

Project Acronym: Edinburgh DAF
Version: 2.0
Contact: Çuna Ekmekcioglu, cekmekci@staffmail.ed.ac.uk
Date: 28/01/09

Edinburgh Data Audit Implementation Project

Final Report

DRAFT

Çuna Ekmekcioglu, Robin Rice

28 January 2009

Contact: Çuna Ekmekcioglu, cekmekci@staffmail.ed.ac.uk



JISC Final Report

Project

Project Acronym	Edinburgh DAF	Project ID	
Project Title	Edinburgh Data Audit Framework Implementation		
Start Date	1 st May 2008 (originally 1 st March 2008)	End Date	1 December 2008 (originally 31 August)
Lead Institution	The University of Edinburgh		
Project Director	Robin Rice R.Rice@ed.ac.uk		
Project Manager & contact details	Dr Cuna Ekmekcioglu cekmekci@staffmail.ed.ac.uk Tel. 0131 650 5108		
Partner Institutions			
Project Web URL	http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2007/dataauditframeworkpilots.aspx		
Programme Name (and number)	Digital Repositories programme 2007-8		
Programme Manager	Neil Jacobs		

Document

Document Title	<i>Project Final Report</i>		
Reporting Period	1 st May – 1 st December 2008		
Author(s) & project role	Dr Cuna Ekmekcioglu Project Manager		
Date	28/01/2009 version 2.0	Filename	edinburghDAFfinalreport1.pdf
URL			
Access	<input checked="" type="checkbox"/> Project and JISC internal		<input type="checkbox"/> General dissemination



JISC Final Report

Table of Contents

Acknowledgements	4
Executive Summary	5
Background	6
Aims and objectives	7
Methodology	7
Implementation	8
Output and results	8
Outcomes	9
Conclusions	9
Implications	10
Recommendations	10
References	11
Appendix A: Data Audit Framework: Edinburgh GeoSciences Pilot	12
Appendix B: Data Audit Framework: Edinburgh case studies	14
Centre for Integrative Physiology	17
School of Divinity	19
Economic and Social History	21
SFC Brain Imaging	23
Institute for Astronomy	25
Appendix C: Online Questionnaire	27
Appendix D: Project Steering Group members list	29

Acknowledgements

The Edinburgh DAF Implementation project was funded by the JISC through a grant from its Repositories Programme. We are grateful to our Steering Committee and MRC Human Genetics Unit for their contributions and DCC/HATII at the University of Glasgow, particularly Sarah Jones, for their continuous support. Thanks also go to all our key contacts who helped us access research active staff in their groups as well as all research staff who took time to complete the online questionnaire and provided detailed information at the interviews.

Executive Summary

The Data Audit Framework (DAF) Implementation projects – along with the Data Audit Framework Development project (DAFD) led by DCC/HATII - were conceived in response to recommendations made by Liz Lyon in the seminal JISC-commissioned report *Dealing with Data* (2007): “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 DAFD project developed the methodology starting in April, 2008, and the Edinburgh project started in May 2008, as one of four JISC-funded projects to test the framework through implementation. The Edinburgh project finished end of November, 2008.

The aim of the project was to provide exemplars of Data Audit Framework adapted to the current needs of data curation activities in University of Edinburgh, and to find what data the research community has, where it is located, and who is (or is not) responsible for it.

In order to be able to clearly address the issues in managing research data assets within the University we set out to implement the Data Audit Framework across three colleges, therefore a range of disciplines.

Our overall approach consisted of an online questionnaire and semi-structured interviews with research active staff across three colleges. In terms of the methodology we used the spreadsheet-based Data Audit Framework methodology developed by the DAFD project team in advance of the online tool. The methodology was implemented in four stages:

1. Planning the audit
2. Identifying and classifying assets
3. Assessing management of data assets
4. Reporting and recommendations

The key outputs from the project were five case studies – the final ‘audit’ reports for each research unit audited. These were the School of Divinity, Economic and Social History, Centre for Integrative Physiology, Brain Imaging, and the Institute for Astronomy. None of the audits was a comprehensive survey of research data but they proved to be a good starting point to auditing research data holdings and investigating data management practices within the University.

The project achieved its aim in terms of providing exemplars of Data Audit Framework activity in a UKHE institution. In addition to carrying out the data audits and writing up the case studies, project staff engaged in a number of dissemination activities such as speaking at the DAFD launch and other events, writing newsletter and journal articles, and writing the Lessons Learned document. The interest in this activity can be seen by the great volume of questions that accompanied the project manager’s presentations about ‘lessons learned’ at two events (lasting over twenty minutes in both cases).

While further awareness-raising with the University and UKHE more generally is still important, staff require pragmatic assistance in the form of guidance on best practice, research unit or school procedures, College or University-wide infrastructure and policy, and identifiable forms of support for data curation in the form of expert support staff, web pages, and discipline-specific guidelines, as well as short, focused, training opportunities.

Fuzziness persists, in terms of defining the data types and currency of data to be audited, and particularly in identifying the locus of responsibility for further curation beyond the goals of using data as a means to the end of publication. If those conducting the audits are not considered to be the data owners, then there is a question about who the audit is being done *for*.

The DCC should continue to support the online DAF tool, and support and development for this should continue to be funded by JISC. Take-up should be monitored (there is evidence of interest from the US) and coupled with further strategic and pragmatic efforts to build infrastructure and support for data curation in UKHE.

Background

The Data Audit Framework (DAF) Implementation projects – along with the Data Audit Framework Development project (DAFD) led by DCC/HATII - were conceived in 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 DAFD project developed the methodology starting in April, 2008, and the Edinburgh project started in May 2008, as one of four JISC-funded projects to test the framework through implementation. The Edinburgh project finished end of November, 2008.

Edinburgh is a research led University; for example, the University won a record £210 million to fund its research in 2006/07. Research data are generated by individuals and research groups in all 21 Schools of the three Colleges, spanning a very wide range of disciplines in the Humanities and Social Sciences, Medicine and Veterinary Medicine, and in Science and Engineering. There is an increasing trend for research groups to work as part of global collaborations which create substantial amounts of data that need to be accessible to researchers based in different institutions. Research funding agencies now usually stipulate that there should be a plan for sharing the data collected during the lifetime of a research project but they allow no provision for the cost of preserving this data after the end of the contract period.

