Digital Archives for Archaeology and the Historic Environment, a report from the 'Bedern Group'

York 17th October 2011

1. Introduction

The Digital Preservation Coalition was invited to facilitate a conversation between English Heritage, RCAHMS, RCAHMW and ADS to explore closer collaboration on the delivery of digital preservation services for archaeology and the historic environment. The agencies recognised that, thoughtfully defined, closer collaboration could deliver advantages to all parties such as improvements in quality, more effective interoperability of processes, greater efficiencies in operation or economies in scale. Moreover, trusted preservation services benefit from peer scrutiny and transparency. So, even if closer collaboration proves to be impractical, these discussions may yet provide the basis for independent review and enhancement.

2. Notes from the meeting

The following were in attendance: Catherine Hardman, Stuart Jeffrey and Julian Richards (ADS); William Kilbride and Carol Jackson (DPC); Duncan Brown, Mike Evans and Keith May (EH); Lesley Ferguson, Kirsty Lingstdat, Diana Murray and Emily Nimmo (RCAHMS); Gareth Edwards and David Thomas (RCAHMW). The meeting was chaired by WK and CJ provided notes on the discussion and actions.

The meeting opened with some preliminary discussion about what partners hoped to achieve with the time and expertise available. There was shared agreement about the growing scale, complexity and expectations associated with the digital data which each collects and a recognition that the data sets produced in archaeological research were manifestly challenging for long term preservation. Issues such as shared standards to support data creators, shared practices to reduce costs and enhance quality or efficiency in preservation, shared and transparent planning of preservation functions, and greater clarification of 'who is doing what' were identified as being of mutual concern. Operational issues around preservation were also raised. There was mutual agreement of the need for consistency and transparency of charging practices, the need to clarify whether the depositor of an archive was the field unit or the museum and the need to develop and train the workforce at all levels. There was mutual agreement on the expectation of interoperability at presentation level and the need for preservation to access subject specialist expertise rather relying entirely on generic archival practice.

This open discussion led the group to agree that an effort was required to develop and refine a shared vision for heritage data archiving. This vision would be expressed in three types of output:

- more localised collaboration in practical preservation actions between the operational staff of the digital preservation facilities
- the clarification and consolidation of practices in order to lower the barriers to the safe and timely deposit of collections, facilitating greater and more effective communication between archives and producers of content
- a clearer public statement about the value of digital data and the need from prompt and adroit action to secure that value. Knowledge is not a by-product of heritage: it is the essence of it. Archives are an asset which, properly managed and open, create opportunities and deliver greater impact from underlying processes which created them.

In order to develop these ideas each organisation gave a short impromptu presentation on their current work and expectations for the next three years

- ADS is core funded from research councils and concentration therefore core concentration is on academic collections, wherever academics chose to operate. Occasional funds from other sources mean that the ADS collections now include a large volume of projects primarily in England through EH funding but also from Scotland, Wales and Ireland. There are all types of data representing all types of archaeological intervention, and the archive is entirely digital. Core research council funding means that a charging policy is in place for those outside the education sector. The funding model will change in the next three years so there is a need to establish a baseline of costs from incoming deposits and/or development and research activities.
- RCAHMW is the national home for Welsh archaeological archives, including those produced by English units working in Wales. They are actively accessioning and curating digital archives, including those produced by their own survey and investigation staff, and are planning to improve their digital curation facilities, through the SWISH partnership with RCAHMS in the near future. There is a recognised need to improve and promulgate standards amongst external data creators and depositors, but no concerted programme of to promote these would take place before technical improvements to ingest and curation facilities are complete. RCAHMW would look to align with

national standards alongside other bodies such as RCAHMS and ADS. When issues of resources and infrastructure have been addressed, more concerted promotion of RCAHMW as the national home for such archives will take place. Currently RCAHMW does not charge but would consider doing so if there were a shared national position. Most work comes from the planning process, so feeding advice into the planning process is critical to their long term success: especially if this could give some advanced notice of what might be coming down the line. All work has a digital component but experience is that the planning process is tightly controlled at the start and trails off at the end.

- RCAHMS is national body for Scotland like RCAHMW in goals. The backbone to the collection is the in-house field group and their own programmes of aerial photography and survey. But external material comes in large quantities, partly through agreement with Historic Scotland, and RCAHMS is a home for all manner of small and large archives. All data types, all forms and sizes of projects are represented and this is only likely to increase. Some basic training is available for projects and there is a recognition of the benefits and need of automated ingest for digital collections. RCAHMS has a programme of visits to field units to help develop skills and tries to engage in new developments such as standards for 3d scanning. Large volumes of architectural material and this and aerial photography are the most popular elements of the collection. Online access is provided principally via CANMORE and because of the SWISH partnership, RCAHMS is able to maintain a programme of development work around access.
- EH NMR is an institutional repository for the work of English Heritage. All manner of collections and formats, and growing collections on account of active fieldwork and estates management. EH NMR also has a records management function for the organisation. EH would not refuse external collections but would not want huge volumes and only under controlled conditions. The wider community is not within the remit. Scan–on-demand is under consideration and be especially popular for photographic collections which is the largest volume of the collection. EH-funded research archives are not submitted to the EH NMR but are directed to use a 'reputable repository' which in practice means the ADS. EH is a large organisation with numerous functions and activities, especially around standards. Work like the Archaeological Archives Forum which EH has sponsored is highly relevant to the discussions.

