSciences using the Data Audit Framework (DAF) methodology developed by DAF Development Team.

School of GeoSciences is a leading international centre for research into GeoSciences, with some 80 academics, 70 research fellows and 130 PhD students, and an annual research grant and contract income of around £4-6 million. In the last UK Research Assessment exercise, the School was rated as internationally competitive, receiving the top grade of 5/5* for its research. The School Staff

contribute to one or more of five Research Groups (Earth Subsurface Science, Global Change, Human Geography, Edinburgh Earth Observatory, Centre for Environmental Change & Sustainability) and may be involved in inter-University Research Consortia and Research Centres.

This report briefly describes how the audit was conducted and the lessons learned from conducting this audit.

The Audit
Stage 1: Planning the audit
This stage involved desk research such as browsing the School website, collecting annual reports and published articles, compiling a list of research active staff with their research responsibilities. Following on from this preparatory work, an initial meeting was arranged with the IT managers of the School to discuss how best the audit could be conducted and whether we could have access to the shared drives. Based on the information gathered from the desk research and the interview with the IT managers, the key staff were identified and invited to provide information about their research and research data via semi structured interviews.

Stages 2 & 3: Identifying and classifying data assets
In these 2 stages interviews were conducted with 35 academic/research staff, and an inventory of 25 data assets was created. The interviews were in the form of semi structured discussions to allow us gather as much information as we could such as data types, size of the collection, software used for analysis, value, storage, back-up, and retention of the data assets. Although this was not a comprehensive survey, the later interviews started to provide information already collected, suggesting the most important data assets had been recorded. Results of the pilot were reported back to the development team at a meeting in Glasgow at the end of May.

Stage 4: Assessing the management of data assets
Of the total 25 data assets only 4 of them were classified as vital assets and the detailed analysis of these assets was carried out buy the auditor contacting the owner of the data assets as and when necessary. It was rather difficult to complete all the metadata fields in audit form 3.

Stage 5: Reporting results and making recommendations
Generally speaking, the audit was useful to identify the gaps and issues in managing data assets in the School. Staff comments and suggestions for improvement of data management were found very useful. The results of the audit were drawn together and a final report was produced to recommend actions for change.

Lessons learned
Time
Time is one of the most important factors in conducting the audit successfully. If time was not a restriction we could have obtained better results for the audit. The planning stage should have been done well in advance, and the key staff should have been contacted at least a couple of weeks before the interview dates. In our case, most staff were out on field trips, or busy with marking exam papers and sitting in exam boards. Although 65 staff were contacted via email for interviews we could only interview 35 of them as the rest were not available for interviews until late June. The interviews themselves were time consuming. Ideally, an online survey could have been circulated to gather general information and then this could have been followed up by detailed interviews. As we did not have enough time to follow this approach we had to rely only on interviews.

Access to information
Access to the shared drives was denied on the basis of data protection. We could have accessed the drives if we contacted every research staff and obtain their written permission for access. However,

this could have delayed the audit for another couple of months. Also, a great majority of the data assets was held on external hard disks, personal PCs and laptops, USB storage devices and CDs/DVDs. Since we could not access either the shared drives or data held on personal storage devices, the audit was conducted on project basis recording the data assets for the projects that were mentioned by staff in the interviews.

Scope and level of granularity
We had trouble with determining the scope and the level of granularity of the audit. We managed to identify only a couple of well described databases. The rest of the data assets we recorded were collection of text files, audio files, images etc. that were used in a particular research.

Documentation
We had difficulty with locating and accessing the documentation where data assets were described. There was no main index or catalogue of the data held in the shared drives – users were expected to search for the required data themselves, or with guidance from the relevant research staff. Searching for the data was difficult as most of the data was undocumented and there was not a well defined folder structure. Data was generally stored using a sub-folder file system where a variety of schemes (by year, by location, by relevant field trip etc.) were used. It was difficult to identify data from the file names only as there was no standard file conventions used while the data was stored.