The University's Information Services (IS) was keen to carry out this project for a number of converging reasons. First, the Data Library's DISC-UK DataShare project¹ (which had been developing a local data repository to operate in tandem with the open access publications repository, ERA) was ongoing during the period. A number of lessons had been learned about engaging researchers early in the research life-cycle (Green, 2007), and the importance of addressing potential depositors' actual needs as well as the less tangible goal of opening up access to research data for the sake of it (Gibbs, 2007). The DAF project offered a chance to work closely with researchers to determine what data had been created in various parts of the University, as a stepping stone to exploring further needs for curating, preserving, or sharing that data as determined by the researchers themselves.

Second, in line with the University's Research Computing Strategy² a similar but wide-ranging survey on research computing needs had been conducted by IS over the summer and autumn of 2007 with a good response. The survey results revealed a number of areas where researchers had demands that were not being fully met, in terms of hardware, software and support. An emphasis on curation echoed the University's overall mission which is “the creation, curation, and dissemination of knowledge.”³

Third, the College of Science and Engineering's IT committee, in cooperation with IS' Academic Liaison Director to the College, had identified the need for a data storage strategy for the College, and the DAF project was seen as a useful vehicle to progress that goal.

Since the locus for conducting data audits within the university was not clear, a broad steering committee was formed as a stakeholder group with representation from: EDINA and the Data Library, Edinburgh Compute and Data Facility, the Digital Curation Centre, the Library, the University Archives (all part of IS), as well as Records Management, the research office (Edinburgh Research and Innovation), and data 'champions' from each of the three colleges.

¹ <http://www.disc-uk.ac.uk/datashare.html>

² <http://www.is.ed.ac.uk/plans>

³ <http://www.ed.ac.uk/about/mission>

Aims and Objectives

The aim of the project was to provide exemplars of Data Audit Framework adapted to the current needs of data curation activities in University of Edinburgh, and to find what data the research community has, where it is located, and who is (or is not) responsible for it.

Specific objectives of the project were to:

- provide exemplars of data audit in research centres and or schools in each of the three Colleges and therefore across a range of disciplines;
- produce and disseminate findings - inform central and school management about the organisational and technical issues associated with the management of research data assets;
- provide guidance documentation and training geared toward discipline-specific and cross-disciplinary data creators;
- make recommendations for school-based and central policy development,
- identify researchers who wish to share their data through the DSpace repository or through web services or 'mash-ups'.

There was no time to carry out the third objective during the short intensive project; this will be done as a follow-up activity after the end of the project, as part of the DataShare project activity and with the input/direction of the existing steering group.

Methodology

In order to be able to clearly address the issues in managing research data assets within the University we set out to implement the Data Audit Framework across three colleges, therefore a range of disciplines. We aimed at examining current practice and identifying what data the research community has, where it is held, and who is (or is not) responsible for it. A decision was made between the Data Audit Framework Development project management and JISC to restrict the scope of the Data Audit Framework to research data assets. As such our audits were not concerned with administrative assets, such as student databases, or inventories of research outputs, nor publications and research papers or web resources.

Our overall approach consisted of an online questionnaire and semi-structured interviews with research active staff across three colleges.

In terms of the methodology we used the spreadsheet-based Data Audit Framework methodology developed by the DAFD project team in advance of the online tool. The methodology was implemented in four stages:

1. Planning the audit
2. Identifying and classifying assets
3. Assessing management of data assets
4. Reporting and recommendations

In the planning stage we defined the purpose and scope of the audit, conducted preliminary research and scheduled meetings with key contacts such as the IT officers and Heads of Schools. An online questionnaire was circulated to research active staff via these key contacts. When we could not get a high response rate for the questionnaire we also conducted semi-structured interviews with members of these groups.

In the second stage and third stages we classified the assets and identified weaknesses in data policy, and current data creation and curation practices. The final stage of the methodology was to report back to the schools with some recommendations to improve existing data management practices.

Implementation

Since our project work depended on the audit methodology developed by the DAFD project team in April and May we could not begin conducting the audits as soon as the project started. It was early June when our work began with implementing the spreadsheet-based methodology. The timeframe for the project was quite short, and so we aimed to work with five schools across three colleges. Initially our chosen sample consisted of the School of Biological Sciences, Institute for Astronomy, Engineering & Electronics, The School of Education, and Brain Imaging. However, we had difficulty gaining sufficient participation for three of the larger units: School of Biological Sciences, Education, and Engineering & Electronics, so we needed to replace them with other groups who agreed to take part in the audit. This of course led to some delays in conducting the audits. Our final sample included the School of Divinity, Economic and Social History, Centre for Integrative Physiology, Brain Imaging, and the Institute for Astronomy. The key contacts from these groups provisionally agreed to take part in the audits.

We first designed an online questionnaire to collect as much information as we could on existing research data and data management practices. Our key contacts circulated the questionnaire within their groups. Unfortunately, the timing was tricky, as the circulation of the questionnaires coincided with summer vacation followed by the new academic year. Some staff were reluctant to complete the surveys as they were too busy. Some of them mentioned that they were having “survey fatigue” as they were being asked to complete too many questionnaires for different projects (such as SCARP). For whatever reasons, the response rate we received to the questionnaires was fairly low in most groups.

On the other hand, some of the groups preferred to have face-to-face meetings rather than completing the online questionnaire. In these cases we conducted semi-structured interviews with these staff. However, arranging the interviews was time consuming. We also had difficulty as outsiders to the department with locating and accessing the documentation where data assets were described, and it was impossible to access shared or personal drives where the data were stored, as suggested by the methodology.

Despite these problems we managed to collect some information about existing research data and data management practices in these groups, and prepared five case studies discussing some of the common data issues. None of the audits was a comprehensive survey of research data but they proved to be a good starting point to auditing research data holdings and investigating data management practices within the University.

Outputs and Results

The key outputs from the project were five case studies – the final ‘audit’ reports for each research unit where we described the existing state of the data creation, management, and curation procedures and discussed some common data issues such as inadequate storage, lack of inconsistencies filing structures, naming convention etc.

In addition to these key outputs the following presentations and articles were disseminated:

- An [article](#) in Edinburgh BITs (July 2008) to promote the project within the University.
- A presentation on the [DataShare and Data Audit Framework projects](#) at Edinburgh at two events presented by Robin Rice:
 - [DCC/Edinburgh e-Science Collaborative Workshop](#), National e-Science Centre, University of Edinburgh, 12 June, 2008
 - [Research Data Management Workshop](#), University of Oxford, 13 June, 2008.