3. Workplan

Discussions around who does what produced a summary of ten points which need to be addressed to create a unified vision for heritage data management. These items were not intended to be self-contained or exclusive, and nor do they require similar amounts of effort to resolve. However the group agreed that each of these themes needed to be pursued through appropriate agencies where these existed and that they would establish mechanisms to address them if appropriate vehicles did not already exist. Moreover the group held an informal vote to rank the issues in order of perceived importance for initial discussion:

- connecting digital archives with fieldwork and other collections (=1)
- developing and clarifying cost models and charging policies (=1)
- automating and standardising the ingest of collections (=3)
- consolidating and clarifying 'upstream' advice for creators contractors and curators (=3)
- understanding users and measuring the costs and benefits of 'value added' use (=5)
- resolving perceived overlaps in collecting policies (=5)
- certifying and quality assuring preservation services (7)
- developing skills for the sector and for digital preservation practitioners (=8)
- provision of 'cloud' storage or services in the cloud to facilitate preservation (=8)
- influencing industry and developers to provide preservation-ready systems (10)

Time did not allow a full discussion on each issue so the group picked the top two issues and had a discussion on how they might address these.

4. 'Connecting the archives'

The group noted that there were a series of discontinuities around the provision of archive services that needed to be addressed to simplify the processes of digital preservation, reduce the effort involved in their creation, to maximise their impact, improve their consistency, validate their content and strengthen their connection to the rest of archaeological process in museums, fieldwork and development control. Typical issues raised included the following example concerns:

- Archives are generated by people and organisations that are unlikely to be thinking about the archival value of the collection at the time they are created.
- The 'life of a monument' and its value is most readily expressed at an information level rather than at the level of the physical object: massive resources are expended conserving the physical object but only very little is spent on maintaining the real value which is in the information content of the archive.
- If fieldworkers don't know where their archives are going to end up, it's very difficult for them to prepare it to a standard that is sufficient for it to be deposited or used easily.
- Project design is the key to making sure that a project archive is well formed. If a research project has a well formed set of questions at the start then the archives will provide a simple means of assessing the value of the answers given and thus develop a well formed and incremental research agenda. But archaeology is not good at this and needs to get better at following questions and validating answers.
- If we can make it really easy for fieldworkers to follow instructions then it will be really easy for development control officers and/or museum curators to check whether the instructions have been followed.
- There is very little validation of the development control process which is very intrusive at the outset but much less effective after the project is complete. Clearer instructions will make it more likely that officers can check the results and take sanctions as they ought.
- Museum curators are supposed to be responsible for the whole archive but they are under a lot of pressure and are not likely to want responsibility to manage the digital archive: so they need to know what is coming, need to know how to validate it, and need to be able to pass it to the right digital repository.

The group was asked, 'What is preventing a clearer statement from repositories about what they will accept, how to present it and to whom?' In response the group agreed the following actions would improve the provision of advice:

- Develop training resources for museum curators and development control officers to help them understand digital archives more fully and helping them write requirements into the planning process **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**
- Encourage the early testing of archives to establish whether standards have been followed and to intervene in the development control process with sanctions if there are failings in this regard. **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**
- Harmonise the standards for deposition between repositories **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**
- Develop clear joint communications on the common standards for deposition, with an active outreach to fieldwork units **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**
- Work with SMA and others to agree a joint statement of what constitutes a conformant submission package and develop a simple checklist which they can readily apply to validate incoming materials **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**
- Issue a joint statement on archives, articulating their value and the opportunities associated with them, and aligning policies with relevant policy expectations regarding open data. **(Action: DPC to draft, appropriate officers in ADS, EH, RCAHMS, RCAHMW to review, endorse and publish)**
- Persuade funders to promote and insist upon a common approach to preservation of research archives, making it easier for the researchers and others to understand and follow a shared requirement **(Action: appropriate officers in ADS, DPC, EH, RCAHMS, RCAHMW)**

The group noted several consequential benefits from this which would have an impact on other elements of the workplan. For example, if common standards could be created then the likelihood was that SIPs would begin to be harmonised. This in turn would facilitate the development of practical collaboration between repositories on things like preservation planning, ingest and metadata creation. It would also contribute to risk assessment and provide some transparency with costing: a submission which conformed to a shared standard would be low risk and low cost whereas a one that did not conform would run at a higher risk and / or face a supplementary charge.

5. 'Charging and cost models for preservation

Discussion then moved to issues around charging and cost models for preservation. It was noted that the charging policies of different agencies and the models used to calculate costs varied between institutions and

this made it hard to recover costs and would be useful to the community in its development of technologies. It was also noted that it was relatively easy to claim to do preservation simply by adding content to a website and that preservation costs were often confused with storage costs: but preservation is more than just storage and therefore represented only a small element of the total costs. Typical issues including the following points:

- The most reliable model for charging was to provide a one-off charge at the point of deposition. Charging for access would not be practical except in unusual circumstances and recurring charges to depositors were unlikely to be feasible in the medium term.
- The 'polluter pays' principal is well established in archaeology so the costs of archiving development control archaeology should reside with the developer that commissioned the original fieldwork. Fieldworkers and development control officers should be clear about this when designing project briefs. Lack of clarity early on serves no-one.
- Charges would never be the same because local circumstances and priorities varied, but charging policies needed to be aligned so that units and developers could more easily understand and plan for archival deposition charges.
- Repositories should work out their own costs and should not act as a cartel: but the cost elements can be harmonised in order to ensure greater transparency and as a way of explain the different functions. Hence the LIFE model breaks preservation costs into elements of 'acquisition', 'ingest', 'metadata', 'storage', 'preservation actions' and 'access'. A list of heading like this could be developed and shared.
- In practice, archives that conform to agreed standards are relatively low cost: archives that do not conform to standards would be relatively high cost.
- Experience shows that the costs of ingest tend to be high so that anything which could reduce the costs of ingest would be welcome – such as automated deposit and quality assurance procedures.
- Experience also shows that the costs of building and maintaining an access systems were the most variable and that simply offering a 'one-size fits all' interface was unlikely to satisfy depositors who often times wanted bespoke or individualised access. Such mechanisms were also generally appreciated by users and were an important element in delivering impact from a project.
- Digital preservation was too often seen as a cost or a problem whereas in fact it is a capacity and an opportunity. This needs to be stated more plainly.
- The charging and processes associated with digital collections needs to align with the processes for the deposition of a physical archive, and therefore there is a need to communicate with museum curators.

