

*Article*

# Migrating 2 and 3D Datasets: Preserving AutoCAD at the Archaeology Data Service

Katie Green <sup>\*</sup>, Kieron Niven <sup>†</sup> and Georgina Field <sup>†</sup>

Archaeology Data Service, Department of Archaeology, University of York, The King's Manor, York YO1 7EP, UK; kieron.niven@york.ac.uk (K.N.); georgina.field@cantab.net (G.F.)

<sup>\*</sup> Correspondence: katie.green@york.ac.uk; Tel.: +44-1904-323-954

<sup>†</sup> These authors contributed equally to this work.

Academic Editors: Constanze Curdt, Christian Willmes, Georg Bareth and Wolfgang Kainz

Received: 25 January 2016; Accepted: 16 March 2016; Published: 25 March 2016

**Abstract:** The Archaeology Data Service (ADS) is a digital archive that has been promoting good practice in the use of digital archaeological data and supporting research, learning and teaching with high quality and dependable digital resources for twenty years. The ADS does this by preserving digital data in the long-term and by promoting and disseminating, open and free datasets, gathered from all sectors of archaeology. An integral component of the ADS remit has been the life-cycle principle of preservation, curation and dissemination of data in order to enable re-use. The ADS practices a combination of normalization, version migration, format migration and refreshment for the active management and ongoing preservation of all archived data types. This paper highlights the importance of the ongoing management of research data for long-term preservation. In particular this paper focuses on the challenges of migrating spatial data, specifically Computer Aided Design (CAD) files. Successful data migration of these files ensures that data is accessible and usable, and provides many opportunities through data re-use to combine and re-interrogate datasets, allowing new archaeological interpretations to be developed.

**Keywords:** digital preservation; data migration; CAD; re-use

## 1. Introduction

An essential component of effective Research Data Management is the preservation and curation of digital data. This paper discusses the importance of digital preservation and active Research Data Management, through a case study of preservation by migration of Computer Aided Design (CAD) data in the archaeological discipline. The issues associated with the long-term preservation of digital data—together with the advantages of doing so—are now becoming increasingly well-known across a wide range of fields outside the archival community. In the archaeological discipline there is a growing awareness of the importance of Research Data Management and online open access resources have become an increasingly important aspect of modern archaeological research practice. Archaeology by its very nature can be a destructive process. Creating a comprehensive record of archaeological research is imperative to the discipline because many of archaeology's primary field methods cannot be replicated by future researchers. It is therefore vital to preserve and facilitate access to this archaeological record in order to test, assess and subsequently reanalyze data and the hypotheses arising from them. Good practice frameworks now require archaeological organizations and practitioners to institute procedures to ensure the long-term preservation and access to digital data. In England the National Planning Policy Framework [1] states that any heritage assets to be lost (wholly or in part) must be recorded in a manner proportionate to their importance and any archive generated must be made publicly accessible. The Chartered Institute for Archaeologists also emphasizes in its standards and guidance “that it is the responsibility of all curators of archaeological

archives to ensure that archives are stored to recognized standards for long-term preservation and made accessible for consultation" [2].

The Archaeology Data Service (ADS) is a discipline-specific digital repository which was established in 1996 in response to the increase in the creation of born digital archaeological data [3] and in recognition of the associated dangers of data loss [4]. The ADS is hosted by the Department of Archaeology at the University of York and was initially funded by the UK Arts and Humanities Research Board (now Arts and Humanities Research Council, AHRC) and the Joint Information Systems Committee (JISC). More recently the ADS has developed an innovative business model that relies on a depositor-charging policy [5]. The ADS remit is to support research, learning and teaching within archaeology with freely available, high-quality and dependable digital resources. The ADS does this by preserving digital data in the long term, supporting the deployment of digital technologies, promoting and disseminating a broad range of data in archaeology, and providing technical advice to the sector via the ADS website. Over the past 20 years, the ADS has become widely recognized for excellence in digital preservation and in developing and disseminating guidance on standards for archiving, not just in the archaeological community, but on a much wider scale. In 2012 the ADS received the Digital Preservation Coalition (DPC) Decennial Award for the most outstanding contribution to digital preservation over the preceding decade [6] and the ADS has held the Data Seal of Approval since 2011 [7].

An integral component of the ADS remit has been the life-cycle principle of preservation, curation and dissemination of data in order to enable re-use. This paper introduces the digital preservation and migration strategies at the ADS before addressing the challenges of a recent large scale migration of CAD files undertaken by the ADS. This is the first time a migration of this scale has been attempted in the 20 years of the ADS's existence. The strategy for migrating CAD data, as used by the ADS, is presented and the necessary steps for a successful CAD migration are laid out. Associated positive and negative outcomes from the migration are presented. This case study will provide valuable lessons for the digital preservation community and those attempting migration as part of their own Research Data Management strategies. The authors argue that although a time consuming, and somewhat laborious process, data migration is a vital step for maintaining a viable and publically accessible digital archive and is an integral Research Data Management strategy.

## 2. Digital Preservation at the ADS

The ADS preserves archaeological datasets within the area broadly defined as Archaeology and the Historic Environment, including the material culture aspects of Ancient History and Classics. Data which are offered for deposit to the ADS are evaluated to assess their intellectual content and evaluate if they may viably be accessioned, managed, preserved, and disseminated [8].