- A presentation on the "[Experiences gained from implementing the Data Audit Framework](#)" presented by Çuna Ekmekcioglu at the [Data Audit Framework and DRAMBORA Online launch event](#), British Academy, 1st October, 2008, London.
- "Lessons learned from implementing the Data Audit Framework", by Çuna Ekmekcioglu at the Data Audit Framework Workshop, [4th International Digital Curation Conference, "Radical Sharing: Transforming Science?"](#), National e-Science Centre, University of Edinburgh, 1st December, 2008
- Jones, Sarah, Ball Alexander & Ekmekcioglu, Çuna, [The Data Audit Framework: a first step in the data management challenge](#). International Journal of Digital Curation, Vol. 3, No.2, 2008.

Outcomes

The project achieved its aim in terms of providing exemplars of Data Audit Framework activity in a UKHE institution. In addition to carrying out the data audits and writing up the case studies, project staff engaged in a number of dissemination activities such as speaking at the DAFD launch and other events, writing newsletter and journal articles, and writing the Lessons Learned document (actually, this was written as part of our participation in the DAFD project led by DCC/HATII, but these projects are inter-related). The interest in this activity can be seen by the great volume of questions that accompanied the project manager's presentations about 'lessons learned' at two events (lasting over twenty minutes in both cases).

In terms of revealing current practices at Edinburgh, the project has been quite successful and will be used by the steering committee to develop strategies and services within the University. As anticipated, discovering 'what data has been created at Edinburgh, where it is located and who is (or is not) responsible for it' could only be done for representative samples, and even in organisational units where activity was focused, it revealed something more akin to the 'tip of the iceberg' than a full audit. On the other hand, particularly in the Humanities, there were many cases where original data simply hadn't been created, and researchers were active by using published sources and secondary data sources.

Conclusions

Generally speaking, the audits were a good starting point and useful to identify the gaps and issues in managing data assets in the schools and units audited. Staff had numerous comments and suggestions for improvement of data management at different levels indicating an awareness of the issues, even where it has not been made a priority to address.

While further awareness-raising is still important, staff require pragmatic assistance in the form of guidance on best practice, research unit or school procedures, College or University-wide infrastructure and policy, and identifiable forms of support for data curation in the form of expert support staff, web pages, and discipline-specific guidelines, as well as short, focused, training opportunities.

The data audit work was extremely intensive for one person to carry out and could not be sustained at the same pace and resourcing levels beyond the project time period. In any case, there did not seem to be reason to carry on conducting data audits, when the locus for data curation does not currently rest with the institution, but with the individual researchers, or at least the Principal Investigators. It was difficult to overcome the sense that the researchers were helping the project staff by participating, rather than vice-versa. This may be partly because the metadata collected as part of the DAF methodology is most useful to an outsider (e.g. title, description, etc), rather than someone deeply familiar with the data, who would require more domain-specific metadata fields for the audits to be useful to them (such as in the GeoSciences, geographic extent, projection, etc.). See the survey done for the GRADE project for more about attitudes to sharing geospatial data (Smith, 2007).

Implications

The DAF can provide a very useful starting point for librarians and others who wish to extend their support for research data into the university community, by providing a vehicle for engaging with researchers to describe the data they hold, and a focus for discussion of current practices for curation. The project team would like to suggest that other universities use the DAF and the new online tool created by the DAFD team.

Nonetheless, the term audit may be misleading, because without stringent procedures in place, the data used by any whole department currently and in the past is unlikely to be entirely known to the department's researchers at the detail required by the DAF. Previous projects such as StORe have already determined that only a small amount of the data generated by UKHE in most disciplines is deposited in a trusted data repository for preservation and re-use by others (Pryor, 2007). Data that are not documented for future use are unlikely to be kept in a form that can be retrieved, catalogued, and audited.

Fuzziness persists, in terms of defining the data types and currency of data to be audited, and particularly in identifying the locus of responsibility for further curation beyond the goals of using data as a means to the end of publication. If those conducting the audits are not considered to be the data owners, then there is a question about who the audit is being done *for*.

The implications of the OECD principles on public access to research funding (OECD, 2007) and other similar top-down pronouncements are that institutions are responsible for the data created by their employees. However, current norms in most disciplines regard the researchers themselves as solely responsible for their data, as research inputs for scholarly publication outputs. Data management plans as currently (and recently) required by some of the research councils can help raise the level of responsibility felt by research staff for curating their data, but there is uncertainty as to how this aspect of grant proposals affects the viability of the research proposal overall (if at all). In any case, without institutional data policies, institutions have little control over how these are written or implemented and researchers will continue to presume that they are solely responsible for the data produced.

Recommendations

The DCC should continue to support the online DAF tool, and support and development for this should continue to be funded by JISC. Take-up should be monitored (there is evidence of interest from the US) and coupled with further strategic and pragmatic efforts to build infrastructure and support for data curation in UKHE.

References

- Gibbs, H. (2007). [DISC-UK DataShare: State-of-the-Art Review](#). DISC-UK, August 2007.
- Green, A. and Gutman, M. P. (2007) [Building Partnerships among Social Science Researchers, Institution-based Repositories and Domain Specific Data Archives](#). *OCLC Systems and Services*, 23 (1), 35-53. <http://deepblue.lib.umich.edu/handle/2027.42/41214> [Open Access version]
- Lyon, L. (2007) [Dealing with Data: Roles, Responsibilities and Relationships. Consultancy Report](#). UKOLN.
- OECD (2007) [OECD Principles and Guidelines for Access to Research Data from Public Funding](#). OECD.
- Pryor, G. (2007) [Project StORe: Making the connections for research](#). *OCLC Systems and Services*, 23 (1), 70-78.
- Smith, R. S. (2007) [Geospatial Data-sharing in UK Higher Education: Informal Repositories and Users' Perspectives](#). University of Sheffield, UK.

Appendix A

Data Audit Framework: Edinburgh GeoSciences pilot study

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 by the auditor contacting the owner of the data assets as and when necessary. It was rather difficult to complete all the meta data 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 was 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.

Appendix B

Data Audit Framework implementation: Edinburgh case studies

Introduction

One of the specific objectives of the Edinburgh Data Audit Framework implementation project was to provide exemplars of data audit in research centres and or schools in each of the three Colleges and therefore across a range of disciplines. To achieve this objective five case studies were undertaken in the following schools or centres:

- Centre for Integrative Physiology
- School of Divinity
- Economic and Social History
- SFC Brain Imaging Research Centre
- Institute for Astronomy

Implementation

Data Audit Framework methodology was tested in pilot audits with academic/research staff in the selected schools and centres.

