

INFORMATION STANDARDS QUARTERLY

FALL 2013 | VOL 25 | ISSUE 3 | ISSN 1041-0031

ISQ

TOPIC

DATA CURATION

DATA CURATION ISSUES IN THE
CHEMICAL SCIENCES

DATA CURATION IN THE OPENAIRE
SCHOLARLY COMMUNICATION
INFRASTRUCTURE

PRESERVING THE GREY
LITERATURE EXPLOSION:
PDF/A AND THE DIGITAL ARCHIVE

ENSURING THE LONG TERM IMPACT OF
EARTH SCIENCE DATA THROUGH DATA
CURATION AND PRESERVATION



Ray
Moore



Tim
Evans

Preserving the Grey Literature Explosion: PDF/A and the Digital Archive

RAY MOORE AND TIM EVANS



Founded in 1996 the *Archaeology Data Service* (ADS)^[1] was established as one of five disciplinary data centers, under the auspices of *Arts and Humanities Data Service* (AHDS),^[2] to provide specialist advice and expertise during the lifecycle of digital data from creation, through preservation, and onward to its potential reuse.

The ADS provides support for research, learning, and teaching within the archaeological sector, providing freely available, high quality and reliable digital resources that can be preserved and disseminated in the long-term. In addition, subject specific expertise has allowed the ADS to aid the sector in the creation and documentation of digital data, through projects such as the *Guides to Good Practice*,^[3] recently updated in collaboration with the Digital Antiquity consortium in the United States. Consequently the focus of the ADS has always been on preserving high quality, well-documented data that holds the greatest potential for reuse.

The ADS works with local and national agencies within governmental, research, and commercial environments, and acts as a digital repository for organizations including the Natural Environment Research Council, the Arts and Humanities Research Council, and the British Academy and English Heritage. In addition the ADS functions as a data broker facilitating the exposure and accessibility of existing datasets; the ADS online catalog *ArchSearch*,^[4] for example,

provides access to monument inventories produced by over 30 regional and national agencies within Britain. One of the key roles of the ADS, however, is the preservation and dissemination of grey literature produced as a consequence of archaeological fieldwork carried out during the planning process. Working with partners, the ADS hosts the OASIS project^[5] and the associated Grey Literature Library^[6] that, respectively, enable the recording of fieldwork activities within England and Scotland, and also provide direct access to over 20,000 unpublished reports, produced by some 140 commercial contracting units and researchers working within Britain (discussed below).

In part, the work of the ADS reflects the complex nature of an archaeological profession that embraces local and national governmental agencies, museums, and councils alongside “traditional” research and academic environments; but typically also includes companies and organizations working within both the commercial and voluntary sectors. At the same time, the scope of research carried out under

the banner of “archaeology” is varied and comprises a diverse range of specialist and related fields that spans the disciplinary boundaries from history through the social sciences and into the “hard” sciences. From the outset, the ADS has embraced this broad community with a collection policy that covers the full spectrum of archaeological research. As a result, an archive is just as likely to include a strontium isotope analysis of human remains from Roman Britain (doi: 10.5284/1000405) as a 3D laser scan of the prehistoric rock art of Yorkshire (doi: 10.5284/1000092). These diverse techniques and methodologies produce an equally broad range of digital outputs. An archaeological excavation may produce databases, spreadsheets, CAD plans, and GIS files alongside standard desktop publishing and image formats. Add to these the outputs of landscape and geophysical surveys and it is easy to see the difficulty faced in archiving digital data produced during archaeological research. Yet despite this complexity, the most common formats remain documents and reports that are predominantly deposited in the Portable Document Format (PDF). By providing an account of the work of the ADS, and specifically its experiences in the curation of a large collection of grey literature, it is hoped to report on some of the problems and issues associated with archiving digital content in the PDF format. With an eye on the future we hope to provide some insight into the impact of the development of PDF/A-3 on the archiving and preservation communities.