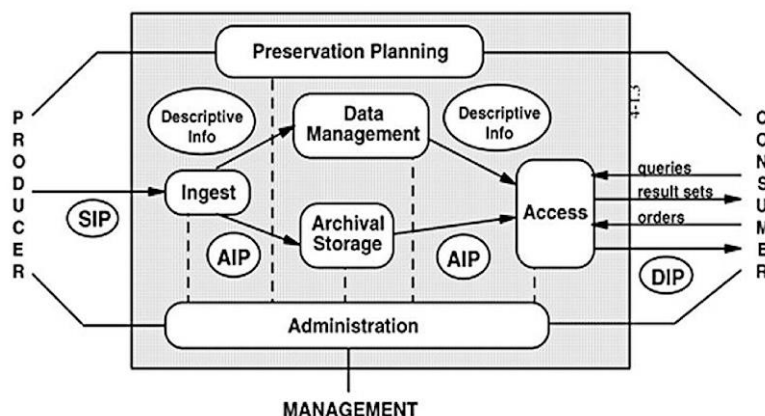
In order to meet these requirements the ADS provides depositors with a series of online Guidelines for Depositors [9] which provide detailed advice on the ADS's collection policy, accepted file formats for deposit, and how to compile metadata specifically for deposition with the ADS. The Guidelines for Depositors are complimented by the ADS and Digital Antiquity's Guides to Good Practice [10]. The Guides to Good Practice are a series of online guides which aim to provide advice and workflows for the planning, creation, and preservation of common archeological data types.

In the archaeological discipline the advancement of digital media has dramatically changed how we communicate and record our heritage. Today archaeologists are generating born digital data in unprecedented volumes during every stage of a project, from fieldwork and assessment, through analysis and reporting, to dissemination. These datasets are increasingly complex, utilize a variety of formats, and can be the accumulated research of individuals, teams and institutions, forming a vast and fragmented corpus. Preservation processes may have to deal with hundreds of file types, from copious types of devices, using a plethora of software packages and a broad range of archaeological techniques which adds to the preservation challenge. The types of data typically archived by the ADS include: text reports, databases, raster images (including aerial photographs, remote sensing imagery, digitized

maps and plans), datasets related to topographical and sub-surface surveys and other locational data, vector images, such as CAD drawings, 3D reconstruction models, video and audio. To mitigate the challenge of dealing with such a variety of different data types the ADS maintains up-to-date data procedures for all common archaeological data types, which include details on the appropriate file formats for deposition, long-term preservation and dissemination.

Data deposited with the ADS is managed within a framework which conforms to the Open Archival Information System (OAIS) reference model [11]. The OAIS reference model is a framework which defines the responsibilities and interactions of data producers, managers and consumers, and maps out the core activities (and relationships between them) that need to be carried out in order for the system to work. Figure 1 depicts the major functions of the OAIS reference model. The reference model specifically outlines the processes required for the ingest, long-term preservation and dissemination of data, through a series of data transformations that should take place as data moves from the data producers through the OAIS and on to the data consumers. These transformations form the following information packages:

- Submission Information Package (SIP): Data supplied by the data producer, including documentation to facilitate archiving and re-use.
- Archival Information Package (AIP): Data generated from the SIP and transformed into a long-term preservation package managed within the OAIS, including administrative, technical and re-use documentation.
- Dissemination Information Package (DIP): Data generated from the SIP or AIP and made available to data consumers, including documentation to facilitate re-use.

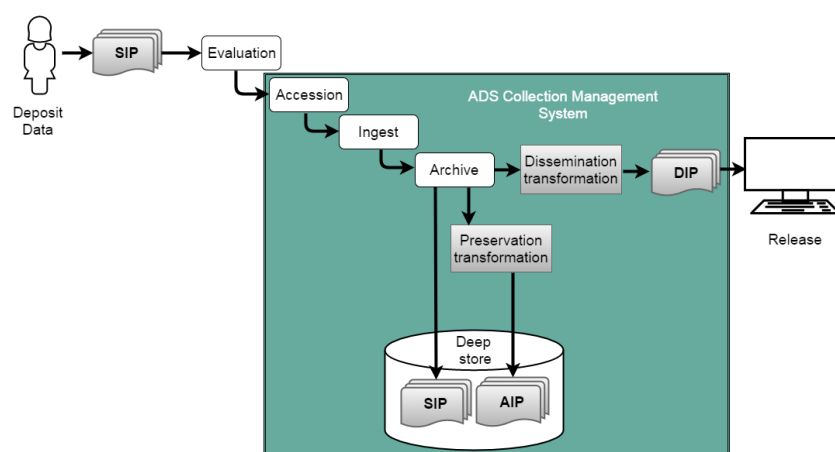


**Figure 1.** The major entities of the Open Archival Information System (OAIS) Reference Model.

© Consultative Committee for Space Data Systems.

The ADS's Preservation Policy [12] and Repository Operations [13] actively follow preservation and management strategies based on the OAIS reference model with the aim of ensuring the authenticity, reliability and logical integrity of all the resources entrusted to its care. The ADS's operations can be briefly outlined as follows: Data producers deposit data in accepted formats with data-type specific, file-level metadata. The data and metadata provided by the data producer, alongside administrative documentation becomes the SIP. The SIP is then accessioned and ingested into the ADS Collections Management System (a bespoke Java Struts application based on an Oracle database which facilitates and records all the ADS's accessioning, preservation and dissemination processes). The accessioning process includes verification of the validity, integrity, consistency, and completeness of all data and metadata received. The SIP is then stored and an AIP and DIP created. This often requires the transformation of data into appropriate file formats for long-term preservation and online dissemination. The AIP and SIP then enter the ADS archive for long-term storage and curation. For all

datasets, the DIP is made available to data consumers as a series of files downloadable from the ADS website. This workflow is depicted in Figure 2. Table 1 also provides examples of the data transformations undertaken to create AIP and DIP formats for data types submitted in a selection of formats. While certain datasets have additional online interfaces (e.g., searchable databases or GIS datasets), all data—including CAD files—at the ADS are stored and disseminated as individual files within discrete project-based collections. While such a strategy does not currently mean that data (e.g., geospatial datasets) is integrated between collections, standardized collection- and file-level metadata is integrated across the entire ADS archive to allow data discovery.