Considering the short duration of the project and the large number of staff to be contacted we decided to implement the audits through an online survey followed up with interviews as and when necessary. In total 15 questions were included in the survey. These questions mainly covered most of the fields given the Form 2 and 3 of the audit methodology together with some extra questions on storage quantities, backup, who's responsible for data and whether there are any policies / plans for data management . The online survey was circulated to academic/research staff via our key contacts in each school and centre. In most cases there was a long delay in response time and the response rate was quite low, and in some cases we had to rely only on interviews with staff as this was their preferred way of participation in the audits.

Once all the responses to the survey were received and the interviews were completed, an inventory of data assets was created for each school and centre. Since most of the assets collected were classified as 'important' we did not carry out a detailed assessment of these datasets.

Generally speaking, rather than being comprehensive audits the implementations were more requirement gathering exercises to learn about current data management practices and identify gaps existing in services.

Success factors

Implementing the methodology was straightforward. Although the implementations were not comprehensive audits, information collected provided a detailed view of the volume of data assets, data types, storage and back-up issues, current skills gaps in data management, and issues with the retention of the data assets.

Some of the common observations and issues which arose from the case studies included:

Storage provision: In most cases there was insufficient storage space available on the servers. Many datasets (mostly arising from small-scale projects) were stored by researchers themselves in a more or less ad hoc manner on personal external storage devices with little chance of effective retrieval. As a result, these datasets were not managed effectively or made readily-accessible to other researchers. Regular back-up facility was only available for the data that is stored on the school server.

Data value and retention period: Majority of the participants' data were very valuable and it would be difficult to regenerate the data in case of a loss. In some cases, particularly interviews and

surveys, it would be impossible to regenerate the data. The data were generally retained up to or more than 10 years.

Lack of a formal data management plan: All the participants –only with a few exceptions - who involved in the interviews and completed the survey acknowledged that they did not have a formal data management plan. Some research staff mentioned that there is pressure from funding bodies, especially, to ensure that data, once created, is properly managed and stewarded. Staff also indicated that they would prefer the issue of data management was raised at the beginning of the research process rather than at the end of the project.

Lack of guidelines and standardised procedures in creating and storing data: In most cases access to data stored on the server was not straightforward as the data were not catalogued or there was little manual cataloguing. 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. Metadata was sparse at best. When such metadata items existed, they were either minimally populated or were the default files generated automatically by the data's host proprietary application (and were hence incomplete).

A clear message that came through was the urgent need to develop greater awareness and understanding of data management within the university as well as guidance on best practice.

Lessons learned

In terms of revealing current practices at Edinburgh, the project was quite successful. The audits were a good starting point and useful to identify the gaps and issues in managing data assets in the schools and units audited. Interviews with staff provided a base for detailed discussion on data creation and curation practices which was useful to build rapport and help staff engage with the audits. During the audits it was observed that some of the schools and units audited had an extensive experience in handling the data they collect and/or produce, whereas in some others, particularly in the Humanities and Social Sciences, researchers were active largely by using published sources and secondary data sources.

Challenges encountered during the implementations:

- It was clear that staff engagement was an important challenge to be addressed. Engagement was achieved but not to the extent hoped for. There was a perception amongst some research staff that they were helping the auditor rather than vice versa.
- Staff perception of auditing their research data. In general research staff view was that anything other than producing research output is unproductive as this is how they are rated.
- The online survey did not return a high response rate. Most staff indicated they were having a 'survey fatigue' due to a large number of surveys they were asked to complete for various projects. Another group wasn't keen on completing online surveys and they preferred to be interviewed instead.
- Time needed to set up and conduct interviews. A number of email correspondences and telephone calls were required to set up the interviews. Most of the time there were no responses to emails at all. It was also difficult to set interview dates and times easily due to lectures, field trips, start of the academic year, exam times etc. Thus, the methodology was time consuming to implement when based on interviews.
- Scoping was an issue. How far back in time to go was a recurrent question, as was what constitutes data – does it include 'paper data' as well, and in terms of content is it analysis rather than any kind of information. The status of derived data was another major question in some cases as it was a vital asset to the researchers but they did not have curatorial control.
- The data audit work was extremely intensive for one person to carry out and could not be sustained at the same pace and resourcing levels beyond the project time period.

Key factors for successful implementation of data audits:

- Support and commitment from the institution
- Provision of sufficient storage space for staff to encourage them store their data on regularly backed-up, reliable and secure storage area networks
- Integration of data management practices into the institutional policies
- Raising greater awareness and understanding of data management within the institution
- Establishing best practice guides and providing advice as and when needed
- Provision of training in data management and curation
- An enthusiastic coordinator/champion in the school/centre to 'kick start' the process and maintain momentum
- Regular audits introduced at the beginning of a research project
- Flexibility in the methodology/tool to allow for different contexts/needs (avoiding a 'one size fits all' approach')
- Do not assume that academic/research staff have time and the skills to deal with their data.

Conclusions

On the basis of the information gathered from interviews and surveys during the audits, the following recommendations were drawn up for the Schools and Information Services:

1. There is an urgent need to develop greater awareness and understanding of data management within the university as well as guidance on best practice.
2. An institutional data policy with guidance on best practice in data creation, management and long-term preservation would provide schools and principal investigators with a basis from which they could develop policies and practices suited to their particular contexts. Templates for writing data management plans could be made available to research grant writers.
3. Information Services needs to consider collaborating with Colleges/Schools to deliver co-ordinated training programmes and supporting materials, targeted at researchers and postgraduate students in specific disciplines.
4. Information Services needs to provide more networked storage services for staff, and a modernised "archive" (off-line, backed up) service.
5. Information Services should continue to offer support for conducting data audits. This needs to be voluntarily adapted by schools or research units. A risk analysis technique could assist with classifying the datasets.
6. The School management in the form of the IT Committee may consider how to resource a more in-depth and complete audit of the existing research data within the school as a precursor to developing strategies for effective data management on the shared servers. Metadata might be gathered at the time that it is reported or submitted to research funding bodies.
7. The School IT Committee needs to consider a consistent approach to data management; e.g. investigating procedures for file naming and version control on shared servers and developing guidelines for good practice to improve data management within the School. Policies for what data is stored and backed up on the servers versus for data stored off-line need to be rationalised.
8. Principal investigators should write a data management plan at the beginning of new projects, taking into account plans for preservation and sharing after the end of the project.
9. Principal investigators should consider the use of EDINA and Data Library services at the University of Edinburgh for documenting and sharing metadata and data.

