

Digital data

Author Andrew Fetherston, LAARC
Version 10
Date 26/04/2013
Status Ready for publication
Change History 9: "2.7_Digital_data.doc" Wylie, Joanna [10/10/2011]
8: "2.7_Digital_data.doc" Wylie, Joanna [13/08/2009]
7: "2_7digital_standards_v02.doc" Wylie, Joanna [23/05/2008]
6: "2_7digt_abridged_stand_v01.doc" Evans, Edward (19/02/2007]
5: "2_7digt_depo_stand_v01.doc" Evans, Edward (13/12/2006)
4: "D3S27_redraft_03.doc" by Evans, Edward (06/12/2005)
3: "D3S27_redraft_02.doc" by Evans, Edward (15/11/2005)
2: "D3S27_redraft_01.doc" by Evans, Edward (24/06/2005)
1: "D3S27.doc" by Grew, Francis

Changes to 9.0 Section 2.7.7 note concerning registered finds images added

2.7 Digital Data

“Modern archaeological projects have the capacity to create large quantities of digital information, whether from the on-site recording of excavation and survey data, specialist datasets developed during pre and post-excavation analysis, or publication of interpretative maps and plans. Digital information is created at every stage of a project, from fieldwork to assessment, analysis, and finally reporting and dissemination.”

Archaeology Data Service / Digital Antiquity Guides to Good Practice (2011) '[About these Guidelines](#)'. Edited by Kieron Niven and Francis Pierce-McManamon

This section of these standards is concerned with just such material; as follows:

2.7.0 General

2.7.1 Transfer method

2.7.2 Documentation

2.7.2.1 *Metadata requirements*

2.7.2.2 *Codes/ Schema etcetera*

2.7.2.3 *Relationships between files*

2.7.2.4 *Checksums*

2.7.2.5 *File lists*

2.7.3 Categories of data and associated metadata requirements

2.7.4 Word processed documents: text reports and other documents

2.7.4.1 *Metadata requirements for word processed material*

2.7.5 Databases and spreadsheets

2.7.5.1 *Spreadsheets*

2.7.5.2 *Relational databases*

2.7.5.3 *Metadata requirements for databases and spreadsheets*

2.7.6 CAD/Vector Graphics

2.7.6.1 *CAD*

2.7.6.2 *Vector graphics*

2.7.6.3 *Metadata requirements for CAD and vector graphics*

2.7.7 Raster images

2.7.7.1 *Digital photographs*

2.7.7.2 *Other raster images*

2.7.7.3 *Metadata requirements for raster images (all types)*

2.7.8 GIS

2.7.8.1 *Metadata requirements for GIS*

2.7.9 Geophysics

2.7.9.1 *Metadata requirements for geophysics data*

2.7.10 Matrices

2.7.10.1 *ArchEd (Bonn matrix drawing programme)*

2.7.10.2 *Vector and raster graphics tools*

- 2.7.11 XML
 - 2.7.11.1 *Metadata requirements for XML*
 - 2.7.12 Moving images
 - 2.7.12.1 *Metadata requirements for moving images*
 - 2.7.13 Audio
 - 2.7.14 SPSS
 - 2.7.15 File naming conventions
 - 2.7.16 Glossary
- Metadata Example

2.7.0 General

The purpose of this section of the Standards is to specify **methods of preparing digital data for archiving**.¹ The specific archive components that are to be deposited in digital format are listed in the 'Types and Contents' Section of these Standards (Section 1):

- Project Planning Archive – see 1.1.5
- Site Archive – see 1.2.4
- Field-walking Archive – see 1.2.5
- Standing Building Archive – see 1.2.6
- Post-Excavation Assessment Archive – see 1.3.3
- Research Archive – see 1.4.3

It is essential that data is deposited in the formats specified in this section of the Standards, although we also **strongly encourage** the deposition of files in their 'native' format **in addition** to the formats specified. The Museum of London works closely with the Archaeology Data Service (ADS; <http://ads.ahds.ac.uk/>), the Collections Trust (<http://collectionstrust.orangeleaf.org/> - formerly known as the MDA) and the English Heritage Data Standards (<http://www.english-heritage.org.uk>), to ensure that the standards for its Archaeological Archive are, so far as possible, compatible with **national** standards. This will help guarantee the sustainability of the archives, while at the same time allowing for their integration with broader information structures.

At the same time, we recognise that contractors may sometimes have found it necessary to use specialised applications which cannot export the data into standard deposition formats. In these instances, depositors:

- **Must provide printouts of finished reports** generated from the data;
- May deposit backup copies of the data.

Please note that administration files such as archive indices or box lists need not be deposited in digital format.

¹ The deposition of digital data relates to data that is born-digital; depositors are **not** required to create digital surrogates of hardcopy material.

2.7.1 Transfer method

For the purposes of transfer and ingest, the museum will only accept digital material deposited in the formats specified in Sections 2.7.4 – 2.7.13. The data, however, may be compressed in Zip format. To assist with the transfer of data generated on Unix systems, we are prepared to accept text files compressed with 'Compress' (.Z files) and 'gzip' (.gz files), as well as 'tar' archives of text files. All other forms of compression are unacceptable.

As of November 2010, digital data should be transferred for deposit using **FTP (file transfer protocol)**. FTP is a commonly used method for transferring files between two computers – one is the host or server computer, and the other is the client. In this case, the server is here at the Museum of London, and your computer will be the client machine. If your organisation has not already been issued with an IP address, user account and password to access the Museum of London's FTP server please contact the LAARC Records Officer (Digital) to arrange this.

Once issued with this information, you will need to download an FTP client – a specialist software programme that enables you to securely connect to a FTP server and upload files. There are various FTP clients to choose from, but the LAARC recommends FileZilla which is freely available and very easy to use. To download FileZilla, see: <http://filezilla-project.org/>. Please make sure that you download the FTP client however, and not the FTP server. If you choose to use a different FTP client, please advise the LAARC accordingly.

Once you are connected to the Museum's FTP server, please transfer the virus-checked digital data and accompanying checksums (see Section 2.7.2.4) into your organisation's folder as shown in the directory tree for the server. Please double check that all the necessary files have been transferred prior to disconnecting.