**Figure 2.** Archaeology Data Service (ADS) operational workflow.

**Table 1.** This table shows the preservation and dissemination formats for data types submitted in a selection of formats. *N.B.* This is not a complete list of all ADS accepted file formats and differences in archival and dissemination formats may occur.

Data Type	Submission Format (SIP)	Preservation Format (AIP)	Dissemination Formats (DIP)
Audio	.aif	.wav	.mp3
CAD	.dxf	.dwg	.dxf and .dwg and .pdf/a
Database	.accdb	.csv	.csv
Geophysics (RAW data)	.segy	.segy	.segy
Geophysics (rendered images)	.png and .pgw	.tif and .tfw	.png and .pgw
Raster Image	.gif	.tif	.jpg or .png
Spreadsheet	.xls	.csv	.csv
Statistics	.sav	.txt	.txt
Text	.doc	.docx	.pdf/a
Vector Image	.ai	.dwg and/or .tif	.dxf and .dwg and dpdf/a
Video	.dvx	.mp4	.mp4

### 3. An Overview of Migration

OAIS does not proscribe specific preservation strategies or guidance for the preservation of specific data types within the AIP, but the active management and lifecycle approaches it advocates tend toward migration rather than emulation or technology preservation [11]. The ADS adopts this approach, and uses a combination of normalization, version migration, format migration and refreshment for ongoing preservation of all archived data types [12] (pp. 4–5).

Planning for the migration of data stored within the ADS is a key component of our larger “Preservation Planning” data lifecycle activities. The various types of digital migration are discussed in detail in the OAIS reference document but, in simple terms, a data migration consists of a transfer or update of a discrete set of data, where an older version is replaced comprehensively with a newer or alternative implementation. All data within the ADS archive is normalized to standard preservation

and dissemination formats on ingest (see Table 1 for examples), however, these formats may become outdated or superseded by newer versions or alternative formats. Ongoing ADS activities, including a “technology watch” and engagement with user communities, aim to monitor and highlight such developments. When a new format emerges that is suitable for both the ADS and the designated community (both data producers and data consumers), it may trigger the need for a digital migration. Migration within the ADS is usually defined in terms of the update or replacement of a single long-term preservation file format across the entire archive. The way in which data is stored within AIP directories at the ADS is geared towards such format-based data migrations [13].

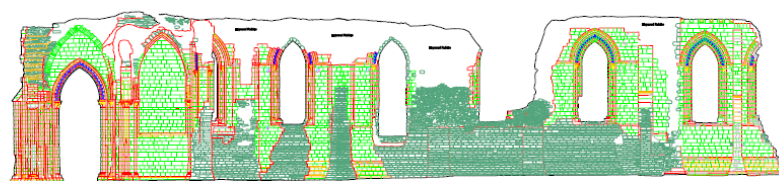
It is worth highlighting at this point that large-scale data migrations are not without risk [11] (pp. 3–5) and should ideally be mitigated by the initial process of data normalization to standardized formats carried out at the ingest stage. A key component to identifying suitable formats for preserving and disseminating data is their stability and resistance to frequent change. When a data migration occurs it frequently signifies a carefully planned and fundamental change in how a repository ingests, stores, and preserves a certain data type *i.e.*, a new preferred format replaces a previously standardized data format used across the entire archive.

- As a simplified step-by step process, a data migration involves: The identification of archived data (e.g., specific file types) that require migration
- The duplication and storage of this data as an “original” AIP
- The migration of the identified data to a new archival format and creation of new AIP

In addition to the creation of “new data”, a data migration also involves a corresponding update to relevant Descriptive Information and Preservation Description Information (PDI) relating to the migrated dataset. These updates are a key component of an ongoing data lifecycle and include documenting what processes have been undertaken on which data (input and output), new locations for this data, and other elements such as fixity values.

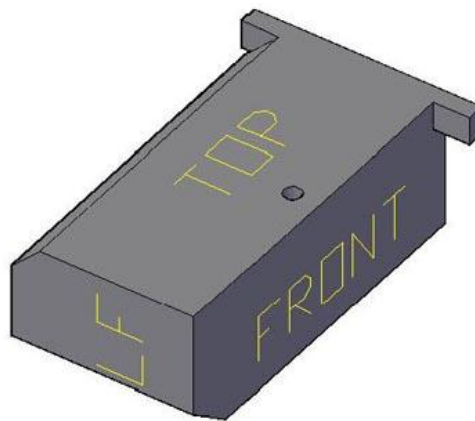
#### 4. CAD Migration Strategy at the ADS

In 2013, through a combination of technology watch and the monitoring of datasets ingested into the ADS archive, it was decided that a change in policy on how the ADS stores and disseminates CAD files was required. Two and three dimensional CAD files, along with a wide range of geospatial and three dimensional datasets, are a common component of many archaeological projects and are regularly deposited at the ADS from a wide range of commercial and research-based data producers. Figure 3 is a representation of a typical two dimensional CAD file recording the elevation of a building including information on building phases. Figure 4 depicts a typical three dimensional CAD model of a block from a historic building. These files are almost always created in Autodesk’s AutoCAD software with the application and associated file types having remained a *de facto* standard for archaeological CAD data for at least the last 15 years [14] While the common adoption of this software has resulted in seemingly straightforward requirements for data deposit—the ADS guidelines specify that CAD files should be deposited as either AutoCAD DWG or Drawing Exchange Format (DXF) [15]—the regular update and release of new versions of these formats by Autodesk has created a number of issues for the long-term storage and preservation of this CAD data.