Archaeology at University of Glasgow by Sarah Jones

Background
From May to July 2008 an audit of data holdings within GUARD at the University of Glasgow was conducted using the Data Audit Framework. This report considers how the methodology was implemented and reports back on lessons learned.

GUARD (Glasgow University Archaeological Research Division) is the archaeological practice within the Department of Archaeology. The unit was founded in 1989 and currently has 33 members of staff. It is a commercial provider offering a wide range of archaeological services, from consultation to fieldwork and post-excavation analysis. GUARD serves a range of clients including commercial developers, central and local government, public utilities and private individuals. Contracts are predominantly based in Scotland though staff also work in England, Ireland, Northern Ireland and further a field.

Summary of progress by stage
Planning the Audit: The initial planning stages were fairly smooth. The Director of GUARD was already aware of data issues within the Unit and keen to take part so it was not necessary to develop a full business plan. An initial meeting took place in early May to determine the scope of the audit and identify expectations. This gave the audit a clear focus and helped direct the initial stages of work. Access to the departmental shared drives was agreed which allowed research into the staff and work of the unit to be conducted. The information provided on the shared drive and website was ample to understand the context in which the Unit operated.

An email explaining the audit was sent to the Director to distribute to staff. This was thought sufficient, however when contacting staff during stages 2 and 3 they weren't always familiar with the work. In light of this a second interview was set up with the Unit's archivist to help gain some internal advocacy. This meeting proved very useful. The archivist was struggling with the Unit's data issues so was enthusiastic about the audit and introduced the auditor to staff members to encourage their participation.

Identifying and Classifying Data Assets: A few interviews were requested to assist in identifying assets but the response rate was low. This was probably for a few reasons: the interview requests were generic rather then being targeted at specific areas of expertise or collections; staff availability was minimal as the Unit's work is largely conducted off-site; and people seemed unfamiliar with the audit. As access to the shared drives had been agreed however, the inventory could largely be created through desk-based research. The scope had been set as all work from the past three years excluding the forensic material due to sensitivities. After a preliminary survey a decision was made to record assets by project. GUARD conducts around 60 projects annually, each generally creating a small amount of data of a number of common types. As such, the inventory quickly became very large and entries were fairly consistent in terms of the type and size of data being created. In light of this a decision was made to complete a comprehensive survey for the first half of the scope and a sample for the 2005-06 period, picking out especially large, multi-stage projects or ones with unusual funding sources, data types or subjects. This made the best use of time as effort was not spent replicating work yet still ensured the range of data assets was represented to allow all data issues to be investigated in the next stage.

The suggested classification was amended slightly to suit GUARD's data assets. While it could be argued that all projects are vital as the nature of archaeology means it would not be possible to reconstruct the data if lost, a classification was desirable to help decide which assets to analyse in greater detail in the next stage. It would have been unsuitable to categorise all active collections as vital as work was still ongoing on 42 collections from an inventory of 65. Instead classification was based on the revenue a project generated, since GUARD is a commercial unit, and whether it was still active or part of another piece of work. This acknowledged the trend for projects to be continued through subsequent pieces of work and the fact that the Unit was not responsible for long-term preservation as completed projects could be archived with RCAHMS.

Assessing the Management of Data Assets: The detailed analysis of vital collections was conducted by interview. As much of audit form 3 as possible was completed in advance so this could simply be verified and enhanced in the interviews. Although it took some time to schedule the interviews, response rates were higher as most of the staff approached had already been introduced to the auditor and the requests were more specific, focusing on a particular data assets they had been involved in. Some general questions on how the member of staff created, managed and used data was asked at the start of the interview. This helped to build rapport and provided a useful overview of the Unit's approach to data that helped guide the interviews.

Final reporting and recommendations: The interviews were very useful for seeing how the Unit created and managed data and identified areas for improvement. Staff were also very open with suggestions of what issues they faced and changes they felt necessary. These aspects helped feed into recommendations for change to improve workflows and minimise the risk of data loss and corruption.