OASIS and the ADS Grey Literature Library

The vast majority of all PDF files accessioned by the ADS are deposited through the OASIS project, borne out of a partnership in 2000 between the ADS, the Archaeological Investigations Project (AIP)^[7] of Bournemouth University, and English Heritage.^[8] The rationale behind the original project was the increasing need to provide an online index to the mass of archaeological grey literature produced as a result of the advent of large-scale developer-funded fieldwork in the country. The issue of grey literature in archaeology is an important and sometimes controversial one;^[9] the existence of so much unpublished and inaccessible information is anathema to any knowledge-based discipline, but especially one where the study of the resource often entails its destruction. Indeed, since the advent of commercially funded archaeology in England in 1990, it is estimated that there are up to 4000 individual archaeological events in any year,^[10] each potentially producing an unpublished report of the results that is lodged with a

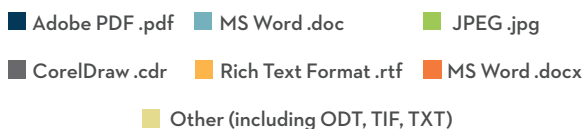
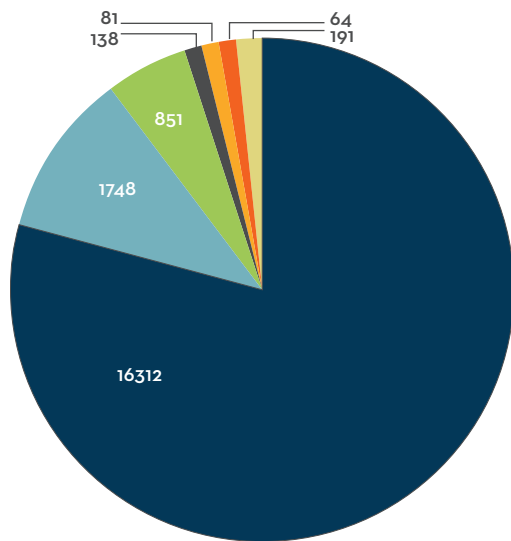
local Historic Environment Record (HER). Originally, these reports would have been produced in hard copy only, but with the advent of wide-scale computing it was common for the reports to exist in both physical and digital formats and increasingly in digital form only.^[11]

Thus, the OASIS system was introduced in England in 2001, and thereafter in Scotland (in partnership with RCHAMS and Historic Scotland^[12]) in 2005. The system itself consists of an online data form that can be used by those involved in archaeological fieldwork to capture and record the data they gather in the course of their investigations. The form can then be submitted to the local HER where information is validated, typically by the HER Officer, and exported to the local archive. In addition, as of 2005, any unpublished or grey literature in digital form can be uploaded with the relevant record, allowing the simple transfer of the digital report to both the HER and the relevant National Monument Record. Although primarily designed to collect information about developer-funded archaeological fieldwork, the system can also be used to record other archaeological activity, such as desk-based assessments, building recording, scientific dating such as dendrochronology or radiocarbon, or events associated with the maritime environment. It also extends to include fieldwork or research undertaken as part of academic research projects as well as the activities and findings of volunteer community groups—in fact the whole spectrum of archaeological work being undertaken in England and Scotland at both a site and landscape level.

At the time of writing, there are currently 466 organizations and individuals signed up to the system, with 36,415 records currently either awaiting completion, being validated, or completed, with a total of 20,785 digital files uploaded (Figure 1). Of this larger total, 19,047 records (representing 292 of the 466 organizations) have been completed and signed off by the relevant National Monument Record. Once a record has been completed (i.e., the report and OASIS metadata has been checked), any attached digital report can be transferred to the ADS archive, assigned a Digital Object Identifier, and disseminated through the Grey Literature Library interface. The transferral of reports with a subset of the OASIS metadata (to aid resource discovery) began in piecemeal fashion early in 2005, but began in earnest with the implementation of an automated transfer system in 2008, and with the implementation of DOIs in 2011. At the time of writing 14,265 OASIS records have been transferred from OASIS to the Grey Literature Library, representing 19,385

CONTINUED »

Figure 1: File types uploaded to OASIS and accessioned by the ADS (as of 01/08/2013)



individual files, 185 ADS collections, and 1512 accessions. It should be noted that the library also includes just fewer than 6000 reports accessioned from other projects, such as scanning of physical grey literature.

The scope of these files is a fascinating cross-section across modern archaeological practice in England and Scotland. One can view the results of a survey of a twentieth century gun emplacement on the Isles of Scilly (doi: 10.5284/1010773), an excavation of a Romano-British cemetery in Cambridgeshire (doi: 10.5284/1001160), or a laser scan and geophysical survey of a Neolithic Stone Circle in Perth and Kinross (doi: 10.5284/1019599). Given the high profile of the “grey literature problem” in UK archaeology—it should be noted that the issues of commercial archaeology and unpublished outputs are not restricted to the UK but in fact are mirrored in Europe and beyond^[13]—it is no surprise that this corpus has attracted significant attention from the archaeological community. This is evident in a number of high profile research projects that have concentrated solely on what this previously unheralded resource can inform us about particular geographic locales or archaeological periods.^[14] It arguably represents a shift away from research based solely on the traditional textbooks and

monographs that focus on the sites of national or international significance, and towards an increased emphasis on a corpus that contains the details of the commonplace.