**Figure 3.** Representation of a two dimensional Computer Aided Design (CAD) file held in the ADS archive depicting the south elevation of the choir aisle’s south wall of Glastonbury Abbey [16] © Trustees of Glastonbury Abbey.





**Figure 4.** Representation of a three dimensional CAD file of the Propylaea [17] held in the ADS archive.  
© Harrison Eiteljorg II.

Prior to the data migration described here, the ADS archival policy has been to ingest CAD data as either native AutoCAD DWG files or in Autodesk’s DXF format. These files have then been migrated to DXF version R14 for both preservation and dissemination purposes (see Table 2). The decision to use DXF R14 as a preservation format was primarily based on its support for textual encoding (ASCII) and its primary purpose as an exchange format which could be used beyond Autodesk software [18]. In reality, however, due to the fast development of the AutoCAD software, the DXF format has seen almost as many version updates as the proprietary DWG format (which has seen eighteen new versions since 1982). As a result, the decision was made in early 2014 to change the ADS archiving policy and adopt DWG version 2010 (AC1024) as the preferred archival format for CAD data. This format was assessed and found to be a relatively stable and well adopted format and, as the use of AutoCAD is commonplace within the community, the move to a native format introduced specific advantages both in terms of compatibility and file size. As the data migration process remained focused on native AutoCAD formats, the move to the DWG format also provided better assurance that the various significant properties and characteristics of CAD files remained both within the preservation and dissemination versions of the data. Such properties include the precision and accuracy of the data, the associated coordinates and geometry alongside visual conventions such as line weights, styles, and colors [19] (these are also documented in the required metadata files deposited with all datasets).

**Table 2.** Numbers of CAD files in the ADS archive by file format pre-migration. (The numbers in this table do not tally due to duplicate files (*i.e.*, in both formats) in both original and dissemination directories. Other non-CAD vector files are also preserved as DXF and may be disseminated in other formats.)

File Format	Original Submitted Files	Preservation Files	Dissemination Files
AutoCAD DWG	1254	0	512
DXF	675	1498	1500

In addition to the migration of the archived CAD datasets, the proposed digital migration of CAD files at the ADS would also involve updating the dissemination versions of these files. Previously all CAD data was disseminated as individual downloadable DXF files (the same format that was stored within the AIP). As part of this process it was decided that the ADS should also aim to increase the accessibility and re-use potential of CAD datasets through the dissemination of both AutoCAD 2010 DWG and DXF files alongside additional Portable Document Format (PDF) files, preview images, and thumbnails of the drawing’s layouts. While the migration of CAD data in existing AIPs was a “behind the scenes” task, that primarily served the long-term preservation needs of the ADS, the updating

and enhancement of the dissemination versions of the data in the DIPs created a separate thread of migration work that directly impacted on the user-facing side of the ADS website.

## 5. CAD Migration Process

The data migration process was undertaken over a number of months during 2015 and can be broken down into a series of clearly defined stages. These stages are summarized below and depicted in Figure 5:

Step 1: Identification of CAD files. The initial stage of data migration involved the identification of collections within the ADS repository that contain CAD datasets. In addition to simply identifying the collection, individual files, their locations, file types (DXF or DWG) and versions were also recorded. CAD files were identified both by file extension and by a digital signature created by the National Archive's Digital Record Object Identification tool DROID. This signature is stored within the ADS CMS at points during the data accession and archiving process and allows all stored file versions to be identified.

Step 2: Removal and storage of original preservation and dissemination versions of files. A key element of the migration strategy was that while the archived CAD datasets were largely to be updated, no data was to be removed or deleted from the archive. In practice this meant that any preservation or dissemination versions of data that were to be replaced were copied to a new "migrated" location within the collection prior to the migration process (see Figure 6 for an example of a collection AIP file structure). As the majority of data migrations are carried out on the existing preservation datasets, this process aims to preserve these as intermediate datasets between the original SIP data and the new, migrated AIP data. Although, as described below, this wasn't always the case in the ADS CAD migration process, these datasets were kept to provide a "history" for how the CAD elements of each collection had been preserved and disseminated.

Step 3: Migration of archival versions of data to a new preservation format (DXF or DWG to DWG 2010). This step formed the core element of the migration process for the preservation datasets and involved the creation and verification of CAD files in the new DWG preservation format. Previous identification of the files, along with the file-type archive structure upon which the ADS collections are based, allowed this work to be automated, or partially automated, per collection.

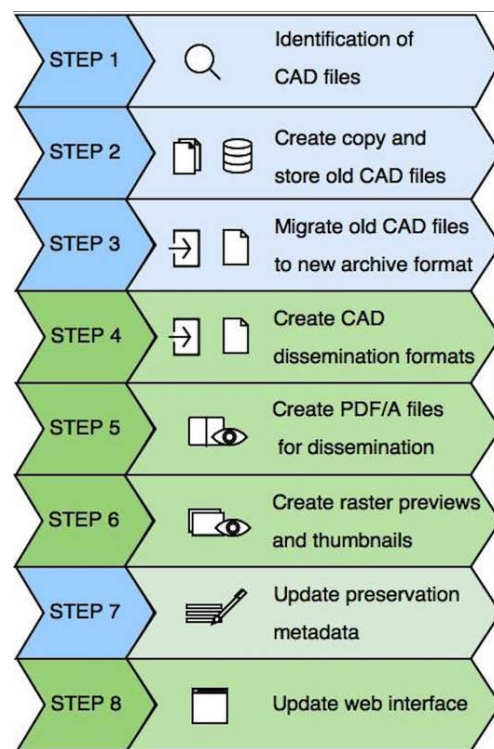
Step 4: Migration of data to new dissemination formats (DWG or DXF to DWG and DXF 2010). As with Step 3, Step 4 formed a core element of the migration work undertaken, but went beyond the updating of existing files to include the creation of multiple dissemination versions.