After a discussion around these issues the group agreed on a number of action points to follow up:

- RCAHMW offered to raise the question of charging for the deposit of digital collections with senior management in an effort to create a level playing field for charging and to advise the others of the results of that conversation **(Action: RCAHMW)**
- Draft and publish a joint statement on the value of digital preservation stressing the need for early intervention at item level, the difference between preservation and storage, and the new types of capacity and opportunity they create **(Action: DPC to draft, appropriate officers in ADS, EH, RCAHMS, RCAHMW to review, endorse and publish)**
- Engage funders of research and persuade them to harmonise their expectations regarding the funding of preservation and the nature of charging policies **(Action: DPC to draft, appropriate officers in ADS, EH, RCAHMS, RCAHMW to review, endorse and publish)**
- Develop a joint cost model between repositories so that charges are comparable and thus more transparent **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**
- Undertake joint communication around charging policies for creators, curators and consultants, and connect charging policies explicitly to the minimum standards for depostition **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**
- Work together to provide verifiable research on preservation actions that improves the quality or reduce the costs of preservation, especially around ingest. **(Action: appropriate officers in ADS, DPC, EH, RCAHMS, RCAHMW)**

6. Closing remarks and next steps

The day ended with a number of comments about how to proceed from here. The following actions were agreed:

- Notes from the meeting will be drafted and circulated for comment and correction **(Action: DPC)**
- Participants should respond within two weeks to confirm that they agree the notes are an accurate reflection and how they intend to resource the actions allocated to them **(Action: appropriate officers in ADS, DPC EH, RCAHMS, RCAHMW)**
- The survey was noted as being of value when it comes to planning practical collaboration. Participants were agreed to review, update, expand or complete their responses to the survey in the same two week period **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**
- Participants agreed to examine the draft workplan and to make recommendations about how the headings in this section of the document might be progressed, by whom, with what resources and on what timescales. **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**
- DPC will collate and report the results of this process once comments and corrections have been received. Documents will be mounted on the members area of the DPC website **(Action: DPC)**

WK thanked participants for their time and enthusiasm. He reminded them of a number of forthcoming DPC events and closed the meeting at 1600.

7. Consolidated list of actions

7.1. Develop training resources for museum curators and development control officers to help them understand digital archives more fully and helping them write requirements into the planning process **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**

7.2. Encourage the early testing of archives to establish whether standards have been followed and to intervene in the development control process with sanctions if there are failings in this regard. **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**

7.3. Harmonise the standards for deposition between repositories **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**

7.4. Develop clear joint communications on the common standards for deposition, with an active outreach to fieldwork units **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**

7.5. Work with SMA and others to agree a joint statement of what constitutes a conformant submission package and develop a simple checklist which they can readily apply to validate incoming materials **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**

7.6. Issue a joint statement on archives, articulating their value and the opportunities associated with them, and aligning policies with relevant policy expectations regarding open data. **(Action: DPC to draft, appropriate officers in ADS, EH, RCAHMS, RCAHMW to review, endorse and publish)**

7.7. Persuade funders to promote and insist upon a common approach to preservation of research archives, making it easier for the researchers and others to understand and follow a shared requirement **(Action: appropriate officers in ADS, DPC, EH, RCAHMS, RCAHMW)**

7.8. RCAHMW offered to raise the question of charging for the deposit of digital collections with senior management in an effort to create a level playing field for charging and to advise the others of the results of that conversation **(Action: RCAHMW)**

7.9. Draft and publish a joint statement on the value of digital preservation stressing the need for early intervention at item level, the difference between preservation and storage, and the new types of capacity and opportunity they create **(Action: DPC to draft, appropriate officers in ADS, EH, RCAHMS, RCAHMW to review, endorse and publish)**

7.10. Engage funders of research and persuade them to harmonise their expectations regarding the funding of preservation and the nature of charging policies **(Action: DPC to draft, appropriate officers in ADS, EH, RCAHMS, RCAHMW to review, endorse and publish)**

7.11. Develop a joint cost model between repositories so that charges are comparable and thus more transparent **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**

7.12. Undertake joint communication around charging policies for creators, curators and consultants, and connect charging policies explicitly to the minimum standards for depostition **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**

7.13. Work together to provide verifiable research on preservation actions that improves the quality or reduce the costs of preservation, especially around ingest. **(Action: appropriate officers in ADS, DPC, EH, RCAHMS, RCAHMW)**

7.14. Notes from the meeting will be drafted and circulated for comment and correction **(Action: DPC)**

7.15. Participants should respond within two weeks to confirm that they agree the notes are an accurate reflection and how they intend to resource the actions allocated to them **(Action: appropriate officers in ADS, DPC EH, RCAHMS, RCAHMW)**

7.16. The survey was noted as being of value when it comes to planning practical collaboration. Participants were agreed to review, update, expand or complete their responses to the survey in the same two week period **(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)**

7.17. Participants agreed to examine the draft workplan and to make recommendations about how these items might be progressed, by whom, with what resources and on what timescales.

- connecting digital archives with fieldwork and other collections
- developing and clarifying cost models and charging policies
- automating and standardising the ingest of collections
- consolidating and clarifying 'upstream' advice for creators contractors and curators
- understanding users and measuring the costs and benefits of 'value added' use
- resolving perceived overlaps in collecting policies
- certifying and quality assuring preservation services
- developing skills for the sector and for digital preservation practitioners
- provision of 'cloud' storage or services in the cloud to facilitate preservation
- influencing industry and developers to provide preservation-ready systems

(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)

7.18. DPC will collate and report the results of this process once comments and corrections have been received. Documents will be mounted on the members area of the DPC website **(Action: DPC)**

8. Response from ADS

Catherine Hardman (ADS) welcomed these proposals. It would like to prioritize two sets of actions and offers to lead them. These two steps are priorities because the other recommendations are to some extent contingent on them.