The ADS has also enhanced access to the grey literature through sharing the resource discovery metadata through portals, most notably Europeana,^[15] thus enabling records to be cross-searched with cultural and scientific heritage collections across Europe. Furthermore, the use of grey literature is not purely restricted to traditional research; recent initiatives have used the ADS Grey Literature corpus as the basis for natural language processing rich semantic indexing of grey literature documents.^[16] Herein is the unique potential of grey literature; far from being the under-used and maligned resource of old, the ADS Library provides an accessible and rich corpus to be used alongside the traditional hard-copy monographs, as well as providing semantic annotation and cross-searching far beyond any physical library search. Clearly, our grey literature is an essential part of the fabric of modern archaeological practice and research, and it is perhaps not an overstatement to say that digital grey literature has moved from being a challenge to an opportunity for the discipline and beyond.

However, of further interest to the ADS is the digital content of these files. Increasingly, as software capabilities of archaeological practitioners become more varied, the final reports from fieldwork incorporate a raft of data types aside from plain text. For example, it is not unusual for a report from a moderately sized archaeological “event” to include a full text description, raster images (color and grayscale), tabular data imported from software such as Access or Excel, and vector data imported from CAD or GIS programs. From an overview of all files uploaded to OASIS, it is clear that the vast majority of these reports are transferred in the PDF format (Figure 1). As well as data content, it is noticeable that as grey literature becomes more accessible via use of the Web, the reports show an increased amount of design and artwork, undoubtedly to showcase the skill and professional standards of the organization responsible. Thus, perhaps ironically, what pre-OASIS would have been short text documents with little or no graphical or stylistic output have now become cutting-edge, sophisticated documents with multiple types of data incorporated (Figure 2). To achieve a secure and sustainable archival version of this report, the digital archivist is faced with a far from simple proposition and must begin to understand the development of the PDF and PDF/A formats.

PDF and PDF/A

Developed by Adobe Systems from the PostScript image file format, the PDF specification was made available from

1993, but remained a proprietary format until 2008 when it was released as an open standard and made available as an ISO (International Organization for Standardization) standard (ISO 32000-1:2008).^[17] The PDF was created as a digital format for representing documents. Like most formats, the PDF has developed somewhat organically—adapting to changing technologies and requirements from amongst its core users. Initially, the format was created for use within the desktop publishing industry as a means for users to share and view documents, which incorporated both text and images created using a variety of different software, in an unchangeable manner across disparate platforms and viewed independently from the environment in which they were created. It is this functionality that has led the format to become the prevailing standard within the publishing industry, as well as the *de facto* form for sharing documents through the Internet. So popular has the format become that many commercial businesses, governments, and other institutions maintain large collections of important information within the PDF format. Many of these documents need to be accessible for considerable periods of time, while others will require permanent preservation. A solution for the preservation of PDF documents was devised through the collaboration of partners within the records