Note: there is no maximum file transfer limit with FileZilla, but large archives will take a long time to transfer and may cause the FTP server to crash. Please contact the LAARC Records Officer (Digital) if the size of your total archive deposit is more than 1GB, so that an alternative transfer method can be arranged.

2.7.2 Documentation

Documentation is an integral part of the digital archiving process, enabling resource management, discovery and use. The term ‘metadata’ is often used interchangeably with documentation, but metadata is best described as a specific type of documentation that is structured and exhibits consistency. Metadata is essentially data about data, summarising a range of elements associated with a digital object’s creation, content, structure and context. Other types of documentation are less structured and don’t fit into standardised metadata schema, but are still as important.

2.7.2.1 Metadata requirements

Metadata requirements are generally conformant with those of the Archaeology Data Service, and are divided into two distinct categories: ‘project-level metadata’ and ‘file-level metadata’. The former relates to the project archive as a whole, whilst the latter relates to the individual files within the archive or data set, as further described below.

Metadata: project level

By the time the archive is ready for deposit most project-level metadata should have been submitted (see Table 1). If this is not the case, please contact the Records Office of the Archaeological Archive. Table 2 presents a number of optional metadata elements that depositors may submit should they wish. If submitting optional metadata elements, please deposit in both hard copy and digital format as a plain text (TXT) file titled ‘Project Metadata’.

Table 1: Pre-submitted project level metadata

<i>Element</i>	<i>Explanation</i>	<i>Example</i>
Project Identifier	Site Code	ABC05
Organisation	The organisation responsible for the work undertaken.	XYZ Archaeology Ltd
Project Name	The name given to the project.	Excavation of 123 High Street
Type of Project	The type of work undertaken.	Excavation, evaluation, watching brief and so on
Site Location	The full address of the site, including postcode.	123 High Street, Atown, AB1
Borough	The borough(s) in which the site is located	Aborough
Project Start Date	The date the project began.	01-03-2005
Project End Date	The date the project concluded	17-03-2005
NGR	The primary grid reference for the site.	512345 198765 or TQ12349876

Table 2: Optional project level metadata

<i>Element</i>	<i>Explanation</i>	<i>Example</i>
Contributor Moral Rights	A statement about who holds moral rights in the resource. This element is optional .	Creator of digital images: Person, A.
Contributor Name	Name of individual associated with performing a particular role in the work undertaken. This element is optional .	Person, A
Contributor Role	Role played by the individual cited above. This element is optional .	Project Manager, Project supervisor, Illustrator
Contributor	Organisation with which the person	Post-Ex Specialists Ltd

Affiliation	cited above is associated if different from element 'Organisation'. This element is optional .	
Related Projects	Identifier for any related projects This element is optional .	XYZ03
Access Statement	A statement about any restrictions imposed on access to a resource. This element is optional .	Full site address cannot be published due to the sensitive nature of the location.

Metadata: file-level

The effective preservation of digital data necessitates the generation of a certain amount of file-level metadata such as file name, file size and file format and version. The file-level metadata requirements for each data category are listed under the general heading of 'Metadata Requirements' in Sections 2.7.4 to 2.7.13.

Depositors must supply at least Level 1 metadata, but as archive users, should also consider supplying metadata at Level 2 or above as this can make the re-use of data significantly easier. Each level builds incrementally on the one that precedes it; i.e. Level 2 should be supplied in addition to Level 1, and Level 3 should be supplied in addition to Levels 1 and 2.

It would be very useful to receive file-level metadata for all submitted files, including 'native' and optional files, but it is mandatory for master files only. For example, if a finds catalogue is submitted for deposition as a tab-delimited text file (TAB), but the native MS Excel version (XLS) is also supplied, metadata will be **required** for the master TAB file, but **optional** for the native XLS file.

N.B: If multiple files are produced using the same combination of software / formats, it is acceptable to record this once, rather than reproduce the information for every file concerned (see example below):

Example:

Word-processed files

All word-processed files were created in Microsoft Word 2003 (v.11), and are RTF (v.1).

"ABC11_method_statement.rtf"
 File size: 70, 822
 "ABC11_watching_brief_report.rtf"
 File size: 111,616
 "ABC11_post_excavation_asst.rtf"
 File size: 234,987

Metadata should be deposited as a plain (TXT) file in both digital format and hard copy, and an example is provided at the end of this section.

2.7.2.2 Codes/ schema etcetera

As prescribed by 'Metadata Requirements', sections 2.7.4 to 2.7.13 - any field names, abbreviations, codes, drafting conventions, XML schema and so on must be fully documented. Please note their use in the archive's metadata file, or if necessary, submit as additional documentation in both hard copy and digital format. With regards to artefacts, environmental material and archaeological feature descriptions however, contractors should be adhering to the **agreed term-lists where available** (see sections 2.3 and 2.5 and Term Lists available from the website). There is no need to provide additional copies of these term-lists as part of the archive's documentation, but please note their use in the archive's metadata file.

N.B: The digital version of any additional documentation submitted with the archive should conform to the relevant requirements (e.g. for word-processed material see 2.7.4, for spreadsheets see 2.7.5).

2.7.2.3 Relationships between files

The preparation of files for deposit may necessitate the breaking down, or grouping, of certain files that require re-constitution or close association to ensure that the data can be understood in the future. Such files may include the export of a number of tables from a relational database, export of a number of worksheets from a spreadsheet or the grouping of the three major file types associated with shapefiles. It is important that these relationships are readily identifiable, and file relationships must therefore be documented as part of the file-level metadata requirements (see 2.7.2.2. above). Further guidance is provided in Sections 2.7.4 – 2.7.16 as appropriate.

2.7.2.4 Checksums

Every file sent for deposit should be accompanied by an appropriate MD5 checksum. A checksum is a numeric value or 'sum' that is used to verify the validity of a file, and MD5 is a particular 'brand' of checksum, with the file extension '.md5'. Once a digital archive has been transferred onto the Museum of London server, new checksums will be computed and compared to the deposited checksums to detect any errors that may have occurred during transfer.