Lessons learned
Timing is key – When investigating the organisational context it would be helpful to consider when is the best time to conduct the audit. Staff in GUARD are often out on fieldwork and as the audit was in the summer period annual leave also affected availability. It was anticipated the audit would be completed in May but delays in setting up interviews extended this to July, meaning it ran in parallel with other work the auditor was conducting, adding to the delays. Where possible an extended period of elapsed time where the auditor's diary is clear should be allowed. Otherwise planning should be moved forward to try to schedule main work in a short time period.

Scope the work carefully – The initial meeting with the Director of GUARD was very useful to scope the audit and identify expectations to ensure the audit delivered. The scope and level of granularity adopted should be flexible as it may be necessary to amend these during the audit.

Get internal advocacy – An email may not be sufficient to inform staff of the work and encourage participation. Attending a staff meeting, obtaining a personal introduction or securing some internal advocacy may be more successful methods. Involving core members of staff who are responsible for data management, such as the archivist, can help ensure the department accepts ownership of the audit results and take the recommendations on board.

Make best use of staff time - Agreeing access to internal documents is preferable as more of the research in the initial stages can be conducted off-site, thereby limit the demands placed on organisational staff in terms of questionnaires and interviews. It will also allow additional information to be collected in the early stages so time can be optimised in the interviews and discussion can focus on day-to-day working practices to see how data is being managed.

Centre for Computing in the Humanities (CCH) at King's College London by Stephen Grace

Background
Stephen Grace at the Centre for e-Research (CeRch) undertook a case study audit of the Centre for Computing in the Humanities (CCH) at King's College London (King's) during October-December 2008. He was helped by having the experience of the other audits undertaken in Glasgow, Edinburgh and Bath universities and thanks Sarah Jones, Cuna Ekmekcioglu and Alex Ball for their insights.

CeRch is a new research centre with a broad remit to work across discipline areas at the intersection between research methods and practice, digital informatics and e-infrastructure. It was established in October 2007 and launched in April 2008. It was based on the experience of the Arts and Humanities Data Service executive and the AHRC ICT Methods Network

Four approaches to other departments (in the School of Medicine) were made by Sheila Anderson, Director of CeRch and by Stephen Grace. These were unsuccessful for different reasons, and CCH was approached to help because of long-standing good relations with CeRch. CCH is a specialist research centre with an international reputation in the application of technology in research in the arts, humanities and social sciences. It has a teaching programme, but its primary focus is research activity which culminates in making digital resources available. The research projects, and their digital assets, are critical to the mission of CCH.

Summary of progress by stage
Each of the stages of the audit is taken in turn. The DAFD project decided to combine the Identifying and Classifying stages, and this case study treated these as a single stage.

Planning the audit
The original intention of the King's work package was to audit a department in the School of Medicine. One had been approached informally by Sheila Anderson before the DAFD project and expressed interest at that stage; it declined an invitation because of timing, as did another department. A third declined on the grounds that it "had no data" (it was a new research institute) and the fourth did not respond. This process was protracted and frustrating to the auditor and the wider DAFD project. CeRch was only recently established when the invitations were made, and its competence to assess data management may not have been clear to departments.

CCH was then approached in the person of the Research Fellow with overall responsibility for data management (DM), and he was willing to take part with the ready agreement of his Director. CCH is adjacent to CeRch and there are good relations between staff in the two centres. In addition, they

share a System Administrator and this made the task of gaining access to servers very straightforward. User permissions were granted, and the auditor was able to gain access from his desktop within a few hours of agreeing the work with the Research Fellow. Travelling between departments was much reduced compared to the Edinburgh case study, for instance.

The Research Fellow was interviewed on data management practice and infrastructure (including plans and aspirations for the future). This gave the auditor a good general understanding of the distinctive culture of data management at CCH. A Research Associate was identified who could speak across a range of projects of his practical experience in managing research data.

Identifying and classifying data assets

Because of the delay in starting, it was decided to save time by scoping the data audit more narrowly than the whole department. At least fifty-six projects are listed on the CCH website, and the consistent DM practice identified at stage one suggested (as did feedback from the other DAFD case studies) that sampling would elicit enough evidence. Twelve projects were selected for audit, four each from the list of completed, stage two and current projects. A Research Associate with responsibility for data in four projects was interviewed about his data management practice.