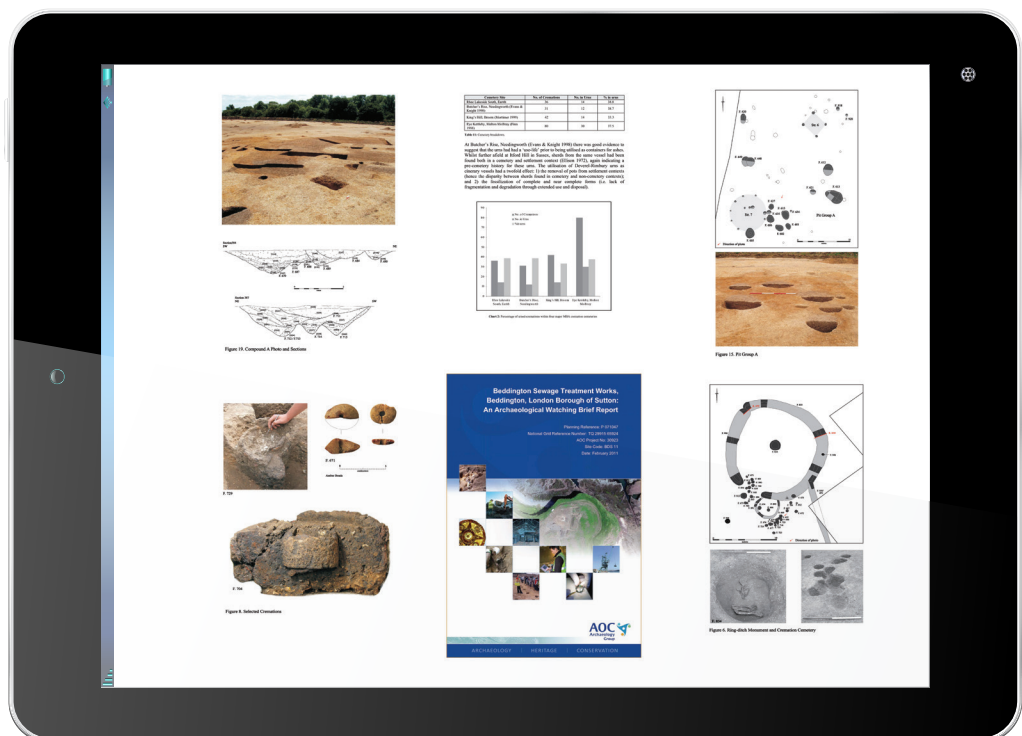
management and archival communities, which was identified as PDF/A (archival).^[18] In essence, PDF/A-1 (ISO 19005-1:2005)^[19] builds upon the PDF 1.4 specification by providing a mechanism for representing digital content in a form that maintains the visual appearance of the electronic document, independent from any hardware and software used to create, store, and render the file. To achieve this, PDF/A-1 embeds all fonts and metadata within the file so that it can be consistently rendered regardless of the hardware and software used to create or view it. Consequently the specification prohibits the use of transparency and encryption, meaning that the file can be easily read using basic text editing software. It does, however, permit hyperlinking to external content, although these links are inactive. At the same time PDF/A-1 requires the inclusion of extensible metadata (XMP) that documents the file and facilitates use. As Sullivan reports,

PDF/A may not be the last preservation format that will be needed, but proper application of PDF/A should result in reliable, predictable and unambiguous access to the full information content of electronic documents.^[18, p.55]

The PDF/A-1 specification offers two levels of compliance: level A and B. Those with the lower level of conformity

CONTINUED »

Figure 2: The typical content of a modern digital archaeological report

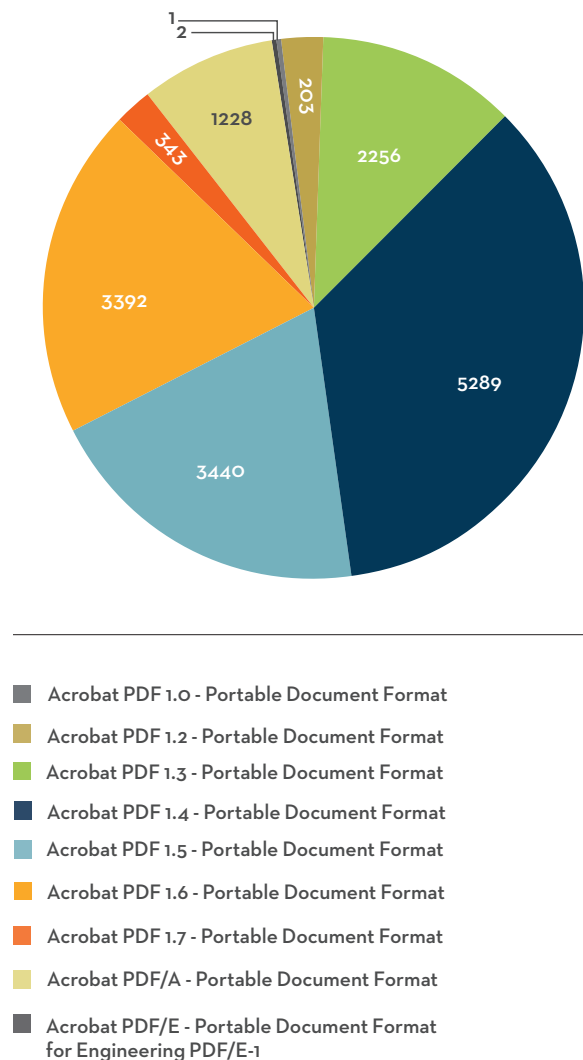


(PDF/A-1b) have basic compliance to the PDF/A-1 standard, ensuring reliable reproduction of the visual appearance of the PDF. The higher level of compliance not only preserves the visual appearance, but also preserves accessibility (for the visually impaired) and makes any content available for reuse. The subsequent extension of the underlying PDF format has seen the development of PDF/A-2 (ISO 19005-2:2011)^[20] and PDF/A-3 (ISO 19005-3:2012),^[21] which address subsequent developments of the specification based upon the later PDF 1.7 specification (ISO 32000-1:2008). In each instance, and when implicated within a format normalization strategy, PDF/A allows repositories to preserve PDF content more comprehensively. The full impact of these developments has yet to be realized amongst the wider archiving community.