There are several **freely available** programmes for calculating checksums, notably Jacksum and Fastsum:

- Jacksum [1.7.0 is available for download from: <http://www.jonelo.de/java/jacksum/>]
- FastSum [the **free** command line edition is available for download from: <http://www.fastsum.com/> and full user instructions can be found at: <http://www.fastsum.com/support/user%20manual.php>]

Please contact the LAARC Records Officer (Digital) for further assistance about checksums.

Note: There is no requirement to provide metadata for these checksum files.

2.7.2.5 File lists

In sections 2.7.4 – 2.7.13 you are asked to provide the name and size (in bytes) of the files deposited as part of the metadata requirements. Should you be submitting a large number of files it is possible to automatically generate this information in the form of a file list which can then be incorporated into the metadata file. One way of producing a list of files in a particular folder is via 'Open Command Window Here':

- Right click the folder containing the files;
- Select 'Open Command Window Here';
- At the prompt type: dir/s>filelist.txt and then press 'enter'²

The file list will appear inside the folder. An example is provided below.

Note: 'Open Command Window Here' is a Microsoft PowerToy (additional software programme) released for Windows XP. If you have Windows XP but not the PowerToy, it can be downloaded from the Microsoft website – see: <http://www.microsoft.com/windowsxp/Downloads/power toys/Xppowertoys.msp>

If you have Windows Vista, 'Open Command Window Here' is accessible via the following steps:

² The '/s' ensures that files in any sub-folders are also listed.

- Right click the folder containing the files, whilst also holding down the shift key;
- Select 'Open Command Window Here';
- At the prompt type: dir/s>filelist.txt

You can also produce a file list by opening a command prompt window in the normal manner – i.e. via the start menu (by default the window will start in your home directory – H:\):

- At the prompt type dir/s myfolder>listmyfolder.txt. For example, if you wished to produce a list of the files in the folder 'XYZ08' which was located under network drive 'P', you would type: dir/s P:\XYZ08>listXYZ08.txt

The file list, comprising file name and size, will appear under your home directory – H:\.

```
Volume in drive J is Shadow
Volume Serial Number is 4A71-8E64

Directory of J:\REPOSITORY\ingest\new_archives\MOLAS\Deposit_11\AMQ03

24/07/2008 12:09 <DIR>      .
24/07/2008 12:09 <DIR>      ..
16/07/2008 14:35      39,954,506 AMQ03dbacover.tif
26/06/2008 15:40      74,732,706 AMQ03dbafig01.tif
26/06/2008 15:40      26,110,899 AMQ03dbafig02.tif
26/06/2008 15:41      26,110,899 AMQ03dbafig03.tif
09/07/2008 15:22          40,448 AMQ03Images Reg.doc
16/07/2008 14:36      39,932,535 AMQ03wbcover.tif
26/06/2008 15:37      22,263,093 AMQ03wbfig01.tif
26/06/2008 15:38      107,146,408 AMQ03wbfig02.tif
03/07/2008 12:47          44,544 AMQ03_digi_checklist.doc
19/05/2008 11:54          15,360 biblio.xls
22/06/2006 17:15          198,144 dta01.doc
11/04/2003 16:43          308,059 ET1 01.JPG
11/04/2003 16:43          303,744 ET1 02.JPG
11/04/2003 16:43          305,960 ET1 03.JPG
11/04/2003 16:43          302,869 ET1 04.JPG
11/04/2003 16:43          304,730 ET1 05.JPG
11/04/2003 16:43          304,937 ET1 06.JPG
11/04/2003 16:43          307,508 ET1 07.JPG
23/11/2006 15:15          116,736 EVA01.DOC
28/08/2007 10:56          270,058 eva01.dwg
24/07/2008 12:09           30 filelist.txt
26/06/2008 16:18           5,388 metadata.txt
23/11/2006 15:18          105,472 MS01.DOC
18/09/2006 16:44           94,720 MS02.doc
22/06/2006 15:03          111,616 wb01.doc
23/07/2007 17:06          286,497 wb01.dwg
                26 File(s)  339,677,866 bytes

Total Files Listed:
26 File(s)  339,677,866 bytes
2 Dir(s)  769,787,801,600 bytes free
```

Size of file in bytes

Figure 1: Example file-list

2.7.3 Categories of data and metadata requirements

The following sections (2.7.4 - 2.7.14) provide a definition for the category of data considered, a list of commonly used software applications, acceptable deposition formats, guidance notes on preparing files for deposit, and metadata requirements. Sections 2.7.4 – 2.7.7 cover the most commonly deposited data categories [text documents, databases, spreadsheets, vector graphics, raster images], whilst sections 2.7.8 – 2.7.14 cover additional categories of data which may be generated by an archaeological project [GIS, geophysics, matrices, XML, moving images, audio, SPSS files]. A glossary of technical terms is provided in section 2.7.16.

Please note the guidance below is solely concerned with the preparation of digital files for deposit. Guidance governing the content and scope of the data that comprise these files can be found elsewhere in these Standards.

2.7.4 'Word-processed' documents: text reports and other documents

Definition: Formatted text-based document produced by word-processing applications.

Common software applications: MS Word, Corel WordPerfect, MS Works, OpenOffice Writer.

Acceptable deposition formats:

- DOCX (Office Open XML format); **or**
- ODT (OpenDocument Text); **or**
- PDF/A (Portable Document Format/Archive); **or**
- RTF (Rich Text Format); **or**
- TXT (Plain Text – only suitable for very simple text documents)

Deposition of files in their 'native' format (in addition to PDF/A, RTF or TXT format) is also strongly encouraged.

DOCX (ECMA Standard: ECMA-376 and ISO/ICE Standard 29500) - part of the Office Open XML (OOXML) format developed by Microsoft and released with MS Office 2007 as the default file format for MS Word. DOCX is a zip-based file format - XML files and other content are compressed into a zip archive.

ODT Part of the Open Document Format (ODF) for XML-based document formats - ISO Standard 26300. As with DOCX, ODT is a zip file containing XML files and other content.