Step 5: Creation of PDF/A files of CAD layouts for dissemination. As well as creating new dissemination files in CAD DWG and DXF formats, additional, more accessible PDF/A format files were also created for those users without access to CAD software.

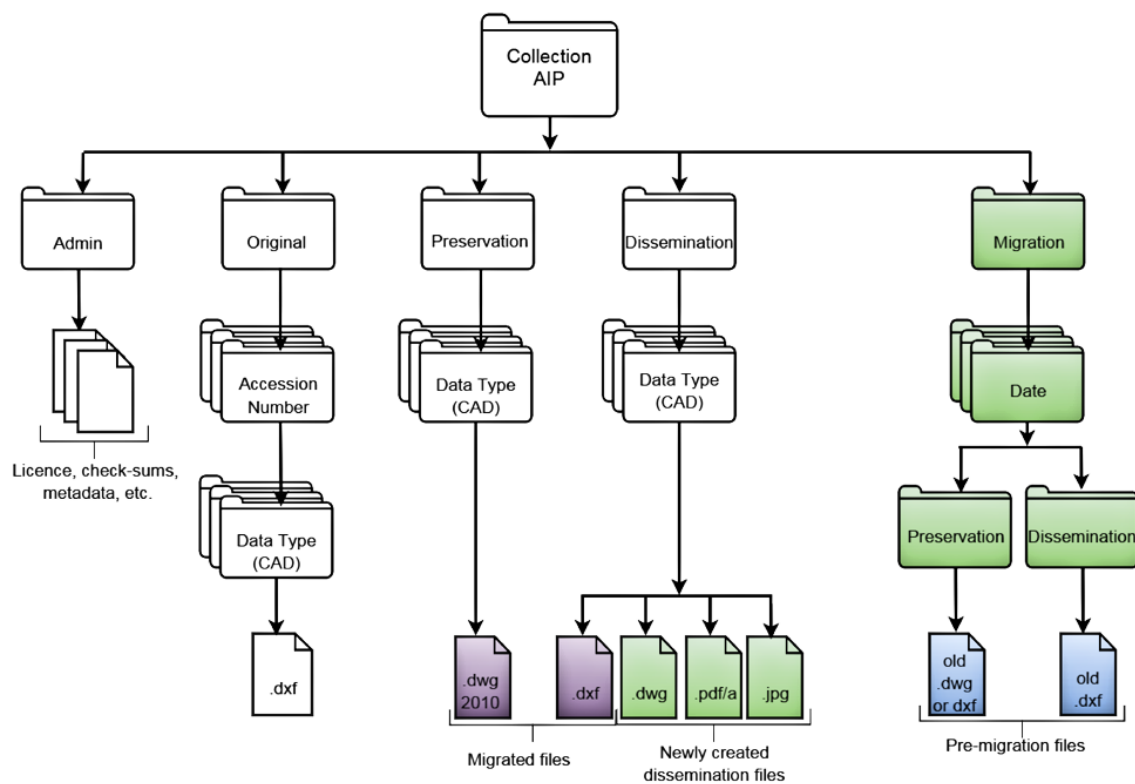
Step 6: Creation of raster preview images and thumbnails for dissemination. As with Step 5, additional raster preview and thumbnail images were created from the CAD files to provide increased data accessibility.

Step 7: Update of corresponding metadata. Aside from Step 8 (providing access) the key final stage of the data migration was the documentation of the process itself. This involved updating the various sections of the ADS CMS to include details of the processes carried out on the data, the locations of both old and new files, the updating of fixity values, and the documentation of new relationships between files.

Step 8: Update of web interfaces to include new files and previews. As a result of providing updated and additional dissemination formats, the final stage of the data migration process involved the updating of the relevant ADS collection webpage to allow access to these new files (see Figures 7 and 8).




**Figure 5.** Workflow diagram of the ADS's CAD migration process. Blue highlights steps taken to ensure long-term preservation, green highlights steps taken to enhance the dissemination of data.



**Figure 6.** Example of the additions to an Archival Information Package (AIP) file structure following a migration. Green indicates new files, purple indicates migrated files, blue indicates moved files.





[HOME](#) [ARCHSEARCH](#) [ARCHIVES](#) [ADS-easy](#) [LEARNING](#) [ADVICE](#) [RESEARCH](#) [ABOUT US](#) [BLOG](#) [myADS](#) [LOGOUT](#)


[ALL](#) | [JOURNALS AND SERIES](#) | [GREY LITERATURE](#) | [PROJECT ARCHIVES](#) | [BIBLIOGRAPHIES](#) | [THESES](#)

## Gwithian, Cornwall: Excavations 1949-1969

Henrietta Quinnell, Anna Lawson-Jones, Jacky Nowakowski, Joanna Sturgess, Carl Thorpe, Charles Thomas, 2007


[Introduction](#)  
[Overview](#)  
[Downloads](#)  
[Metadata](#)  
[Usage Statistics](#)

[Aggregates Levy Sustainability Fund](#)  
 Data copyright © Cornwall Council unless otherwise stated

  
 ENGLISH HERITAGE

**Primary contact**  
 Jacky Nowakowski  
 Historic Environment Service  
 Cornwall Council  
 Kennall Building, Old County Hall  
 Station Road  
 Truro  
 TR1 3AY  
 UK  
 Tel: 01872 323604  
[Send e-mail enquiry](#)

**Resource identifiers**  
 ADS Collection: 756  
 ALSF Project Number: 3354  
 Collection doi:10.5284/1000200  
[How to cite using this DOI](#)