Like other commercial and academic environments, the PDF has become a pervasive reality within the archaeological profession. A recent report produced by the ADS has shown that just over 50% of unpublished grey literature held by archaeological organizations is principally in a digital form, with 43% maintained in a PDF format.^[11] As the figures in this article suggest, this number may be conservative. These figures for the digital reports have been augmented by programs of digitization of physical archives, both within the academic and commercial sectors of the archaeological profession, where the principle format is similarly the PDF. Consequently PDF remains the most common format uploaded through OASIS, and transferred into the ADS Grey Literature Library (Figure 1 and Figure 3). These figures must be increased through the inclusion of other PDF files deposited and archived in other collections with the archive.

Despite the popularity of the PDF within archaeological workflows, the ADS does not actively encourage its use as a deposition format; rather depositors are encouraged to preserve the original data streams. Such a policy has developed as a consequence of concerns over the PDF and PDF/A's suitability and sustainability for long-term preservation; an awareness that is becoming more apparent within the digital archiving sector generally. Despite this encouragement and the concerns over sustainability, PDF is still very clearly the preferred method of transfer for a digital report (Figure 3 and Figure 4). Indeed, a basic analysis of the files uploaded to OASIS demonstrates that PDF has become almost the default format for deposition at the time of writing, dominated by PDF 1.4 - 1.6 and in recent years a growing number of PDF/A files (Figure 4). Of interest in these figures is the rise in the use of PDF, often at the expense of original formats such as Microsoft Word. Thus, in the light of this unavoidable prevalence, the ADS has had to take a pragmatic view and look towards the adoption of the PDF/A as a preservation format.

Figure 3: Versions of PDF uploaded to OASIS and accessioned by the ADS (as of 01/08/2013)



While PDF and PDF/A specifications are published and recognized as ISO standards, questions have been raised over the true openness of a format that is still essentially proprietary—being owned by Adobe. Some have noted inconsistencies and discrepancies within the published specification that make development problematic; consequently developers are forced to “fill in those gaps” present within the schema to make it usable.^[22] Related to this issue, concerns have also been raised over the accuracy and precision of software designed to create and validate files that ostensibly adhere to the PDF/A specification,