PDF/A (ISO Standard 19005-1, published September 2005) was created to provide a file format based upon PDF that was suitable for long-term archiving. As stated in *White Paper: PDF/A – The Basics*, the ISO Standard 19005-1:

"...identifies a "profile" for electronic documents that ensures the documents can be reproduced in years to come".³

PDF/A is based upon PDF Reference Model 1.4 (i.e. PDF 1.4) and it comprises two levels: PDF/A-1a (fully compliant with the ISO standard) and PDF/A1-b ('minimal compliance' level). PDF/A-2 (ISO Standard 19005-2, published July 2011) is based upon PDF 1.7 and is not widely used at the present time. Deposition of PDF/A-1 files is therefore preferred.

There are currently a number of software packages that can create PDF/A files, most notably *Adobe Acrobat Professional* (versions 8 onwards) but also *Microsoft Office* 2007 and 2010. Please contact the LAARC Records Officer (Digital) for further guidance about depositing PDF/A files. Please also see the PDF/A Competence Center's website: <http://www.pdfa.org>

Preparation guidelines

- Heavily formatted documents can present many problems and consequently, with a view to archiving, it is advisable to construct documents using minimal formatting and apply the desired 'finish' to a copy of this 'master' for printing. **If possible:**
 - Restrict formatting to the structuring of the document;
 - Use numbered sections rather than page numbers
 - Meaning should not be encoded through formatting;
 - The use of special characters should be avoided;

³ PDF Tools AG, 2007, *White Paper: PDF/A – The Basics*. Accessed on 17/04/2009 from PDF-tools.com: <http://www.pdf-tools.com/public/downloads/whitepapers/whitepaper-pdfa.pdf>.

- Footnotes should be rendered as endnotes.

- Avoid the inclusion of external links or dynamic content as these links can break over time
- If saving a PDF version of a document for dissemination purposes, ensure that it conforms to PDF/A or at least PDF 1.4 standards, and that the original version is also retained (i.e. *.doc file)

Embedded content:

If depositing in RTF or TXT format:

- Embedded content such raster graphics should be prepared for deposit as a series of individual files as set out in the appropriate section(s) of these standards.
- Captions and titles should be left *in situ* within the document **and** the 'extracted' file's name should be inserted immediately beneath; for example:

Figure 1 Site Location Plan

"ABC10_evaluation_rpt_fig1.tif "

If depositing a document with embedded content as a PDF/A or XML-based file format (DOCX or ODT), it is also *recommended* that the embedded content is provided separately as individual files in appropriate preservation formats.

Note: word-processed drafts of published texts should not normally be deposited (it is the responsibility of the contractor to maintain adequate backups until the work has been published). Camera-ready copy, in proprietary typesetter or desktop publishing formats, cannot be archived; however, we encourage contractors to convert and deposit such files in PDF/A format. In the case of major reports, the title and author entries should match the bibliographical entries (see Section 2.1.21).

2.7.4.1 Metadata requirements for word-processed material

Level 1: Minimum Requirements

- File name;
- File size (bytes);
- File format (and version if known);
- Software (and version if known) used;
- File relationship (if applicable; see section 2.7.2.3).

Level 2: Preferred

- Nothing further required.

Level 3: Best Practice.

- Nothing further required.

2.7.5 Databases and spreadsheets

2.7.5.1 Spreadsheets

Definition: An application that manipulates numerical and string data in rows and columns. Comprised of one or more 'rectangular objects' (i.e. worksheets). May contain embedded images (charts, graphs and so on). Data can be directly entered or calculated. Cells can be formatted in a number of ways.

Common software applications: Microsoft Excel, OpenOffice Calc, Lotus 1-2-3 and Quattro Pro

Acceptable deposition formats:

- Delimited text (tab-delimited ASCII or UNICODE files preferred); **or**
- XLSX (Office Open XML format); **or**
- ODS (OpenDocument Spreadsheet); **or**
- PDF/A (Portable Document Format/Archive) – only acceptable for spreadsheets containing significant formatting which will be lost during export to delimited text or migration to XLSX or ODS format.

If depositing delimited text or PDF/A files, please also deposit the original 'native' versions of the spreadsheet(s).

XLSX (ECMA Standard: ECMA-376 and ISO/ICE Standard 29500) - part of the Office Open XML (OOXML) format developed by Microsoft and released with MS Office 2007 as the default file format for MS Excel . XLSX is a zip-based file format - XML files and other content are compressed into a zip archive.

ODS Part of the Open Document Format (ODF) for XML-based document formats - ISO Standard 26300. As with XLSX, ODT is a zip file containing XML files and other content.

PDF/A [See Section 2.7.4 above]

Preparation Guidelines:

When creating spreadsheets:

- Ensure data accuracy through the use of controlled vocabularies and agreed terms where possible (see also Section 2.7.2.2);
- Ensure that the data is atomised (see Glossary) as far as **practicable**;
- Ensure that field names are on the first row only, as database, spreadsheet and statistical package applications commonly assume field names are restricted to the first row of imported text files;
- Make sure that Cell 'A1' does not contain text "ID" ("id" is acceptable), otherwise when importing a text file MS Excel (if the chosen package for rendering the data) will assume the file is in SYLK (Symbolic Link) format;
- Avoid the use of special characters or delimiters within the data as these can create problems during data migration;
- Avoid the use of blank rows as separators within a worksheet;
- Avoid the use of significant formatting to convey meaning;
- Avoid the use of text qualifiers
- Avoid the use of hidden or protected data

To prepare a spreadsheet for deposit:

- Save/export the worksheet as an ASCII or UNICODE tab-delimited text file;

- If necessary (this will depend on the software the worksheet was created in), manually change the file's extension from 'txt' to 'tab' by right-clicking on the file and selecting 'Rename';
- **Repeat** this process if the spreadsheet contains several worksheets of data that all require preservation. The exported files should be named as follows:

Spreadsheet name + worksheet name

For example:

"ABC10_pottery_data_form.tab"
"ABC10_pottery_data_fabric.tab"
"ABC10_pottery_data_decoration.tab"

Embedded content:

If depositing delimited text files:

- Embedded images (i.e. graphs or charts) should be extracted and prepared for deposit as individual files in suitable preservation formats;
- Complex macros, formulae or queries may require separate preservation, and should be documented in a text file. The relationship between this text file and the exported data must be documented in the metadata and the relationship between the files should be clear from their file names;
- Any comments or notes contained within the data should also be extracted and stored either in an additional field (preferable) or in a separate plain text file;
- Ensure that the relationship between any extracted embedded content and the source file(s) is documented in the metadata. Logical file names should also be used, e.g.:
"ABC10_pottery_data_fabric.tab" (source file) > >
"ABC10_pottery_data_fabric_queries.txt"
"ABC10_pottery_data_fabric_notes.txt"
"ABC10_pottery_data_fabric_chart1.png"

Note: it is also *recommended* that any embedded content is provided separately as individual files when depositing data in XML-based formats (i.e. XLSX or ODS).