CCH maintains an online list of its projects (overwhelmingly of curated digital assets), and this provided much evidence for Stage Two. Access to the CCH servers and directories made it simple to identify assets, but the process of collating and uploading information to the online tool is slow. CCH organises its servers into project directories and typically encompassing three sets of assets

Digitised assets (images, digital texts, sound, etc)

Marked-up copies of text in XML

Files and scripts needed to render webpages

These hierarchies are established at the outset of a project, and helped the auditor in identifying and making sense of the data.

Assets were classified by project using the standard DAFD schema. Most were considered "vital" since the project websites were publically available or the resource was still being compiled.

Assessing the management of data assets

Because CCH has a coherent data management practice, it is easy to understand for each project where the digital assets exist and what forms they take. The Research Fellow establishes server and directory requirements at the outset, and user permissions limit the ability of an individual to alter standard practice in the centre. Researchers are aware of their responsibility to manage data, committed as they are to seeing the fruits of their work reach a wide audience. A second interview was held with the Research Fellow to gather more information and share preliminary assessments.

Much of the Assets register in Stage 3 (and Audit Form 3) was populated by the same sources of information used for Stage 2, especially the project descriptions on CCH's website. These two stages may effectively be undertaken in parallel. It was time-consuming to enter the data online, and in future it may make sense to create a spreadsheet to hold the data compiled during an audit.

Final reporting and recommendations

The online tool generated a technical appendix and template for the final report which helped to reinforce the professional nature of the audit. These were delivered to CCH in January, and a debriefing interview was arranged.

Lessons learned

Identifying a department for audit
CeRch had four false starts before finding a willing audit partner. Partly this was because the centre itself was newly established, and maybe needed to establish its credibility across King's. There was no central endorsement for the data audit project, which may have put the work in the context of other initiatives at King's to improve infrastructure and support for research. A couple of departments found the initial timing of the audit in conflict with their work plans at the end of the academic year. The PI-DAF project at King's will ensure that the benefits to the department are made explicit.

The final approach to CCH worked at least in part because the invitation was not sent in the first instance to the head of department but to someone known to have a role in data management. This approach will be used in the King's PI-DAF project where known contacts in departments will be approached. With their support, the head of department will be asked to consent to the audit, approve any permissions (such as for server access) and confidentiality requirements. The DAFD publicity leaflet helped in explaining the audit process and the benefits of participating for researchers.

Organising time
There were no major problems with arranging interviews for the small cohort in CCH, although a wider schedule may have offered depth to the findings. The good relations between CCH and CeRch (and sharing a System Administrator) eased access to the private networks of CCH: granting this permission may be less willing, and involve more administrative burdens, in other departmental settings. It is critical to budget sufficient time, including lead time, in this as with arranging staff interviews.

Collating data and logging it on the online tool was time-consuming, even with a reduced sample of data. If the Data Audit Framework is to be widely used, it is essential every opportunity is made to speed up the process of the audit. This may be by importing data from a spreadsheet compiled by the auditor, or by delegating the collation task to others (see 4 below).

Need for documentation when using tool
The tool presumed the information is to hand when entering data for Stages 2 and 3. It would help if this data (collected from annual reports, documentation, web pages, etc) could be uploaded as a record of the evidence used by the auditor.

Availability of tool to manage audit process
The online tool is a useful way to manage the audit, and it could be enhanced to manage the full audit process. Interview dates, collation of survey information into Audit Forms 2 and 3, actions on recommendations, dates of reviews or follow-up audits could all be accommodated in a tool for an auditor (or team of auditors). In a devolved organisation like King's it is easy to imagine ways that the whole process may be undertaken by a range of actors – from a graduate student collating Form 2 in a single department, to a Records Manager overseeing the College's compliance with data security issues. Some of these issues are explored in the Scenario Test document compiled by CeRch for the DAFD team.