Case Study 1: Data Audit Framework implementation in the Centre for Integrative Physiology (CIP)

Background

This study presents the findings from a data audit pilot study undertaken in the Centre for Integrative Physiology at the University of Edinburgh. The study is set in a context where the amount of digital data being created and gathered by researchers is increasing rapidly; and there is a growing recognition by researchers, of the potential value in managing data efficiently.

The Centre for Integrative Physiology (CIP)⁴ was established in 2004 as an interdisciplinary research centre of the College of Medicine and Veterinary Medicine (CMVM), hosted within the School of Biomedical Sciences with a mission 'to develop and promote international excellence in research and training in Integrative Physiology relevant to human and mammalian physiology, development and disease'.

The scientific remit of the CIP is to investigate key physiological mechanisms across multiple scales of granularity to determine the role of gene products in the integrated function of cells, organs and whole organisms. Investigators exploit appropriate model systems and focus activities that directly facilitate the understanding of human and mammalian physiology, development and disease. Thus CIP investigators span the areas of genomics, proteomics, imaging, whole animal physiology and informatics.

The research activities of the CIP are focused around the 'four cornerstones' of the Integrative Physiology approach with investigators working across these domains:

- Manipulation and regulation of genes and their products
- Molecular imaging and quantitation
- Complex systems function and behaviour
- Predictive modelling of complex systems

The majority of investigators are in the interdisciplinary groups in Genes & Development, Membrane Biology and Neural Control Systems with other investigators based at other sites across the CMVM and College of Science & Engineering (CS&E).

Implementing the DAF methodology

The Data Audit Framework methodology⁵ was implemented in a pilot audit in CIP through a mixture of an online survey directed at projects run by individual researchers and semi-structured interviews.

The online survey based on questions about the projects, data structures and existing data management practices in the Centre. Also, semi-structured interviews were held with a key staff and the IT Services Manager of the School to discuss the best options for piloting the audit, and also current data creation and data management practices in the Centre.

After the initial interview our key contact circulated the online survey within the Centre encouraging colleagues to complete and return it. 14 completed surveys were returned in less than 2 weeks with a response rate of 50%.

Analysis of these surveys together with the results of the desk research and the analysis of the interviews with our key contacts led to an inventory of 14 data assets. The audit results were also invaluable in highlighting the current data management practices within the Centre. These are reported as key findings in the following section.

Key findings

⁴ <http://www.cip.ed.ac.uk/>

⁵ http://www.data-audit.eu/DAF_Methodology.pdf

Data characteristics

The inventory of data assets mainly includes simulated, derived, reference, experimental, and observational data. The variety of data types ranges from documents, reports, MS Word, MS Excel Sheets, laboratory notes, video tapes and microscope slides to digital video files, images, scans, x-rays, data collected from sensors or instruments (including questionnaires) and data automatically generated from or by computer programs.

Storage and backup issues

CIP uses 2 methods to store data:

- Access to the College servers. Staff have access to folders on the shared network drives. Staff access these folders on an ad hoc way (approximately 30 staff). Servers are backed up by the college therefore staff are encouraged to keep a regular back-up of their own data.
- Several individual groups own their own servers to store their data and purely for use within their groups. The rest is storing data on a very large server (windows machine). There is some sort of cataloguing available but this is restricted to only manual cataloguing. It is usually up to the PIs to decide how to structure the data on the servers. The volume of data stored on all the servers for the entire school is approximately 3 Tb and is constantly growing.

Most staff –i.e. about 60% - are not producing large volumes of data therefore data produced are usually suitable for in-house storage. About 40% of staff works with large volumes of data. Most of the data generated in the Centre are in the form of spreadsheets, documents (laboratory notes) and PowerPoint presentations. Images (very high resolution –tiffs, jpegs) and movies make up 60% of the data. The School is also keen in buying in SAN (Storage Area Network) and waiting for the costing.

In terms of the value of data every piece of data is very valuable and loss of it may cause serious problems.

Data management

Individual researchers or the IT staff in the school/centre is responsible for managing the data, and there is a formal research data management plan in place in the unit. Data are generally retained up to 10 or more than 10 years.

The School is keen in using data management tools but the way staff work varies hugely from one project to another therefore these tools do not suit the needs of all staff in the Centre. Also, the appropriate tools usually come in the form of commercial packages and the cost of using these packages is very high. Therefore it is out of reach of most of the research groups. One of the research groups in the Centre bought an auditing commercial package (Biotrue⁶ – Collaborative Data Management System (CDMS) which is a server based software system for biomedical research laboratories that enables scientists to easily store, manage and share all types of instrument and analytical data files) which is capable of recording every aspect of the data generated. It is highly specialised and most suitable for biology.

Data sharing

Data assets are accessed within the research team. There are privacy and confidentiality issues with the data generated.

⁶ <http://www.biotrue.net/site/main/section/36>

Case study 2: Data Audit Framework implementation in the School of Divinity

Background

This case study presents the findings from a data audit study undertaken in the School of Divinity at the University of Edinburgh.

The School of Divinity⁷ is one of the largest single site centres for the study of theology and religious studies in the UK. Founded on a thriving 500-year tradition of creative thought and practice, the Schools research includes the work of three research centres:

- Centre for the Study of Christianity in the Non-Western World
- Centre for Theology and Public Issues
- Centre for the Study of Christian Origins

Academic research lies at the heart of the School. There are 30+ academic staff who are engaged upon their individual research and writing projects. Many are also involved in the three research centres or in a range of collaborative projects with colleagues in this university or throughout the world.

There are several active research projects⁸ in the School of Divinity, some of which are cross-disciplinary, and all of which act as centres of excellence for their subjects. Some also offer important source material and other texts for download.

Implementing the DAF methodology

The Data Audit Framework methodology⁹ was implemented in a pilot audit in the School through a series of semi-structured interviews with six key staff. The online questionnaire was not used in this case study as staff preferred to participate in interviews only.

The first stage of the audit involved some desk research -browsing the School website and compiling a list of research active staff with their research responsibilities. Following on from this preparatory work, an initial meeting was arranged with the School Computer Officer to discuss existing data processing and data management practices in the School and to identify the key staff to interview. Based on the information gathered from desk research and the interview with the Computer Officer, key staff were selected and invited to provide information about their research and research data via semi structured interviews.

In the following two stages interviews were conducted with academic/research staff, and an inventory of seven 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.

Detailed information gathered at the interviews was invaluable in highlighting the current data management practices within the Institute. These are reported as key findings in the following section.