Finally, exported data should be validated to ensure that:

- Each worksheet has been exported;
- Each data file contains the correct number of columns and rows, and that no record breakage has occurred;
- Any numbers are precise and in the correct format;
- Any dates are precise and in the correct format;
- Text fields do not contain any unknown characters (indicate by '□');
- Any coded fields and data are fully documented (see 2.7.2.2)

2.7.5.2 Databases

Definition: Data is stored in tables consisting of records and fields (rows and columns). Primary keys are used to uniquely identify records and to link rows in different tables. Desktop versions typically contain easy to use tools for searching and querying the database. There are different types of databases, but archaeological data is most commonly stored in relational databases.

Common software applications: Microsoft Access, OpenOffice Base, Claris FileMakerPro, MySQL and Oracle

Acceptable deposition format:

- Delimited text (tab-delimited ASCII or UNICODE files preferred)

If the database was created in Microsoft Access or OpenOffice Base, please also deposit the original 'native' version (.accdb, .mdb or .odb file).

Preparation Guidelines:

When creating a database:

- Ensure data accuracy through the use of controlled vocabularies and agreed terms where possible (see also Section 2.7.2.2);
- Ensure that the data is atomised (see Glossary) as far as **practicable**;
- Ensure that field names are on the first row only, as database, spreadsheet and statistical package applications commonly assume field names are restricted to the first row of imported text files;
- Avoid the use of special characters or delimiters within the data as these can create problems during data migration;
- Avoid the use of significant formatting to convey meaning;
- If used, ensure that memo fields do not contain any end of line code that could cause line-breakages;⁴

To prepare a database for deposit:

- Save/export each table as an ASCII or UNICODE tab-delimited text file (do not use text qualifiers during the export process);
- If necessary (this will depend on the software the worksheet was created in), manually change the file's extension from 'txt' to 'tab' by right-clicking on the file and selecting 'Rename';
- Note the relationship between these exported files in the archive's metadata document or through the provision of an Entity Relationship Diagram (ERD).
- Ensure that the exported files are named as follows:

Database name + table name

For example:

"ABC10_human_bone_completeness.tab"

"ABC10_human_bone_metric_data.tab"

"ABC10_human_bone_pathology.tab"

In some cases you may need to alter table names if they are too long or contain special characters or punctuation prohibited within filenames, but please ensure that they are as close to the original table names as possible.

Embedded content:

- Embedded images (i.e. graphs or charts) should be extracted and prepared for deposit as individual files in suitable preservation formats;
- Any comments or notes contained within the data should also be extracted and stored either in an additional field (preferable) or in a separate plain text file;
- Ensure that the relationship between any extracted embedded content and the source file(s) is documented in the metadata. Logical file names should also be used – e.g.:

⁴ Memo fields are used to store large amounts of text in a database.

“ABC10_human_bone_pathology.tab” (source file) > >
“ABC10_human_bone_pathology_notes.txt”,
“ABC10_human_bone_pathology_113.tif”.

Forms, reports, queries and macros:

With databases, preservation focuses primarily on the data tables and the relationship between them. In general, any forms, reports, queries or macros that have been generated within a database are considered non-significant, and do not need to be archived. Please check with the LAARC Records Officer (Digital) if you are unclear about which elements of a database should be archived.

Exported data should be validated to ensure that:

- Each table has been exported;
- Each data file contains the correct number of columns and rows, and that no record breakage has occurred;
- Any numbers are precise and in the correct format;
- Any dates are precise and in the correct format;
- Text fields do not contain any unknown characters (indicate by ‘□’);
- The correct number of characters is present in any memo fields
- Any coded fields and data are fully documented (see 2.7.2.2)

2.7.5.3 Metadata requirements for Databases and Spreadsheets

Spreadsheets

Level 1: Minimum Requirements

- File name;
- File size (bytes);
- File format (and version where applicable);
- Software (and version if known) used;
- Purpose and content of each spreadsheet;
- Purpose and content of each worksheet;
- Field descriptions (except in cases where you have used a MOL proforma such as the MOL finds register proforma);
- Column and row counts for each worksheet;
- Key for any codes and abbreviations employed (see also section 2.7.2.2);
- File relationships (when multiple worksheets have been exported from a single spreadsheet)

Level 2: Preferred

- Data type for each column;
- Calculations checked

Level 3: Best Practice.

- Nothing further required

Databases

Level 1: Minimum Requirements

- File name;
- File size (bytes);
- File format (and version if known);
- Software (and version if known) used;
- Purpose and content of database;
- Purpose and content of each table;
- Field descriptions;
- Key for any codes and abbreviations employed (but see 2.7.2.2);
- File relationships (if applicable; see section 2.7.2.3).

For relational databases, please also document:

- The primary key for each table;
- The foreign key(s) for each table;
- The relationships between the tables - please provide an Entity Relationship Diagram or document the relationships in the metadata file

Level 2: Preferred

- Data type and field length for each field;
- Column and row counts

For relational databases, please also document:

- The purpose and cardinality (e.g. 'one-to-one' or 'many-to-one') of relationships

Level 3: Best Practice.

- Nothing further required.

2.7.6 CAD/ Vector Graphics

2.7.6.1 CAD

Definition: Computer Aided Design: The use of applications, usually vector-based (see below), for the design and rendering of graphical data. May be a 'hybrid'; containing some raster components (see 2.7.7).