- **Harmonise the standards for deposition between repositories:** many of the other actions flow from this one. CSH is happy to be the facilitator for this action
- **Develop a joint cost model between repositories so that charges are comparable and thus more transparent**: Again this is the key action from which other actions would flow, e.g. there is no point issuing joint statements until the cost model is developed. But we should be clear that a joint approach to cost modelling does not necessarily mean that the charges have to be the same across GB.

9. Response from English Heritage

Duncan Brown welcomed the proposals and report on behalf of English Heritage and method through which they had been derived. English Heritage has no specific additional comments at this time but supports progress towards the stated goals

10. Response from RCAHMS

Kirsty Lingstadt (RCAHMS) welcome the proposals and offered a timetable of actions.

Phase one – statements of intent

1. Issue a joint statement on archives, articulating their value and the opportunities associated with them, and aligning policies with relevant policy expectations regarding open data. (Action: DPC to draft, appropriate officers in ADS, EH, RCAHMS, RCAHMW to review, endorse and publish)
2. Draft and publish a joint statement on the value of digital preservation stressing the need for early intervention at item level, the difference between preservation and storage, and the new types of capacity and opportunity they create (Action: DPC to draft, appropriate officers in ADS, EH, RCAHMS, RCAHMW to review, endorse and publish)

Timescale: within 4-6 months

Phase two – practical work

3. Deposit

a) Harmonise the standards for deposition between repositories (Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)

b) Develop clear joint communications on the common standards for deposition, with an active outreach to fieldwork units (Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)

4. Cost:

c) Develop a joint cost model between repositories so that charges are comparable and thus more transparent(Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)

Timescale: within 6-9 months

Phase 3 – engaging other bodies with, and disseminating the results of the practical work

5. Deposit

a) Work with SMA and others to agree a joint statement of what constitutes a conformant submission package and develop a simple checklist which they can readily apply to validate incoming materials (Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)

b) Develop training resources for museum curators and development control officers to help them understand digital archives more fully and helping them write requirements into the planning process (Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)

c) Encourage the early testing of archives to establish whether standards have been followed and to intervene in the development control process with sanctions if there are failings in this regard. (Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)

6. Cost

a) Undertake joint communication around charging policies for creators, curators and consultants, and connect charging policies explicitly to the minimum standards for deposition (Action: appropriate officers in ADS, EH, RCAHMS, RCAHMW)

7. Funders

a) Persuade funders to promote and insist upon a common approach to preservation of research archives, making it easier for the researchers and others to understand and follow a shared requirement (Action: appropriate officers in ADS, DPC, EH, RCAHMS, RCAHMW)

b) Engage funders of research and persuade them to harmonise their expectations regarding the funding of preservation and the nature of charging policies (Action: DPC to draft, appropriate officers in ADS, EH, RCAHMS, RCAHMW to review, endorse and publish)

Timescale: 12 months +

General

8. Documentation and Standards

a) There are a number of these to be pulled together for 3a, 5a and 7a and a useful approach to this would be to undertake an audit of existing standards and arrange for these to be circulated with a view to adapting these to meet the groups needs. (Action: identify who should investigate standards for review by appropriate officers in ADS, EH, RCAHMS, RCAHMW)

9. Costs and charging policies

a) This will require each of the organisations to engage with key funders and have initial discussions to gage response to these proposals before further detailed work can be undertaken.

Timescale: within 6 months

11. RCAHMW

Gareth Edwards (RCAHMW) welcomed the proposals, the report and the positive meeting which gave rise to them. He requested a number of specific changes to how the report presented the situation in Wales which needed to be included before the report could be progressed further. These have been incorporated in the final version.

12. About this document

Version 1	Document initiated by WK and distributed to members as draft	30/09/2011	WK
Version 2	Updated with comments and distributed	04/10/2011	WK
Version 3	Return individual survey responses	13/10/2011	SJ, EW, ME
Version 4	Compile responses and distribute to members	14/10/2011	WK
Version 5	Draft notes from meeting added and distributed for approval	19/10/2011	WK
Version 6	*Updated with comments and approved as final copy*		

1. Appendix: Preliminary Survey

In order to make the most efficient use of time, members were asked to prepare a short overview of their digital preservation processes. This analysis was based on the functional areas of the Reference Model for an Open Archival Information System (OAIS), an ISO standard which assembles the building blocks of a long term preservation facility. The standard assumes the existence of six functional areas: Ingest, Archival Storage, Data Management, Administration, Planning and Access. A seventh area – Common Services – provides the local platform on which an OAIS is constructed. Each functional area in OAIS is composed of a group of components which work together to provide preservation services. Although the function of each component and the relationships between components are described, the standard offers only limited guidance on how any component is configured. From an organisational perspective, OAIS outlines a preservation architecture and common vocabulary, but it makes no assumption on whether the functional components are offered by a single agency or are shared between multiple partners, or outsourced. Therefore, even a preliminary mapping against OAIS enables agencies to plan how they might contract services, outsource them or collaborate as appropriate.

This appendix compiles all three responses received prior to the meeting from ADS, English Heritage and RCAHMS. It was intended as an informal guide to help the meeting rather than a comprehensive statement but its utility in helping to plan practical collaboration was noted. Participants assessed their own strengths and areas for improvement against each functional component of OAIS, identified whether they think there is scope for collaboration in this area, outlined the sort of collaboration they thought possible and what they would hope to gain from it. Answers for each functional area have been compiled together.