**Downloads**  
[Project Documents](#) | [Image Gallery](#) | [Data](#)  
 A selection of project reports are available to download. Select a title from the drop-down box below and click the 'submit' button:  

Please select: submit

**Assessment Report PDF files:**

Assessment Report on excavations at Gwithian volume I	PDF	534 Kb
Assessment Report on excavations at Gwithian volume II	PDF	1523 Kb
Figure 08 and 09	PDF	258 Kb
Figure 10 and 11	PDF	193 Kb
Figure 12 and 13	PDF	393 Kb
Figure 14 and 15	PDF	479 Kb
Figure 19 and 20	PDF	434 Kb
Figure 22	PDF	179 Kb
Figure 23 and 24	PDF	240 Kb


**Assessment Report PNG images:**

	Figure 01	PNG	902 Kb
---	-----------	-----	--------

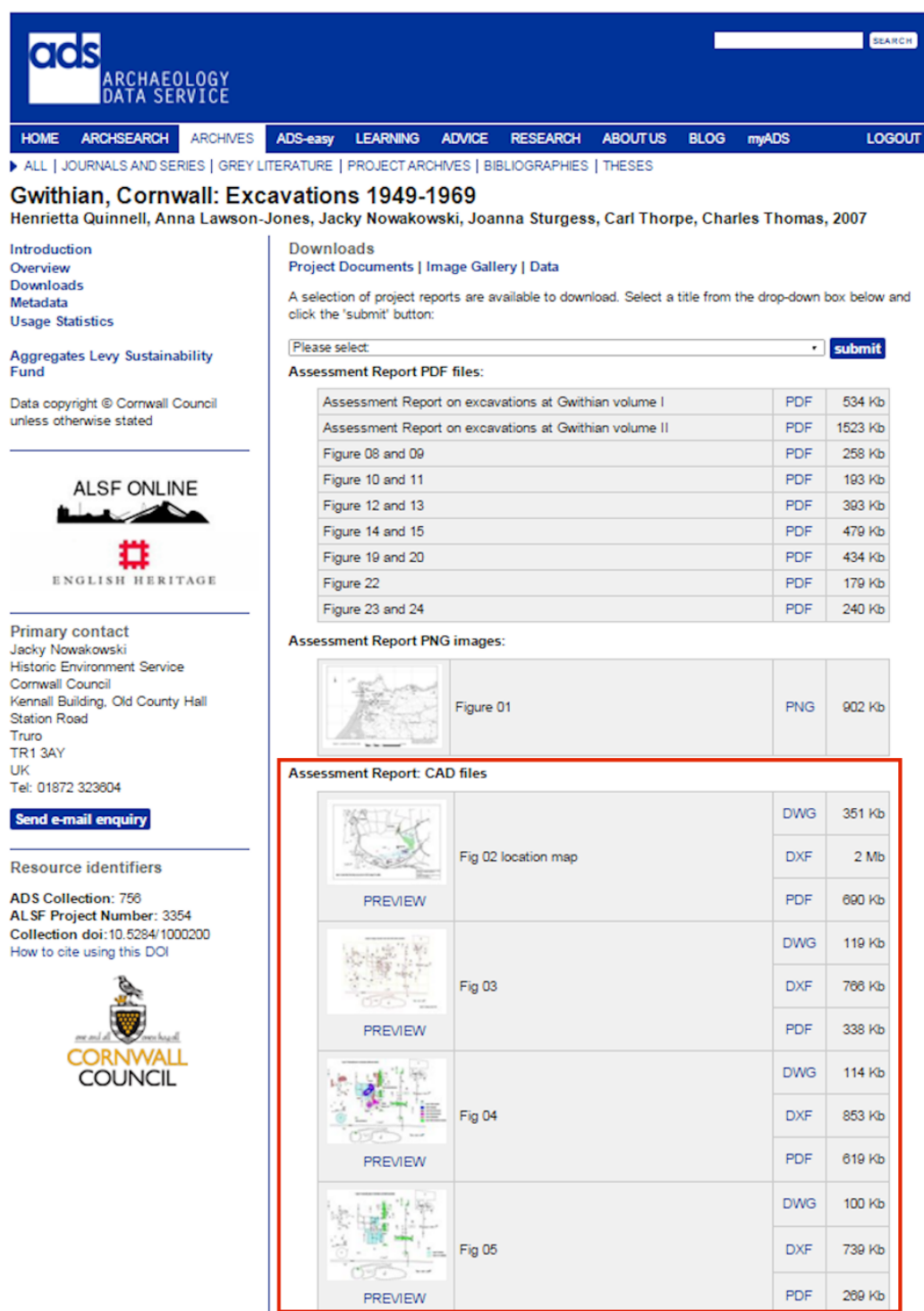
**Assessment Report: DXF files**

Fig 02 location map.dxf	DXF	2467 Kb
Fig 03.dxf	DXF	766 Kb
Fig 04.dxf	DXF	853 Kb
Fig 05.dxf	DXF	739 Kb
Fig 06.dxf	DXF	1127 Kb
Fig 07.dxf	DXF	794 Kb
Fig 16.dxf	DXF	2751 Kb
Fig 17.dxf	DXF	537 Kb
Fig 18.dxf	DXF	691 Kb
Fig 21.dxf	DXF	721 Kb
Fig 25.dxf	DXF	479 Kb

University of York legal statements | [ADS terms of use and access](#) | [Cookies](#)



**Figure 7.** Example of an ADS archive download webpage before the new formats were created. The red boxes highlights the old Drawing Exchange Format (DXF) dissemination files.