Common software applications: AutoCAD, DesignCAD, TurboCAD, FastCAD and EasyCAD.

Acceptable deposition formats:

- Drawing Exchange File (DXF) (preferred); **and/or**
- Scalable Vector Graphics (SVG)

Deposition of files in their 'native' format (in addition to the DXF format) is also strongly encouraged.

Preparation Guidelines:

To prepare the data for deposit:

- Purge drawing of unused layers, blocks, line-type and text;
- Ensure all externally linked elements are present and can be loaded;
- Bind external elements where possible;
- If externally referenced material remains, prepare for deposit according to the appropriate section of these standards and record the relationship(s) between these files and the primary drawing file in the metadata file.
- All drawings to use metres as drawing units and be on Ordnance Survey National Grid coordinate system (OSGB36)

It is also recommended that the use of hatching is avoided (where possible) when creating CAD drawings.

Conversion:

CAD files contain two different 'data' types, the model and the layout(s). These data types should be saved separately:

- Save the model as a DXF file (so that CAD functionality is retained).
Use the most recent version available.
- Save the layout(s) as TIFF or PDF/A files.

If **possible**, it is **advisable** to also create, and deposit, a Scalable Vector Graphics (SVG) version of the model, as this format is less sensitive to version change than DXF.

OS (Ordnance Survey) Material:

All deposited data should be purged of any OS material, but must be **fully geo-referenced onto OS National Grid coordinates**. Raster scans of first edition maps that are out of copyright may be retained.

2.7.6.2 Vector Graphics

Definition: An approach to graphics that deals with individual shapes such as lines, polygons and text, and groups of such objects. Objects are represented as geometric entities, stored as

co-ordinates with formulae describing the path between them. These properties are stored in a database that is transformed by appropriate software into a meaningful image. Can describe two- or three-dimensional geometries and may be 'hybrid'; containing some raster components (see 2.7.7).

Common software applications: Adobe Illustrator, Macromedia Flash, CorelDraw, Paint Shop Pro (see also CAD).

Acceptable deposition formats:

- Drawing Exchange File (DXF) (preferred); **and/or**
- Scalable Vector Graphics (SVG)

Deposition of files in their 'native' format (in addition to the DXF format) is also strongly encouraged.

Preparation Guidelines:

See 2.7.6.1

Also, any **location plans** produced using vector graphic software such as *Adobe Illustrator* and *CorelDRAW* must include at least two marked points labelled with full OS National Grid coordinate values (e.g. 6 figure eastings/6 figure northings), since vector graphic formats are not intrinsically geo-referenced.

2.7.6.3 Metadata requirements for CAD and Vector Graphics

Note: if the required metadata is supplied as part of the digital file (e.g. in a header/footer), it does not need to be reproduced again in the metadata document. Instead, please write 'metadata provided in drawing header/footer' as appropriate.

Level 1: Minimum Requirements

- File name;
- File size (bytes);
- File format (and version if known);
- Software (and version if known) used;
- Caption;
- Scale of plot layout
- Explanation for the significance of any drafting conventions used (e.g. layers, colours, line weight, line type, hatch styles, symbols and so on). If your organisation has adopted a standard layer-naming convention, please reference this, and provide a copy of the convention as additional documentation. For further information on layer-naming conventions see the ADS publication - *CAD: A Guide to Good Practice* (current revision: http://guides.archaeologydataservice.ac.uk/g2gp/Cad_Toc)
- File relationships (if applicable; see section 2.7.2.3).

Level 2: Preferred

- Summary of surveying methodology used

2.7.7 Raster Images

Definition: Raster images (also known as ‘bitmap’ images) are composed of connected ‘dots’ or pixels. Each pixel contains a defined value about its colour, size, and location in the image. The most common types of raster images produced during the course of an archaeological project are digital photographs, document scans, and output files from vector software applications (for example, a TIFF export of an *Adobe Illustrator* or *CorelDRAW* file).

Further guidance on raster images can be obtained from JISC Digital Media - <http://www.jiscdigitalmedia.ac.uk/>

Common types: TIFF (Tagged Interchange File Format), PNG (Portable Network Graphics), JPEG (Joint Photographic Experts Group), GIFF (Graphic Interchange File Format), BMP (Bitmap).

2.7.7.1 Digital photographs

Acceptable deposition formats:

Primary Photographic Record –

- Uncompressed baseline Tagged Interchange File Format (TIFF) v.6 [preferred]; **or**
- Adobe Digital Negative (DNG) [allowed]

Other Photographs (other fieldwork, studio and external photographs, see Section 2.2.1.2)

- Uncompressed baseline Tagged Interchange File Format (TIFF) v.6 [preferred]; **or**
- Adobe Digital Negative (DNG); **or**
- Portable Network Graphics (PNG); **or**
- Joint Photographic Experts Group (JPEG)*

Registered finds images

- Joint Photographic Experts Group (JPEG)
(see separate Guidance for recording registered finds document)

‘Surrogate’ versions

Joint Photographic Experts Group (JPEG)*

* JPEG2000 is not currently an acceptable deposition format.

Preparation Guidelines:

Most digital cameras capture images in JPEG format, but the Primary Photographic Record must be compiled using a high quality camera that can capture in RAW format. The camera must **at least** have the comparable resolving power of a traditional 35mm SLR film camera with interchangeable lenses. For archive deposition, the RAW files must be exported directly from the camera as uncompressed TIFF files, and not converted to uncompressed TIFF files using image processing software which can introduce a lot of variables. If you have chosen to convert RAW files to DNG for archive deposition, please note the details of the Adobe conversion software used in the archive’s metadata document, for example “Adobe DNG Converter 6.3”.

Wherever possible, please also deposit a 'surrogate' version (see 2.2.1.1) in JPEG format for any TIFF or DNG files deposited.⁵

For other photographs (see Section 2.2.1.2), PNG files or the highest quality / lowest compression JPEG files practicable may be deposited.

If you wish to deposit edited digital images please supply these in one of the acceptable deposition formats outlined above, and fully document the alterations undertaken in the metadata file. See also Sections 2.2.2.3 and 2.7.15 of these standards.