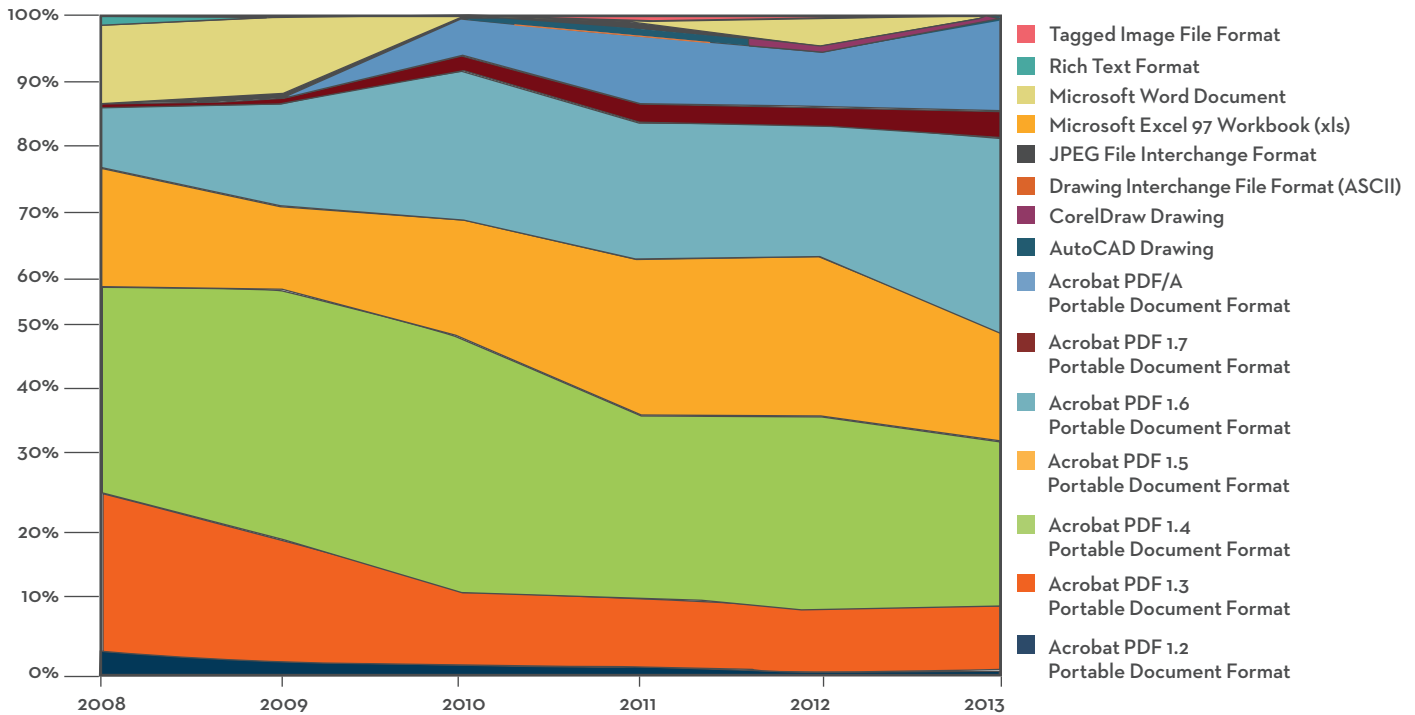


Figure 4: Percentage of file types accessioned through OASIS, by year 2008-2013 (as of 01/08/2013)

but which may not always be the case;^[23] something the experience at the ADS would seem to substantiate. These factors undermine confidence in the longevity of the format for preservation.

While the principle focus of many archiving strategies is the preservation of files and data streams, planning for the reuse of data should also remain high on the agenda. PDF and PDF/A files, when properly implemented, certainly fulfill many of the requirements for long-term preservation, but an emphasis on structure and formatting means that their data reuse is often marginalized. The GLADE report,^[11] for example, has shown that archaeologists producing grey literature for upload to OASIS typically use desktop publishing software like Microsoft Word or Apache OpenOffice to create reports that incorporate a variety of data streams, before being exported into a PDF format. As previously mentioned, archaeological field reports comprise a variety of data types alongside the “traditional” text and image formats. When archived within the PDF/A format, these data streams are preserved in a “flat” form where the emphasis is on visual appearance of the data—something that undermines any computational reading of the data stream and a process that can undermine potential reuse.

PDF/A-3 and future archiving strategies

The publication of the PDF/A-3 specification in October 2012 certainly attests to the continued development of a preservation solution for the PDF format. Essentially the PDF/A-3 standard handles the issue by acting as a container that allows data creators to embed original data, in almost any file format, within a PDF/A compliant document. The development is a headline grabbing one and will certainly appeal to data creators, allowing them greater flexibility in the sharing of data and digital content, whilst simultaneously addressing the issue of preservation. With regard to the latter, the publicity contends PDF/A-3:

eliminates time-consuming hybrid archiving processes in which additional documents (Excel tables, image files, CAD drawings) had to be managed separately from the archived PDF/A file in their original formats. Thanks to PDF/A-3, all relevant information is now contained within a single file.^[24]

This additional feature will certainly open up the format to new applications amongst data creators, but digital archivists and preservation specialists will have already recognized the obvious flaw in the new standard in that it fails to regulate the suitability and sustainability

CONTINUED »

of these embedded files for long-term preservation. This is less of a problem for the more common and open XML-based formats, for example the Microsoft Office or Apache OpenOffice suites, but should certainly raise concern for the outputs of more specialist and proprietary software. Within the archaeological profession, for example, excavation reports typically comprise multiple data streams—from the traditional desktop publishing, databases, and spreadsheets through to more complex data types including CAD, GIS, geophysics, laser scanning, etc.—each requiring discreet preservation strategies.^[3] The specification, therefore, overlooks the need for preservation strategies for this related content—something which conflicts with constraints of the other schemas within the PDF/A family. Of course the standard does acknowledge this shortcoming,^[28, p.8] reporting that any embedded content should be considered “non-archival” and only of short-term or temporary use;^[25] yet its marketing as a long-term archiving solution does little to highlight what could be a fundamental misapprehension amongst those casual users less familiar with digital preservation.^[26] With little control over content, a warranted concern is that the data creators will simply use the format as a “trashcan” for data under the false impression that long-term archiving has been achieved. While endorsing it as an ISO standard under the high profile PDF/A banner, data creators familiar with the headlines may well come to regard this as good archiving practice. More problematic from the preservation perspective is the absence of any obligation within the specification for a compatible reader application that enables the rendering and extraction of embedded objects.^[25]