⁷ <http://www.div.ed.ac.uk/>

⁸ <http://www.div.ed.ac.uk/research>

⁹ http://www.data-audit.eu/DAF_Methodology.pdf

Key findings

Data characteristics

Research staff we interviewed work with either raw data created or gathered in the course of experiments or observations; or derived data, resulting from processing other data. The variety of data types ranges from text corpus, documents, reports and transcripts to images, scans, video tapes and photo collections.

Storage and backup issues

The school doesn't have its own server. All data were stored on the college (CHSS) server and on the SAN. Some staff access SAN from home via VPN (Virtual Private Network). Regular back-up facility is only available for the data that is stored on the school server.

Access to data stored on the server is not straightforward as there is no main index or catalogue of the data held in the shared drives. Users are expected to search for the required data themselves, or with guidance from the relevant research staff.

Data management

All staff involved in the interviews acknowledged that they do not have a formal data management plan. Generally, the Research Project Manager or the IT staff in the School is responsible for managing the data. There is currently no formal requirement for researchers to have a data management plan. Some research staff mentioned that there is pressure from funding bodies, especially, to ensure that data, once created, is properly managed and stewarded. Others indicated that they would prefer that the issue of data management is raised at the beginning of the research process rather than at the end of the project.

Most of the participants' data are very valuable and it would be difficult to regenerate the data in case of a loss. In some cases, particularly interviews and surveys, it would be impossible to regenerate the data.

Data sharing

Some data assets from various projects are available on the Web offering important source material, images and other texts for download:

- Collection of Scottish Letters¹⁰
- Scottish Book of Hours MS42¹¹
- Religion in Late Medieval Scotland (RLMS)¹²
- Missionary Collections in the UK: Mundus project¹³
- Nepal Digital Images¹⁴
- Semantics of Ancient Hebrew Database¹⁵

¹⁰ http://www.div.ed.ac.uk/scottishlett_1.html

¹¹ <http://www.div.ed.ac.uk/ms42>

¹² <http://www.div.ed.ac.uk/rlms>

¹³ <http://www.div.ed.ac.uk/missionaryco.html> and <http://www.mundus.ac.uk/>

¹⁴ <http://www.div.lib.ed.ac.uk/community-list>

¹⁵ <http://www2.div.ed.ac.uk/research/sahd/>

Case study 3: Data Audit Framework implementation in the Economic and Social History

Background

This case study presents the findings from a data audit pilot study undertaken in the Economic and Social History at the University of Edinburgh.

Economic and Social History¹⁶ in Edinburgh is a team of eleven permanent, full-time academic staff based in the School of History, Classics and Archaeology, making it one of the largest Economic and Social History units in the United Kingdom.

Economic History is the study of the way in which economies develop, why that development differs between countries and over time, and how individuals, households and communities contribute to, and are affected by, economic change. Social History is concerned with how people have lived their lives and how and why their experiences and behaviour have changed over time, asking a wide range of questions about social behaviour, organisations and identities in the past.

The research interests within the unit are wide ranging, including the study of economic development, energy policy, globalisation, slavery, demography, urban history, consumption, material culture, museums and collecting, leisure, religious belief, popular culture, medicine and disease, gender, sexuality and the family.

Implementing the DAF Methodology

The Data Audit Framework methodology¹⁷ was implemented in a pilot audit in the School through an online questionnaire and a series of semi-structured interviews with six key staff. In this case study only one staff completed and returned the online questionnaire. The rest preferred participating in interviews.

The first stage of the audit involved desk research to gather as much information as possible about the type of research undertaken in Economic and Social History unit, and research responsibilities of the academic/research staff working in the team. Following on from this preparatory work, an initial meeting was arranged with the School Computer Officer to discuss existing data processing and data management practices in the school and to identify the key staff to interview. Based on the information gathered from the desk research and the interview with the Computer Officer, the key staff were selected and invited to provide information about their research and research data via semi structured interviews.

In the following two stages interviews were conducted with 5 academic/research staff, and an inventory of 8 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.

The results of the desk research together with the analysis of the online survey and detailed information gathered at the interviews were found very useful to highlight the current data management practices within the Unit. These are reported as key findings in the following section.

¹⁶ <http://www.shc.ed.ac.uk/esh/index.htm>

¹⁷ http://www.data-audit.eu/DAF_Methodology.pdf

Key findings

Data characteristics

Research staff we interviewed work with simulated, derived, reference, and observational data. The variety of data types ranges from historical archives in local and national collections, documents or reports, MS Word files, MS Excel Sheets and SPSS files to slides, photo collections, fieldwork data, and websites.

Storage and backup issues

The school has one main server however staff are encouraged to store their data on the college (CHSS) server. Many datasets (mostly arising from small-scale projects) are stored by researchers themselves in a more or less ad hoc manner on personal PCs, laptops or external storage devices with little chance of effective retrieval. As a result, these datasets are not managed effectively or made readily-accessible to other researchers. Regular back-up facility is only available for the data that is stored on the college server.

Most of the data is very valuable and it would be difficult to regenerate the data in case of a loss. In some cases, particularly interviews and surveys, it would be impossible to regenerate the data.

Data management

Usually, the Research Project Manager, individual research staff or the IT staff in the School is responsible for managing the data.

All staff involved in the interviews acknowledged that they do not have a formal data management plan. Although there is pressure from funding bodies, to ensure that data are properly managed and stewarded, there is currently no formal requirement for researchers to have a data management plan. However, at the time of study a guidance document on data storage was being written to provide staff with good practice guidelines.

Data sharing

Some data have already been deposited to national archives. Some others, particularly interview data, were not deposited due to confidentiality issues. As for sharing data, one staff raised the issues around the fitness for purpose, particularly for social science data, the quality of data deposited, and high level requirement for documenting data. Data are expected to be documented well if they have to be deposited. However, there is a trade off between researcher's time to document the data and time to carry out more research.

Case study 4: Data Audit Framework implementation in the SFC Brain Imaging Research Centre

Background

This case study presents the findings from a data audit pilot study undertaken in the SFC Brain Imaging Research Centre at the University of Edinburgh.

Brain Imaging¹⁸ is one of the smaller interdisciplinary Research Groups (IDGs) established within the School of Molecular and Clinical Medicine with a mission to use advanced techniques to image brain structure, blood flow and function, particularly in disease states.