[Please see also Section 2.2 [*Photographs*] of these Standards]

Other raster images

Acceptable deposition formats:

- Tagged Interchange File Format (TIFF); **or**
- Portable Network Graphics (PNG); **and/or**
- Joint Photographic Experts Group (JPEG)

Portable Document Format/Archive (PDF/A) is also acceptable

Preparation Guidelines:

- Capture as uncompressed TIFF (v.6) of appropriate resolution where **practicable**, or PDF/A;
- Layers should be combined into a single image, where applicable;
- If **practicable**, please supply a copy in JPEG format for any TIFF submitted.

2.7.7.1 Metadata requirements for raster images (all types)

Depositors are **strongly** encouraged to deposit at least Level 2 metadata.

Level 1: Minimum Requirements

- File name (for digital photographs this will generally be as assigned by the camera. Image re-naming is permitted but adequate documentation must be provided so that the file names can be matched against those recorded in the on-site photographic logs/ indexes);
- File size (bytes);
- File format (and version if known);
- Where applicable, software (and version if known) used;
- Descriptive caption (this can be copied and pasted directly from the Images Register);
- Date of capture (for digital photographs this is the date that the photograph was physically taken - see screen shot below);
- Alterations (if depositing edited images)

⁵ The deposition of JPEG 'surrogate' versions will become a requirement in the next revision of the Standards

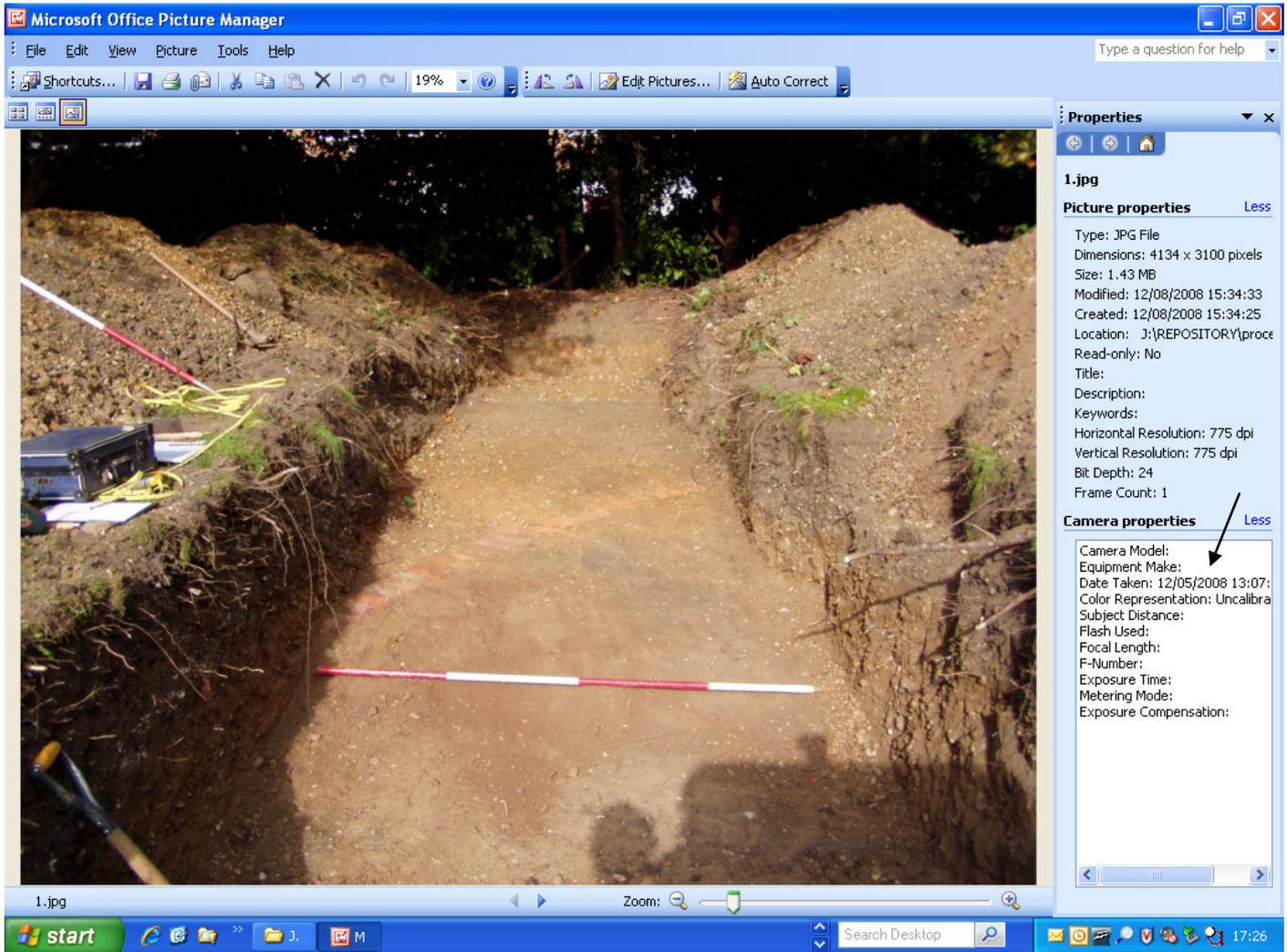


Figure 2: Screen shot illustrating 'date captured'

Level 2: Preferred

- Capture device,
- Device manufacturer,
- Model name;
- Model code;
- Number;
- Capture software;
- Colour management system;
- Resolution;
- Colour space;
- Bit depth;
- Photographer

Level 3: Best Practice

- Technique;
- Period;
- Monument type (if applicable);
- Object type (if applicable).

Digital photos and embedded metadata:

A number of the required Level 1 and Level 2 metadata fields are captured automatically by the camera (mostly in the EXIF format), and additional custom metadata fields (in IPTC or XMP formats) may also be stored internally within digital photos, though this depends upon an organisation's digital photography procedures.

If your organisation **does not add** any custom metadata fields:

- **List** any automatically captured embedded metadata fields in the archive's metadata document – please see the metadata example in the Appendix.