From a data management perspective, the ability of the PDF/A-3 to act as a container, or digital document folder, for associated data streams will be welcomed by many. Within environments where dedicated data management is practically, technically, or economically difficult, the format will be well received as a sustainable solution to the problem; whilst established infrastructures seeking to rationalize strategies will be attracted by the opportunity to reduce complexity and cost. Ostensibly the ability to create PDF/A compliant documents that can be stored alongside associated data “archives” would seem to be an obvious benefit, particularly to those less familiar with data management strategies; yet, in reality, such functionality offers little more than good data management practices currently offer. Of course the requirement for data creators to assign tags (*source*, *data*, *alternative*, *supplement*, and *unspecified*, along with the MIME type) to identify the nature of embedded content will certainly assist in the evaluation of the significant properties of these associated data streams in terms of their management. A proposed enhancement to the schema that would allow creators to explicitly tag those attachments needing preservation would certainly be welcomed amongst

digital archivists. While this metadata is invaluable for those making assessments about the significant properties of the file, the specification only makes limited provision for any other metadata necessary to make any associated data meaningful. At the same time the PDF/A-3 specification presents itself as a versioning tool, one that means the document can include the current working version alongside a final archive version. This will certainly be convenient for creators, but the prospect of developing strategies to deal with multiple versions of the same data stream will fill data managers and digital archivists with dread. Unfortunately at the time of writing, full details of this aspect of the schema remain unclear; consequently making a full appraisal is difficult, but obviously how it deals with the issue of updating content and its relationship with the final “archive-ready” PDF/A version would seem fundamental.

On the positive side, a rather incidental outcome of the new format may be that by allowing creators to add associated content to the PDF/A-3, it may well indirectly encourage fuller preservation of data streams. The PDF/A-3 specification, much like others in the PDF/A family, focuses on the preservation of the visual appearance of the document; consequently any data is essentially “locked in” a fixed textual form that is viewable by the human eye, but is degraded to a form that is difficult to extract and reprocess digitally. By allowing the association of original data streams, creators can extend the potential for preservation and reuse of information in both the short and long term. Of course this new specification will require the development of new strategies to preserve these associated data streams, but as many of these files will have been added in formats familiar to digital archivists and preservation specialists, this should not be considered impossible. Yet difficulties will certainly arise when creators append data streams that are in non-standard formats, or that are not accepted files at the repository receiving the data. What is more problematic, however, is that much of this appended content will lack the appropriate metadata that can provide important contextual information about complex data streams, assist in the assessment of the significant properties, and aid the development of preservation strategies. More significantly, this metadata can provide information that can facilitate use, and reuse, of data—as demonstrated with the ADS Grey Literature Library. It is only by continuing to raise awareness of the importance of metadata that digital archivists can effectively deal with the problem of this embedded content.

Planning for the future

As the PDF takes an increasingly pervasive role within contemporary workflows, digital archivists have been faced with developing appropriate strategies that deal with the long-term preservation of the format. Initially these

approaches were reactive and unsustainable involving the use of open and uncompressed image-based formats. It has been the development of the PDF/A standard that has provided a technically sophisticated, open, and self-contained archiving solution to the preservation of the PDF. While subsequent developments of the PDF/A standard (PDF/A-2 and PDF/A-3) have certainly extended the flexibility of the format, most notably through the ability to embed an increasingly diverse range of files and data streams within the file, but have brought into question the sustainability of the format for preservation. For those working within the archiving community, the principle advantages of the original standard—its freedom from external dependencies—has been seriously undermined. But should digital archivists be overly concerned by these developments? From a technical perspective, the archiving of this embedded content seems relatively straightforward. Experience in working with collections, like the Grey Literature Library, has allowed the ADS to develop preservation strategies for more complex archives that include a diverse range of data types. It is, however, the inadequate metadata requirements for this embedded content that causes most alarm. An ability to tag a digital archive for preservation is not as useful as it might seem, as without appropriate metadata these data streams are virtually useless. In marketing PDF/A-2 and PDF/A-3 as archiving solutions, further work will be necessary to document and make content accessible in the long term. Of course, educating data creators about these shortcomings will be left to archivists and preservation specialists and will no doubt cause much consternation with data creators. While new formats are always treated with some suspicion amongst the archiving community, the issues arising from the developments of the PDF specification should not be considered insurmountable and may actually provide an opportunity for more complete archiving of data streams, even if this is something of an unintentional outcome.

IP | doi:10.3789/isqv25no3.2013.04

RAY MOORE (ray.moore@york.ac.uk) and **TIM EVANS** (tim.evans@york.ac.uk) are Digital Archivists with the Archaeology Data Service, Department of Archaeology, University of York, in the U.K.