2. More details

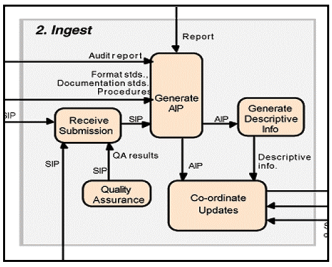
For more details on the OAIS standard, a good place to start is the DPC Technology Watch Report by Brian Lavoie of OCLC online at: http://www.dpconline.org/component/docman/doc_download/91-introduction-to-oais

Survey of Digital Archive Functions

3. Introductions

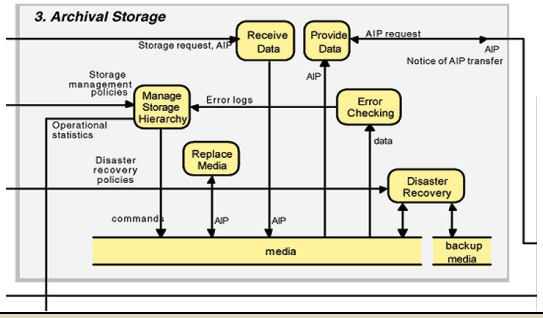
Archaeology Data Service	English Heritage	RCAHMS
http://archaeologydataservice.ac.uk	http://www.english-heritage.org.uk/	www.rcahms.gov.uk
Stuart Jeffrey	Mike Evans	Emily Nimmo and Kirsty Lingstadt
Stuart.jeffrey@york.ac.uk	Mike.evans@english-heritage.org.uk	Emily.nimmo@rcahms.gov.uk; kirsty.lingstadt@rcahms.gov.uk
Wednesday 12th October	7/10/11	12.10.2011

4. Ingest

The 'Ingest' functional area is the interface between the OAIS and its producers and is primarily concerned with the taming of content and preparation for transfer into the archive. It includes five components: 'Receive Submission; Quality Assurance; Generate Archival Information Package; Generate Descriptive Information; and Co-ordinate Updates.		
Please briefly describe the typical elements of your ingest processes?		
We follow the OAIS reference model so cover all the elements above using our internal Collections Management System. Our ingest manual is available here: http://archaeologydataservice.ac.uk/advice/preservation/	We have an ingest system targeted primarily at our own internal research teams: • We get a submission spreadsheet from the photographic or research teams, to a specified format • QA of information on spreadsheet carried out, but no QA of embedded metadata • Descriptive information then developed in our catalogue (AMIE), within 6 weeks	• Negotiate submission • Receive submission • Virus check • Transfer Data and Metadata to Temporary Storage Area • Audit/Appraise Data and Metadata • Complete Accession Record • Transfer Data and Metadata to Digital Archive • Generate and Send Acknowledgement Letter • Create Catalogue Records with Digital Instances • File Paperwork • Store Original Media in Negative Room
Which parts of your ingest processes do you think are particularly strong?		
The ADS puts a lot of effort into negotiation with depositors to ensure a well formed SIP – this is time consuming, but essential.	• Submission spreadsheet is robust and works reasonably well with depositors • Good flow through to descriptive information, ensuring no backlog	• Negotiation with depositors • Deposit agreements and licenses
Which parts of your ingest processes would you like to improve given the resources?		
Streamlining of this process via a more automated ingest, particularly with regard to controlled vocabularies and data formats.	• Need to develop processes to cope with external deposits • Increase amount of automatically generated metadata and reduce need for handcrafting • Improve communication with depositors to establish proper	• Quarantine of data and metadata before virus check • Creation and (automatic) checking of fixity values • Transfer of metadata in a machine-readable format

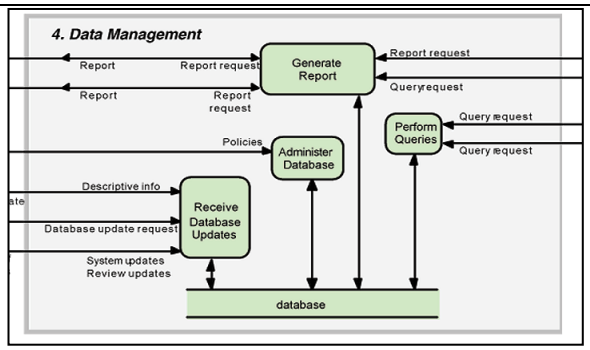
	audit trail	• Automation of file transfer, file structure, naming and migration to preservation format • Automation of cataloguing
Are there any elements of the ingest process which you think would be improved through collaboration?		
Collaboration with LA and Museum Services have led to some streamlining of this process. However collaboration with other repositories in the form of the SIP would ultimately make it easier to aggregate resource discovery metadata (i.e. all deposits, wherever held, are formed in mutually intelligible and searchable forms).	• Possibly development of solutions for automating metadata generation (both at submission point and as part of generating descriptive info) • Standardisation of ingest requirements/formats could help encourage deposit by external depositors	• Sharing examples of best practice in relation to all areas of ingests process. • Developing automation of processes • Standardization of accepted file formats and required metadata
What would you hope to gain from such collaboration?		
Archive portability (a sustainability benefit) and potentially increased visibility of archives through external metadata aggregators.	• Access to technical expertise • Improvement of deposit across sector	• Benefit from experience of others, avoid issues they may have experienced when undertaking similar work. • Economies to be found sharing resources to address issues relevant to all. • Standardization of ingest requirements will facilitate more efficient negotiation and transfer from depositors and hopefully an increase in compliance.
Are there any elements of the ingest process which can only be carried out locally?		
Negotiation can currently only be done per repository as each has its own metadata requirements and archival storage procedures.	• QA in any but the most technical sense benefits from a relationship with record creators and an understanding of their work. • Generating descriptive info is for us integrated with our cataloguing of non-digital material, which has to be local	• Negotiation • Appraisal • Accessioning • Cataloguing • Storage