**ads** ARCHAEOLOGY DATA SERVICE

HOME ARCHSEARCH ARCHIVES ADS-easy LEARNING ADVICE RESEARCH ABOUTUS BLOG myADS LOGOUT

ALL | JOURNALS AND SERIES | GREY LITERATURE | PROJECT ARCHIVES | BIBLIOGRAPHIES | THESES

## Gwithian, Cornwall: Excavations 1949-1969

Henrietta Quinnell, Anna Lawson-Jones, Jacky Nowakowski, Joanna Sturgess, Carl Thorpe, Charles Thomas, 2007

Introduction  
Overview  
Downloads  
Metadata  
Usage Statistics

Aggregates Levy Sustainability Fund

Data copyright © Cornwall Council unless otherwise stated

**ALSF ONLINE**  
ENGLISH HERITAGE

Primary contact  
Jacky Nowakowski  
Historic Environment Service  
Cornwall Council  
Kennall Building, Old County Hall  
Station Road  
Truro  
TR1 3AY  
UK  
Tel: 01872 323804

[Send e-mail enquiry](#)

Resource identifiers  
ADS Collection: 756  
ALSF Project Number: 3354  
Collection doi:10.5284/1000200  
How to cite using this DOI

**CORNWALL COUNCIL**

Downloads  
Project Documents | Image Gallery | Data


A selection of project reports are available to download. Select a title from the drop-down box below and click the 'submit' button:

Please select [submit](#)



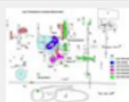

**Assessment Report PDF files:**

Assessment Report on excavations at Gwithian volume I	PDF	534 Kb
Assessment Report on excavations at Gwithian volume II	PDF	1523 Kb
Figure 08 and 09	PDF	258 Kb
Figure 10 and 11	PDF	183 Kb
Figure 12 and 13	PDF	383 Kb
Figure 14 and 15	PDF	479 Kb
Figure 19 and 20	PDF	434 Kb
Figure 22	PDF	179 Kb
Figure 23 and 24	PDF	240 Kb

**Assessment Report PNG images:**

Figure 01	PNG	902 Kb
		

**Assessment Report: CAD files**

Fig 02 location map	DWG	351 Kb
	DXF	2 Mb
	PDF	680 Kb
	DWG	119 Kb
	DXF	766 Kb
	PDF	338 Kb
	DWG	114 Kb
	DXF	853 Kb
	PDF	619 Kb
	DWG	100 Kb
	DXF	739 Kb
	PDF	269 Kb

**Figure 8.** Example of an ADS archive download webpage following the migration. The red box highlights the new popular preview images and the new dissemination formats available to a user.

## 6. CAD Migration Issues and Problems

While the eight steps above provide a general overview of the CAD migration carried out at the ADS, the process as a whole was not without its problems. A major issue arose from the decision to replace the existing DXF R14 preservation format with DWG 2010, essentially reversing an initial preservation process that had been carried out on a significant proportion of our deposited data.

As a large component of the ADS archive had originally been deposited as DWG (1254 files, *ca.* 78%) prior to preservation as DXF R14, it was considered preferable to return to the original SIP files as the source for migration rather than use the normalized DXFs stored in the homogenous AIP. The result of this was that Steps 3 and 4 of the migration process, rather than simply working from a single, normalized input format (*i.e.*, DXF 14), had to cope with the conversion of a mixture of CAD formats from the original SIPs, in places replicating the original ingest and normalization process. In a number of cases this was further complicated by the fact that the original, un-normalized SIP sometimes included multiple versions of the same file in both DXF and DWG formats, and thus required an element of prioritization and selection (DWG over DXF) prior to migration. This reduced the level of automation that could be built into the process and increased the time taken to complete the migration. This issue, however, provides a valuable lesson, in that, the ideal migration activity depicted in the OAIS model is not always achievable or appropriate, and strategic decisions have to be made on a case-by-case basis to achieve the best possible result for the long-term preservation of data. It is important to note the importance of recording these decisions alongside the documentation of the processes undertaken (Step 7) to support future comprehension of the archive.

The creation of PDF/A files for dissemination also caused a further issue, as not all CAD files were generated with appropriate printing layouts. This meant that many of the DWG formats required manual conversion—a highly time consuming process, and also led to some dissemination PDF files that were not wholly representative of the original CAD files. Prior to the migration this had not been a problem, as ingested files were converted to another CAD format and were therefore not required to be formatted appropriately for PDF/A conversion. The positives of increased dissemination capability through the use of PDF/A have been somewhat diminished by the need to check newly ingested CAD data for PDF/A compatible layouts.

## 7. Conclusions

The lessons learnt during the largescale CAD migration process presented in this paper provide an important insight into the digital preservation component of Research Data Management practice.

While the overall migration process presented in this paper was not a strict migration according to the OAIS model and in many cases essentially involved “re-archiving” data, the exercise itself was necessary for the long-term preservation of the data and was undertaken in such a way as to achieve the best possible outcome for both the ADS and data consumers. While elements of the process were both laborious and time consuming (and therefore costly), as a result of having to reassess original files in the SIP, this highlights the benefits of normalizing data at the point of ingest and the production of homogenous AIPs to stable, reliable standards and formats, reaffirming the importance of professional Research Data Management and preservation practices. The importance of the documentation of files at ingest, and the data management and processes carried out on files during archiving, were also emphasized as invaluable assets to the later data migration process. In particular the recording of file location, file type, and the relationships between files within individual collections allowed the overall migration task to be assessed and planned in terms of scope and time required. This allowed the key steps in the migration process to be planned alongside the identification of elements which could be problematic or straightforward. This allowed an assessment of the degree to which each step could be automated to be possible.