REFERENCES

1. Archaeology Data Service (ADS). archaeologydataservice.ac.uk/
2. Arts and Humanities Data Service (AHDS). www.ahds.ac.uk/
3. Archaeology Data Service/Digital Antiquity. *Guides to Good Practice*. guides.archaeologydataservice.ac.uk/g2gp/
4. ArchSearch. archaeologydataservice.ac.uk/archsearch/
5. OASIS project. oasis.ac.uk
6. Grey Literature Library. archaeologydataservice.ac.uk/archives/view/greylit/
7. Archaeological Investigations Project (AIP). web.bournemouth.ac.uk/aip/aipintro.htm
8. English Heritage. www.english-heritage.org.uk/
9. For example, see Bradley, Richard. "Bridging the Two Cultures - commercial archaeology and the study of Prehistoric Britain." *The Antiquaries Journal*, September 2006, 86: 1-13. <http://dx.doi.org/10.1017/S0003581500000032>
10. OASIS III: 21st Monitoring Report, April 19, 2013. oasis.ac.uk/monitoring/OASIS_15apr13.doc
11. Hardman, Catherine, and Evans, Tim. *GLADE: grey literature - access dissemination and enhancement. The pilot assessment phase final report*. York: Archaeology Data Service, August 25, 2010. <http://archaeologydataservice.ac.uk/attach/research/GLADEreportv5.pdf>
12. RCAHMS and Historic Scotland. www.rcahms.gov.uk/rcahms-and-historic-scotland
13. Seymour, Deni J. "In the Trenches Around the Ivory Tower: Introduction to Black-and-White Issues About the Grey Literature." *Archaeologies*, August 2010, 6 (2): 226-232. <http://dx.doi.org/10.1007/s11759-010-9130-z>
14. Ford, Matt. "Hidden Treasure." *Nature*, April 8, 2010, 464: 826-827. <http://dx.doi.org/10.1038/464826a>
15. Europeana. www.europeana.eu/portal/
16. Vlachidis, Andreas, Ceri Binding, Douglas Tudhope, and Keith May. "Excavating grey literature: A case study on the rich indexing of archaeological documents via natural language-processing techniques and knowledge-based resources." *Aslib Proceedings*, 2010, 62 (4/5): 466-475. <http://dx.doi.org/10.1108/00102531011074708>
17. ISO 32000-1:2008, *Document management - Portable document format - Part 1: PDF 1.7*. www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=51502
18. Sullivan, Susan J. (2006). "An archival/records management perspective on PDF/A." *Records Management Journal*, 2006, 16 (1), 51-56. <http://dx.doi.org/10.1108/09565690610654783>
19. ISO 19005-1:2005, *Document Management - Electronic Document File Format for Long-Term Preservation - Part 1: Use of PDF 1.4 (PDF/A-1)*. www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=38920
20. ISO 19005-2:2011, *Document management - Electronic document file format for long-term preservation - Part 2: Use of ISO 32000-1 (PDF/A-2)*. www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=50655
21. ISO 19005-3:2012., *Document management - Electronic document file format for long-term preservation - Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3)*. www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57229
22. Morrissey, Sheila M. "The Network is the Format: PDF and the Long-term Use of Digital Content." In: *Archiving 2012: Final Program and Proceedings*, Copenhagen, Denmark; June 2012, pp. 200-203. Springfield, VA: Society for Imaging Science and Technology. Abstract at: www.imaging.org/IST/store/epub.cfm?abstrid=45333
23. Koo, Jamin, and Carol C.H. Chou. "PDF to PDF/A: Evaluation of Converter Software for Implementation in Digital Repository Workflow." *New Review of Information Networking*, 18 (1): 1-15. <http://dx.doi.org/10.1080/13614576.2013.771989>
24. Oettler, Alexandra. *PDF/A in a Nutshell 2.0*. Berlin: Association for Digital Document Standards, 2013, pp. 10. www.pdfa.org/2013/04/pdfa-in-a-nutshell-2_0/
25. PDF/A-3, PDF for Long-term Preservation, Use of ISO 32000-1 with Embedded Files. Library of Congress, last significant update November 19, 2012. www.digitalpreservation.gov/formats/fdd/fdd000360.shtml
26. Fanning, Betsy A. *Preserving the Data Explosion: Using PDF. DPC Technology Watch Series Report 08-02*. Silver Spring, MD: Digital Preservation Coalition and AIIIM, April 2008. www.dpconline.org/docs/reports/dpctw08-02.pdf