5. Archival Storage

The Archival Storage area is oriented around the management of robust storage, placing data on media, ensuring the integrity of data stored and recovering data from media as required. It includes six functional components: Receive Data, Provide Data, Error Checking, Disaster Recovery, Replace Media, Manage Storage.		
Please briefly describe the typical operation of your Archival Storage?		
All elements mentioned above are again covered by the ADS including data integrity (e.g. MD5), deep storage on and off site and a well formed disaster recovery plan.	Files are stored on a raided disk array, managed by an off-the-shelf digital asset management system Portfolio v9.5. Files are also backed up on 2 sets of hard drives created at time of ingest. There is no systematic system for error checking. Disaster recovery is via our outsourced IT providers – but is based on the hard drives and a tape of the Portfolio indexes, not a tape of the whole system	• All digital archive material will be retained permanently on multiple drives within RCAHMS' network storage. This storage area is read-only to RCAHMS staff except those who are directly involved in accessioning and cataloguing the digital archive • As part of pre accessioning of externally generated material this data is copied to a temporary archive location for evaluation. A temporary storage location for evaluation of internally generated RCAHMS material is provided on network storage so that it is accessible to creators and the digital archivist. • A major upgrade of RCAHMS storage systems timetabled for 2012/13 that will enable the creation of 8Tb volumes • The current storage solution (EVA SAN) is out of warranty in 2012 and will either need to be replaced or storage moved to the cloud • A daily incremental backup is performed alongside a weekly and monthly full backup. Back-ups are written to disc array and tapes which are stored on site in a fire proof safe. Off-site storage is currently under negotiation. • A rolling program of storage media re-fresh is performed over a 3-5 year timetable dependent on project pressures, budget and warranty expiration.

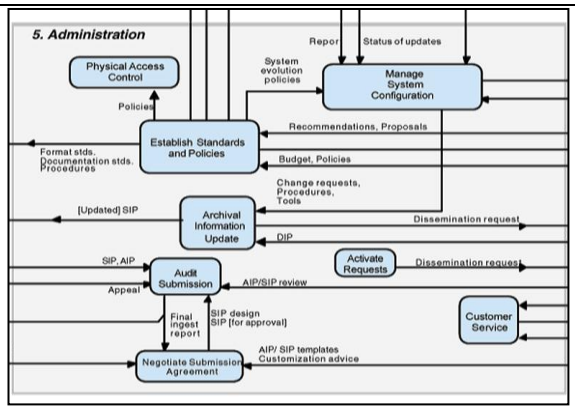
Which parts of your archival storage operations do you think are particularly strong?		
Deep storage is provided both by UoY systems and by the UKDA. The ADS are very committed to ensuring migration of the AIP to non-proprietary formats for long term preservation.	It has worked!?	• Access control • Back-up • Disaster recovery
Which parts would you like to improve given the resources?		
ADS are in the process of moving to a FEDORA based repository which has a number of technical advantages.	• The backup mechanism is not robust or fast. The system has survived a major failure of the raid array, but it took over a month to fully restore. I would like to move to full a mirrored system • System needs to be fully scaleable to allow for ongoing growth • We need a data checking regime	• We are investigating utilizing cloud storage for backup and off-site storage of tapes. • The planned upgrade will facilitate a separation of archival and dissemination copies of digital objects which will aid performance.
Are there any elements of the archival storage which you think would be improved through collaboration?		
Consensus on migration paths and preservation formats would improve alignment between repositories.	• Possibly sharing of backups to improve survivability • Possibly storage of large specialist file types and/or rarely used data	Mirroring of data
What would you hope to gain from such collaboration?		
Repositories could cooperate on disaster recovery procedures.	• Improved survivability • Use of shared expertise to preserve particularly technical or complex data • More cost effective storage, near or off line	Lowering the cost of externally supplied storage and minimizing the risks of any data loss through multiple copies.
Are there any elements of archival storage which can only be carried out locally?		
Local back-up, storage management and disaster recovery.	• QA in any but the most technical sense benefits from a relationship with record creators and an understanding of their work. • Generating descriptive info is for us integrated with our cataloguing of non-digital material, which has to be local	At least one networked copy of data

6. Data Management

The Data Management functions are primarily concerned with ensuring that descriptive information about the contents of the archive is maintained and made available for internal administrative and reporting. In this context 'data' is distinguished from the information packages which the archive receives, preserves or distributes. This functional area is therefore oriented around a collection management database. It includes four components: Receive Database Updates, Administer Database, Perform Queries, and Generate Reports.		
Please briefly describe the typical elements of data management in your preservation facilities?		
The ADS has a well-developed CMS that allows us to perform all the stated data management functions, including the tracking of migration events.	Our cataloguing system AMIE is used as a collections management database in the sense of holding descriptions of archive packages, allowing queries and reports to be run	• Descriptive metadata identifying and describing the collection of archived material for RCAHMS both physical and digital is stored in the catalogue tables of an Oracle database. • Metadata recorded is ISAD-G compliant and it is supplemented with additional 'Digital Instance' information relating to individual files • The database is primarily structured around geographical location and also collection.
Which parts of your data management processes do you think are particularly strong?		
The ADS CMS is fairly integrated in that it tracks collections, people (and organizations) and objects from the point at which a negotiation is opened with a depositor right through the preservation lifecycle. The CMS is also the (live) source of the majority of data used in the ADS delivery system, i.e. the website is generated dynamically from CMS content.	• We use robust data standards, which are common to other EH systems and help support cross searching of descriptive info • Our query tools are powerful (but not very user friendly) • Our cataloguing systems integrate descriptive records for digital and non-digital material	Robust, well-structured database system adhering to established standards.
Which parts of your data management processes would you like to improve given the resources?		
Closer integration between CMS and delivery – particularly with regard to the creation of web services (this is in train). There is also the potential for conversion to LD formats for appropriate datasets – but perhaps this is best pushed back to the ingest/negotiation function	AMIE and the DAMS Portfolio are not properly integrated, leading to them falling out of step, the double handling of information, etc etc. They need to work much more closely together.	Increased commonality between descriptive datasets, data annotation and improvement.