This CAD data migration has been a product of a necessary and worthwhile change in policy and has provided a case study on how to proceed with further conversions and data migrations where these are deemed necessary. The additional increase in dissemination formats made possible by the migration has also provided a clear benefit in making the CAD datasets even more accessible for the general public. The new dissemination formats depicted in Figure 8 have allowed a new demographic of data users, who would not have normally utilized CAD files or those who do not have access to appropriate software, to view and re-use CAD data. The new JPG preview formats have also reduced the effort need to assess if a file available is appropriate for their purposes as they no longer need to

download and open the CAD DFX file in appropriate software. The ADS website's user access statistics have seen a significant change to the behavior of users interacting with these archives following the change. Users are now spending more time in a CAD archive collection viewing preview images and then downloading a smaller number of files. Despite adding additional complication, time and cost to the CAD migration process the added benefits for users have resulted in a significant enhancement to the service the ADS provides. This draws attention to the consequences of dissemination data formats on the reusability of data during Research Data Management Planning. It is important to note that this exercise has emphasized that while a certain data type and format may be best for long-term preservation requirements, dissemination of data in formats outside of the original data type can add to the re-use potential of data. The issues experienced and resultant outcomes presented in this paper will significantly influence future migration and data dissemination strategies at the ADS, and the authors hope this case study will be of value to others undertaking data migration as part of their own Research Data Management.

**Acknowledgments:** The authors would like to thank all the ADS curatorial and technical staff for their valuable advice and guidance during the migration process.

**Author Contributions:** Kieron Niven and Georgie Field carried out the steps of the CAD Migration Process described in Section 2 of this article. Katie Green, Kieron Niven and Georgie Field wrote the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References and Notes

1. Department for Communities and Local Government. National Planning Policy Framework. London, 2012. Available online: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/6077/2116950.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6077/2116950.pdf) (accessed on 25 November 2015).
2. Chartered Institute for Archaeologists. *Standard and Guidance for the Creation, Compilation, Transfer and Deposition of Archaeological Archives*; Chartered Institute for Archaeologists: Birmingham, UK, 2014.
3. Richards, J.D. Preservation and re-use of digital data: The role of the Archaeology Data Service. *Antiquity* **1997**, *71*, 1057–1059. [CrossRef]
4. Waters, D.; Garrett, J. *Preserving Digital Information, Report of the Task Force on Archiving of Digital Information*; The Commission on Preservation and Access: Washington, DC USA, 1996.
5. Hardman, C.; Richards, J.D. Digital Curation at the Archaeology Data Service. *Archäol. Informationssyst.* **2013**, *42*, 76–78.
6. Digital Preservation Coalition. Saving the Digital Decade: DPC Rewards Organizations Helping to Safeguard Our Digital Memory. Available online: <http://www.dpconline.org/newsroom/latest-news/945-saving-the-digital-decade-dpc-recognizes-major-accomplishments-to-safeguard-our-digital-memory> (accessed on 25 November 2015).
7. Data Seal of Approval. Implementation of the Data Seal of Approval. 2013. Available online: [https://assessment.datasealofapproval.org/assessment\\_96/seal/html/](https://assessment.datasealofapproval.org/assessment_96/seal/html/) (accessed on 25 November 2015).
8. Archaeology Data Service. Collections Policy. Available online: <http://archaeologydataservice.ac.uk/advice/collectionsPolicy> (accessed on 25 November 2015).
9. Archaeology Data Service. *Guidelines for Depositors*; Archaeology Data Service: York, UK, 2014.
10. Archaeology Data Service. Guides to Good Practice. Available online: <http://guides.archaeologydataservice.ac.uk/> (accessed on 25 November 2015).
11. Open Archival Information System Reference Model. Available online: <http://public.ccsds.org/publications/archive/650x0m2.pdf> (accessed on 25 November 2015).
12. Preservation Policy. Available online: <http://archaeologydataservice.ac.uk/attach/preservation/PreservationPolicyV1.3.1.pdf> (accessed on 25 November 2015).
13. Repository Operations. Available online: [http://archaeologydataservice.ac.uk/attach/preservation/ADS\\_Repository\\_Operations\\_V2.pdf](http://archaeologydataservice.ac.uk/attach/preservation/ADS_Repository_Operations_V2.pdf) (accessed on 25 November 2015).
14. Condron, F.; Richards, J.; Robinson, D.; Wise, A. Strategies for Digital Data. Available online: <http://archaeologydataservice.ac.uk/legacy/strategies/> (accessed on 25 November 2015).

15. Archaeology Data Service. CAD and Vector Images. Available online: <http://archaeologydataservice.ac.uk/advice/FilelevelMetadata#section-FilelevelMetadata-CADAndVectorImages> (accessed on 25 November 2015).
16. Archaeology at Glastonbury Abbey on-Line. Available online: <http://dx.doi.org/10.5284/1000292> (accessed on 29 February 2016).
17. Harrison, Eiteljorg, II. The CSA Propylaea Project. Available online: <http://dx.doi.org/10.5284/1022574> (accessed on 29 February 2016).
18. Anderson, S.; Pringle, M.; Eadie, M.; Austin, T.; Wilson, A.; Polfreman, M. Arts and Humanities Data Service Digital Images Archiving Study. Available online: <http://www.ahds.ac.uk/about/projects/archiving-studies/digital-images-archiving-study.pdf> (accessed on 25 November 2015).
19. Coyne, M.; Duce, D.; Hopgood, B.; Mallen, G.; Stapleton, M. The Significant Properties of Vector Images. JISC. 2007. Available online: [http://www.jisc.ac.uk/media/documents/programmes/preservation/vector\\_images.pdf](http://www.jisc.ac.uk/media/documents/programmes/preservation/vector_images.pdf) (accessed on 25 November 2015).



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).