SFC Brain Imaging Research Centre¹⁹ provides imaging and technical support for studies in:

- Psychiatry
- Psychology
- Stroke
- Oncology

Approximately, 12,000 subjects were studied at the centre in 11 years. They acquire imaging from NHS patients, NHS patients in research studies and volunteers.

Implementing the DAF methodology

The Data Audit Framework methodology²⁰ was implemented in a pilot audit in the School through an online survey directed at projects. The survey based on detailed questions about the projects, data structures and data management practices in the unit. Two research staff completed the survey for 21 projects with a response rate of 5%.

Analysis of these surveys and the results of the desk research led to an inventory of 21 data assets. The audit results were also useful to reveal the current data management practices within the Institute. These are reported as key findings in the following section.

Key findings

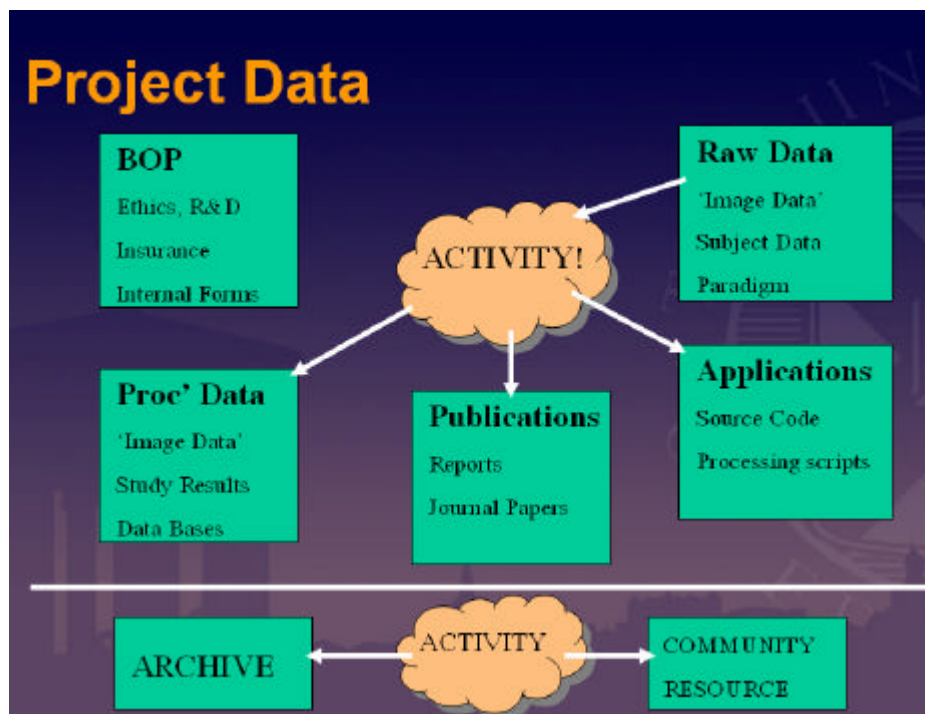
Data characteristics

The inventory of data assets mainly includes experimental, observational, reference data. The variety of data types ranges from images, scans, x-rays, and MS Excel Sheets to patient records. Project data consist of raw data which is usually in standard format (DICOM- e-Prime) and processed data which can be in various formats (often MS Excel or MS word) varying from project to project. The software used for analysis is often closed source and there is no version control. Processed data is usually the responsibility of the Principal Investigators. The diagram below describes the project data.

¹⁸ <http://www.mvm.ed.ac.uk/idg/brainimg.htm>

¹⁹ http://www.nesc.ac.uk/talks/898/CARPENTER_DataCur.ppt#257.3.SBIRC

²⁰ Data Audit Framework Methodology: http://www.data-audit.eu/DAF_Methodology.pdf



Description of the project data: Trevor Carpenter²¹

Storage and backup issues

Data assets are mainly stored on the offsite archive, tapes, CDs, SAN (Storage Area Network), and on occasional case on the external hard drive. There is a regular back up facility for the data collected. Data are very valuable and it would be difficult to regenerate the data in case of a loss.

Data management

Individual researchers or the IT staff in the school/centre is responsible for managing the data, and there is a formal research data management plan in place in the unit. Data are generally retained up to 10 or more than 10 years. Current developments with regard to data management involves the development of Standard Operating Procedures which cover archiving of project data including:

- Creation of data dictionaries
- Documenting analysis procedures
- Specifying preferred data formats

Data sharing

Data assets are accessed only within the research team as raw data. Researchers outside the team are not allowed to access the data assets due to privacy and confidentiality issues. There is one data asset mentioned in the inventory as 'unique in the world'. At present there is no access to this data asset but the researcher is willing to make some or all of it available only some years after the end of the project.

²¹ [http://www.nesc.ac.uk/talks/898/CARPENTER_DataCur.ppt#289,7,Project Data](http://www.nesc.ac.uk/talks/898/CARPENTER_DataCur.ppt#289,7,Project%20Data)

Case study 5: Data Audit Framework implementation in the Institute for Astronomy

Background

This study presents the findings from a data audit pilot study undertaken in the Institute for Astronomy at the University of Edinburgh. The study is set in a context where the amount of digital data being created and gathered by researchers is increasing rapidly; and there is a growing recognition by researchers, of the potential value in managing data efficiently.

The Institute for Astronomy (IfA)²² is a research and teaching group within the School of Physics and Astronomy, University of Edinburgh. It is one of the UK's major centres of astronomical research.

There are currently eleven permanent academic staff, four visiting or emeritus staff, and a large number (roughly twenty) of research and technical staff, as well as around thirty postgraduate research students. The research programmes pursued at the IfA are extremely varied, but the main strengths are in cosmology, active galaxies and the formation of stars and planets, and in survey astronomy and e-Science undertaken by the Wide Field Astronomy Unit (WFAU) on behalf of astronomers in the UK and other ESO member states.

Implementing the DAF methodology

The Data Audit Framework methodology²³ was implemented in a pilot audit in the Institute through a mixture of an online survey directed at projects run by individual researchers and a semi-structured interview with a group leader.

The online survey based on questions about the projects, data structures and existing data management practices in the IfA. Also, a semi-structured interview were held with one of the group leaders to discuss the options for piloting the audit and current data creation procedures and data management practices in the Centre.

After the initial interview our key contact circulated the online survey within the IfA encouraging his colleagues to complete and return it. He also used DAF online tool to record information about WFCAM Science Archive (WSA). His comments on using the tool were positive except that he had difficulty in selecting a 'classification' from the drop-down menu.