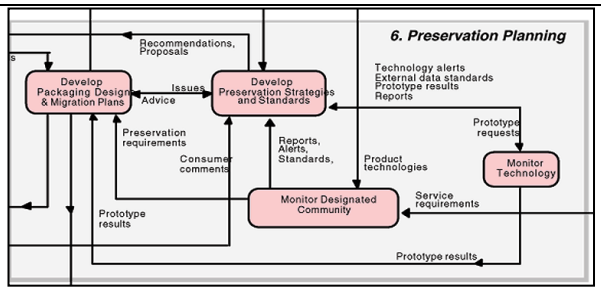
Are there any elements of the data management process which you think would be improved through collaboration?		
No answer	No answer	Aligning practices to facilitate interoperability where this could enhance user interaction with collections. Utilizing the OASIS submission process to capture user generated metadata. This could expedite ingest and cataloguing of digital materials and eliminate possible duplication of effort for our depositors.
What would you hope to gain from such collaboration?		
This would be contingent on a more automated ingest system and consensus on AIP form and other data standards, I think.		Enhanced user and depositor experience Efficiencies in time and/or money.
Are there any elements of your data management processes which can only be carried out locally?		
Even with full collaboration on data standards, automated ingest and functions such as disaster recovery – part of the definition of a repository would be its ability to carry out some degree of data management locally.	Specialist nature of subject matter and integration with non-digital holdings argue for local processes	Some scope for data upgrading to be carried out externally through an online interface (see SURE project), other elements must be managed within RCAHMS.

7. Administration

The Administration functions ensure that the OAIS remains aligned with the goals of the agencies which sponsor it. It is a relatively complex area and interfaces with technology and resources as well with the administrative relationships an archive is required to develop with consumers and producers. It includes eight components: Physical Access Control, Establish Standards and Policies, Manage System Configuration, Archival Information Update, Audit Submission, Negotiate Submission Agreement, Activate Requests and Customer Service.		
Please briefly describe the typical elements of Administration in your preservation facilities?		
As above, the ADS administers ingest and data management via its CMS and associated procedures.	• Our standards and policies are incomplete and the process for managing them currently adhoc and not properly integrated with wider EH management structures. • Reasonable system for controlling access to files by allowing users different levels of access using passwords • We have good mechanisms for negotiating submission agreements through a dedicated flowlines post – but primarily for internal deposits as previously noted.	The administrative function of our digital archive is performed by the Digital Archivist reporting to the Operational Manager for collections and Head of collections, alongside a steering committee for the development of a Trusted Digital Repository. The Digital Archive Policy sets out objectives and responsibilities. The Digital Archivist maintains open lines of communication with external depositors, undertaking user education and facilitating negotiation of submission agreements alongside implementing and maintaining archive policies and standards.
Which parts of your Administration do you think are particularly strong?		
It is a real strength of the ADS that it has well developed administration policies and process documentation. It is also a strength that this is all made freely available to depositors/potential auditors (e.g. DSA). The ADS are very involved in data standards development nationally and internationally - not necessarily a core repository role, but essential where this is not being done by other bodies. We make great efforts to adopt and follow existing standards where they exist. We are also involved more generally in archiving policy (e.g. AAF, EAC) Being University based is a real advantage for the ADS , both in terms of international contacts, but also in terms of access to	Negotiation of submission agreements and policies around this	Oversight, inclusion in RCAHMS' strategic plan and digital archive policy.

national and international research funding to develop procedures and infrastructure.		
Which parts of your Administration would you like to improve given the resources?		
Customer service – in the sense that the repository should not rely on its own access points, but engage fully with the wider range of organizations, projects and initiatives which provide access to distributed data.	We need to put more structured and systematic management of the system in place – to support both adoption of standards and policies and a joined up approach to systems configuration	• Greater granularity in policies and standards. • Clearer definition of management structure and budget.
Are there any elements of the Administration which you think would be improved through collaboration?		
A real need for agreement on licensing agreements arises when thinking about sharing data (even metadata). Since the inception of the ADS the licensing environment has moved on and a sector wide consensus on the use of, say, CC licensing, at least for metadata would be very useful and avoid confusion for users.	A common approach to core standards and policies	• Sharing examples of best practice, examples of policy/planning documents. • Greater consistency across policies and standards of participating organizations.
What would you hope to gain from such collaboration?		
Clarity for users (and also for depositors and repositories!)	A consensus would help with making the case locally for resources to support this.	Benefit from experience of others, avoid issues they may have experienced when undertaking similar work. Help to eliminate the duplication of effort and also ensure the most robust policies possible through critical peer review and wider range of expertise in-putting in the process.
Are there any elements of Administration which can only be carried out locally?		
As with data management there is a core subset of administrative functions that need to be carried out locally – this is especially true without the sector wide consensus on data standards, formats and licensing mentioned above.	Establishment of standards and policies has to align with local business priorities to get management buy –in	At this time all of the administration function can only be carried out locally, although there is scope for collaboration as outlined above.