6 completed surveys were returned at the end of the audit with a response rate of 30%. Analysis of these surveys together with the results of the desk research and the analysis of the interview with our key contact led to an inventory of 12 data assets. The audit results were also invaluable in highlighting the current data management practices within the Institute. These are reported as key findings in the following section.

Key findings

Data characteristics

The inventory of data assets mainly includes simulated, derived, reference, experimental, and observational data. The variety of data types ranges from documents or reports, to images and data collected from sensors or instruments (including questionnaires) and data automatically generated from or by computer programs.

²² <http://www.roe.ac.uk/ifa/>

²³ http://www.data-audit.eu/DAF_Methodology.pdf

Most of the flat-file data stored at IFA is stored in FITS files. FITS is a widely used standard data format in astronomy, which was initially used for image data, but which is now used for most types of astronomical data.

It is worth noting the distinction between the data that individual IFA researchers curate which originate in their own research and the archives which WFAU runs on behalf of the UK –and increasingly, European- astronomical community.

Storage and backup issues

Telescopes and standard code are used on raw data, and standard data products are stored in archives. Data are also stored on tapes, SAN (Storage Area Network), third party (including commercial data storage) storage, school server, CDs, and mirror disks, and regularly backed-up. Most data are updated monthly.

Data management

Individual researchers, the IT staff in the institute or designated person on project is responsible for managing the data. Data are generally retained up to 10 or more than 10 years. There is no formal data management plan in the Institute.

Data sharing

Access to data is typically restricted for a year or eighteen months after ingest into the archives. At the end of this period it is open access and data are shared with other institutions. IFA holds on to a large number of metadata exposed over the web which will ultimately be open access.

Some of the archives curated by WFAU on behalf of the community are:

- WSA - WFCAM Science Archive²⁴ holds the image and catalogue data products generated by the Wide Field Camera (WFCAM) on UKIRT.
- SSA - SuperCOSMOS Science Archive²⁵ holds the object catalogue data extracted from scans of photographic Schmidt survey plates.
- SuperCOSMOS Sky Surveys (SSS)²⁶ provides on-line access to the digitised data derived from SuperCOSMOS scans of photographic Schmidt survey plates.
- SuperCOSMOS H-alpha Survey (SHS)²⁷ provides access to SuperCOSMOS digitised data derived from scans of survey films of the Southern Galactic Plane taken for the AAO/UKST H-alpha survey
- *6dF Galaxy Survey Database*²⁸ provides on-line access to the database built around the 6dF Galaxy Survey.
- WFCAM/VISTA Science Archive²⁹ development at WFAU is part of the VISTA Data Flow System project.
- Archive³⁰ of original photographic plates observed with the UK Schmidt Telescope and copies of photographic sky atlases produced from the UK, Palomar and ESO Schmidt telescopes

The units within IFA are responsible for maintaining these archives and planning to link them to Astrogrid next.

²⁴ <http://surveys.roe.ac.uk/wsa/>

²⁵ <http://surveys.roe.ac.uk/ssa/>

²⁶ <http://www-wfau.roe.ac.uk/sss/>

²⁷ <http://www-wfau.roe.ac.uk/sss/halpha/>

²⁸ <http://www-wfau.roe.ac.uk/6dFGS/>

²⁹ <http://www.roe.ac.uk/~nch/wfcam/>

³⁰ <http://www.roe.ac.uk/ifa/wfau/ukstu/>

Appendix C

Online questionnaire

Section I. Personal details

1. Full name
2. Academic role
3. Research group or research active area

Section II. Details of your research and research data

Please describe here your most recent research project and provide information on the data generated or used in this research project by answering the questions below.

4. Project details
5. Description of the data
6. Ownership: who owns the data?
7. Characteristics of the data (select all that apply)
 - a. Observational
 - b. Experimental
 - c. Reference
 - d. Derived
 - e. Simulated
 - f. Not Applicable
8. Data types (select all that apply)
 - a. Data automatically generated from or by computer programs
 - b. Data collected from sensors or instruments (including questionnaires)
 - c. Images, scans or x-rays
 - d. Websites
 - e. MS Word files
 - f. Excel sheets
 - g. SPSS files
 - h. Digital audio files
 - i. Digital video files
 - j. Fieldwork data
 - k. Laboratory notes
 - l. Photo collection
 - m. Video tapes
 - n. Audio tapes
 - o. Slides
 - p. Text corpus
 - q. Documents or reports
 - r. Patient records
 - s. Other (*please specify*):
9. Size of the data
10. Importance of the data
 - a. Vital
 - b. Important
 - c. Vital

11. Retention period
 - a. Only over the project period
 - b. Up to 5 years
 - c. Up to 10 years
 - d. More than 10 years
 - e. Don't know
 12. How frequently do you update your data over the project period?
 - a. Never
 - b. Daily
 - c. Weekly
 - d. Monthly
 - e. Annually
 - f. Don't know
 13. Is your data backed up regularly?
 - a. Yes
 - b. No
 - c. Don't know
- If it is, where is it backed up? (Select all that apply)
- a. School server
 - b. Storage area network
 - c. DVDs
 - d. CDs
 - e. USB/Flash drives
 - f. External hard drives
 - g. Tapes
 - h. Third party (including commercial data storage)
 - i. Don't know
 - j. Other (*please specify*):
14. Do you currently have a formal Research Data Management Plan in place in your school/centre?
 - a. Yes
 - b. No
 - c. Don't know
 15. Who is currently responsible for managing the data? (select all that apply)
 - a. Research project manager
 - b. Designated person on project
 - c. External project partners
 - d. IT staff within your school, centre or research institute
 - e. Research assistant
 - f. Yourself
 - g. National data centre or data archive
 - h. Nobody
 - i. Don't know
 - j. Other (*please specify*):

Appendix D

Project Steering Group members

Dr Chris Adie	Academic Liaison Director for the College of Science & Engineering
Dr Alastair F Brown	Head of Computing, MRC Human Genetics
Peter M Burnhill	Director, EDINA
Dr Trevor Carpenter	Division of Clinical Neurosciences
Susan Graham	Records Manager
Hamish MacAndrew	Head of Research Support & Development / Business Manager - Humanities & Social Science, Edinburgh Research & Innovation
Graham Pryor	Associate Director & eScience Liaison, Digital Curation Centre
Jean Ritchie	Service Director, IS Research Computing
Elize Rowan	Acquisitions & Metadata Services Manager
Dr James K Stewart	Research Fellow, Research Centre for Social Sciences
Arnott Wilson	Special Collections Manager & University Archivist