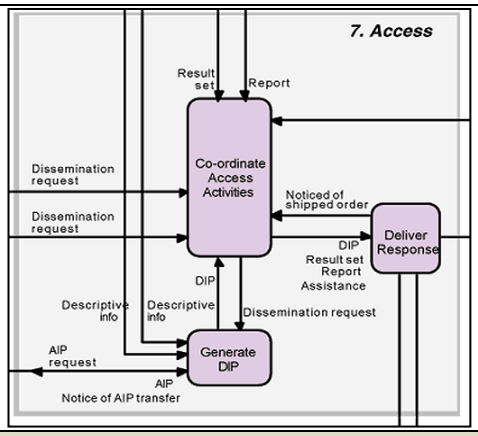
8. Preservation Planning

The Preservation Planning functions are primarily concerned with ensuring that the actions and standards followed by the OAIS remain current and effective through time, and that any actions are assessed and validated where possible. It includes four components: Monitor Technology, Monitor Designated Community, Develop Preservation Strategies and Standards, and Develop Packaging Design and Migration Plans. OAIS makes only rudimentary recommendations for what ought to be included in a preservation plan.		
Please briefly describe the typical elements of preservation planning in your preservation facilities?		
Apart from OAIS recommendations the ADS has worked hard to develop (and publish) data preservation plans covering the elements above. See: http://archaeologydataservice.ac.uk/advice/preservation	• There is adhoc identification of formats or media at risk – often sparked by a particular business need. Eg we are transferring Images of England project digital files from CD to disk array to improve accessibility as well as preservation. • We have good links with internal depositors and regularly discuss preservation and dissemination requirements eg for GIS and websites.	We are still in the planning and development stages of our TDR and have not yet agreed a preservation planning strategy.
Which parts of preservation planning do you think are particularly strong?		
Monitoring of technologies, again it is a strength of the ADS that our technical team are active in the repository community more broadly both in technical discussions and in management practice (e.g. DPC, RDMF), (although there are not always the resources to respond immediately to changes in technology)	Links with internal depositors	N/A
Which parts of preservation planning would you like to improve given the resources?		
Migration plans are more easily created than enacted given other pressures on resources.	Systematic review of formats/media ie monitor technology	Implementation of a preservation plan for each object type.
Are there any elements of preservation planning which you think would be improved through collaboration?		
Consensus on migration paths and standards. Also many (sector specific - i.e. Cultural Heritage) repositories share the same of similar designated communities and given that monitoring these is a specialist function in itself this could be shared between repositories.	Monitoring technology and development of preservation strategies	Sharing of existing preservation plans and migration strategies and systems. Collaborative development where policies/systems are not in place for common formats and object types.

What would you hope to gain from such collaboration?		
Shared responsibility for migration paths and standards as well as community monitoring could reduce the pressure on resources.	Sharing information would give access to expertise. Seeking consensus on preservation strategies would help develop understanding of options and implications – even if we didn't all come to same conclusion	Help to eliminate the duplication of effort and also ensure the most robust policies possible through critical peer review and wider range of expertise in-putting in the process.
Are there any elements of preservation planning which can only be carried out locally?		
Those elements of preservation planning that are contingent on local infrastructure will always need to be done locally.		RCAHMS deals with a variety of materials which may be out of scope for other organizations making collaboration in those areas inappropriate.

9. Access

The Access functions manage the relationship between the OAIS and its consumers. In doing so it interfaces with the administration functions and archival storage. The scale and nature of access is not defined so the consumer could be broker who provides access to the public, a customer, or the public via the Internet. There are only three components: Generate Dissemination Information Package, Co-ordinate Access Activities and Deliver Response.



Please briefly describe the access functions within your preservation facilities?

ADS delivers all its data on-line via its website. The website is complex and utilizes some sophisticated technologies (e.g. NLP and faceted classification). The DIP is generated from the resource discovery metadata developed at ingest and held in the CMS.	Access to digital archives is either via an inquiry to our Research Service staff or our websites (Viewfinder, EH Archives). Websites provide access to versions of a limited range of material. Some records can only be obtained via staff	Canmore is the web front end of the RCAHMS database. It is at the heart of the RCAHMS archive, providing searchable, map-based information on over 280,000 buildings and archaeological sites throughout Scotland, as well as a catalogue of the collection items held. Over 130,000 digital images are available to browse and purchase online. It brings together the results of the survey and collections material into one place and combines location information, site details and images on more than 300,000 archaeological, architectural, maritime and industrial sites throughout Scotland. Canmore offers users the ability to: • Conduct searches and advanced searches to find information on specific sites • Identify where types of site are located • Search for digital images on sites or locations throughout Scotland • Collect direct references to specific sites from a wide range of textbooks and journals • Know which of these we hold in our Search Room • Users can now share their own information and images with the national collection by registering with Canmore

		Access to RCAHMS' Collections material through Canmore will be radically improved by planned developments within the SWISH program. This will include access to hierarchical catalogue records, PDFs files and dissemination surrogates for CAD drawings.
Which parts of access do you think are particularly strong?		
We have developed a search interface and supporting systems intended to make discovery and access easier – however this could still be enhanced (always the case). Our development of web services, bibliographic, Monument Inventory and archive based will be a key strength in the future – potentially delivery of LoD datasets also.	We provide effective retrieval of material. Both ERS staff and websites provide good searching facilities, with fast delivery of well documented files	• Variety of search facilities • Delivery of digital images. • User engagement through adding their own images to a Flickr collection
Which parts of access would you like to improve given the resources?		
Online discovery and access can always be improved.	• The relationship between our catalogue (AMIE), the DAMS (Portfolio) and the Catalogue needs to be much more integrated to allow for quick automatic updating of content • Development of automated on-line delivery mechanisms for a full range of files , including sale and licensing options where appropriate • Development of "scan-on–demand" system for creating digital surrogates for analogue material	• Further innovative ways for our depositors and consumers to interact with our collections. • Automated generation of dissemination surrogates for more object types, access to original data.
Are there any elements of access which you think would be improved through collaboration?		
Yes, in a dream scenario, where data standards and shared repository processes are well developed then cross-searching of repositories and deep and stable linking of distributed datasets could benefit all repositories – essentially created a seamless pool of mutually intelligible data sets wherever they are held.	Only where a partner can provide access to specialist expertise (eg for a certain type of archive) or to a specialist audience	Cross organization/ collection search facilities might bring benefits for consumers. Automated generation of dissemination surrogates.
What would you hope to gain from such collaboration?		
The ADS is neutral on branding of datasets and careful to appropriately attribute data they present that is ultimately drawn		Enhancement of the user experience

from other sources. So even where collaboration is desirable in terms of shared infrastructure, the 'public face', and access modes for data are not our key concern – except where there the deposition of data is funded directly by an organization with a very specific remit and audience.(e.g. via a research council)		
Are there any elements of access which can only be carried out locally?		
Certain types of access to ADS data (i.e.) via the ArchSearch interface cannot be replicated elsewhere so would not be abandoned even though broadcast of data availability via web-service will become more widespread.		e-commerce

10. Any other comments

Are there any other areas of strength you would like to mention?	None
Are there any other areas for improvement you would like to mention	None
Are there any other areas for collaboration you would like to propose	None
Do you think this short survey has been useful? What should we do with the results?	None
Is there anything else you would like to add?	None