[It may be necessary to check that each deposited image does contain embedded metadata – especially if different cameras have been used to capture the images].

If your organisation **does add** custom metadata fields:

Please **advise** the LAARC prior to depositing any archives with digital photos. We can then discuss with you the most appropriate way of preserving this data with reference to your specific image management procedures and software.

2.7.8 GIS

Definition: Geographical Information Systems. A computer hardware and software system designed for capturing, storing, checking, integrating, analysing and displaying data about the earth that is spatially referenced. Combines relational databases with spatial interpretation and outputs; often in the form of maps.

Common software applications: ESRI ArcInfo; ArcView and ArcGIS; MapInfo Professional, GRASS; IDRISI Kilimanjaro; Intergraph

Acceptable deposition formats:

Preferred

- Shape files (SHP, SHX and DBF (all **required**));

Accepted

- Georeferenced Tagged Interchange File Format (TIFF) and Tiff World File (TWF); **or**
- DXF (graphics) **and** DBF (database) files

Preparation Guidelines:

The AHDS GIS Guide to Good Practice provides a valuable reference for projects deploying GIS technology (current revision: http://guides.archaeologydataservice.ac.uk/g2gp/Gis_Toc)

- Ensure all three required shape files are present and the relationship between them is readily apparent;
- Ensure that the scale and resolution of all components is consistent;
- Ensure that the scale is appropriate to the particular data type;
- Use standard notation for all co-ordinate values;
- For attribute data, ensure field names have not been truncated and that floating number points have not been cut off at the decimal point;
- Please supply a hardcopy printout of graphical elements for validation purposes.

OS (Ordnance Survey) Material:

All deposited data should be purged of any OS material, but must be **fully geo-referenced onto OS National Grid coordinates**. Raster scans of first edition maps that are out of copyright may be retained.

2.7.8.1 Metadata requirements for GIS

Level 1: Minimum Requirements

General:

- File name;
- File size (bytes);
- File format (and version if known);
- Software (and version if known) used;
- Purpose of the GIS;
- Coverage[s] contained within the GIS;
- File relationships (see section 2.7.2.3).

For each layer or coverage:

- Method of capture;
- Data source;
- Scale / resolution of data capture;
- Scale / resolution of data storage;
- Assessment of data quality (e.g. root mean square error);
- Date of capture;

For **attribute data**, please follow the metadata guidelines for relational databases (see section 2.7.5.3).

Level 2: Preferred

- Date of any change/modification;
- Reason for change / modification;
- Record numbers affected by change / modification.

For **attribute data**, please follow the metadata guidelines for relational databases (see section 2.7.5.3).

Level 3: Best Practice

No further documentation required, but please see below.

For **attribute data**, please follow the metadata guidelines for spreadsheets and relational databases (see section 2.7.5).

Other Documentation

If aerial photographs have been incorporated into the GIS, where **practicable** please provide:

- Full photographic details;
- Details of scanning processes;
- Rectification details and software used;
- Ground control points.

For terrestrial survey data, where **practicable** please provide:

- Instrument details;
- An assessment of the reliability of the survey base;
- The scale of original maps, factors of enlargement / reduction, map name and series title

2.7.9 Geophysics

Definition: A non-intrusive prospecting technique used to determine the location and character of buried archaeological features. Data is immediately stored in digital form by the recording instrument, and later downloaded to a computer. This data is captured as x, y and z coordinate values, and commonly rendered as a grey-scale raster image for analysis.

Common software applications: Geoplot, InSite, Contor

Acceptable deposition formats:

- *Coordinate component:* tab-delimited text files (TAB - ASCII or UNICODE (preferred);
and
- *Raster rendering:* Tagged Interchange File Format (TIFF)

Preparation Guidelines:

A geophysical survey will generally consist of a set of contiguous grids, each stored as a separate data file. Data should be prepared for archiving at this file level. Although exportable as ASCII, co-ordinate data contains inclusions specific to particular software products, effectively rendering it proprietary. It is therefore **essential** to also deposit a raster rendering of the data in an archival format (TIFF). 'Database' and pictorial data should be prepared for deposit according to the appropriate section of these standards. In addition:

- The data *per se* should be exported as a tab-delimited (ASCII or UNICODE) text file;
- Plots should also contain a scale bar, co-ordinate lines or pass marks, a legend, and be **securely** geo-referenced.

2.7.9.1 Metadata requirements for Geophysical Data

Level 1: Minimum Requirements

- File name;
- File size (bytes);
- File format (and version if known);
- Software (and version if known) used;
- Purpose of survey;
- Survey instrument;
- Instrument settings;
- Land use of survey area;
- Soil condition;
- Weather conditions;
- File relationships (if applicable; see section 2.7.2.3).

Level 2: Preferred

- Surveyor's name (can also be recorded in the optional project metadata field 'contributor')
- Relationships between grids/transects;
- Direction of travel for each pass;
- Sampling interval within the grids/transects;
- Underlying solid geology;
- Drift geology;
- Classification of any monument(s), if applicable;
- Schedule Ancient Monument Number(s), if applicable.

Level 3: Best Practice

- Images detailing location, orientation, grids, and so on of the survey.

Other Documentation

Please see sections 2.7.5 and 2.7.7 regarding the metadata required for databases / spreadsheets and raster images respectively.

2.7.10 Matrices

The following procedures should be adopted for matrices drawn digitally:

2.7.10.1 *ArchEd (the 'Bonn matrix-drawing program')*

Definition: A program for producing Harris Matrix diagrams, developed from the earlier DOS-based Bonn Seriation and Archaeological Statistics Package (BASP).

Acceptable deposition formats:

- Export as a Bonn Program *.lst file (ASCII text file) **and** a *.bmp file.

Additionally, a hardcopy plot (see Section 2.1.8.2) should be deposited.

Note: if you have produced a matrix using BASP, export it as a *.lst file and then import it into *ArchEd* to export a *.bmp version.

2.7.10.2 *CAD/Vector Graphics or Raster Image tools*

See Sections 2.7.5 and 2.7.7 respectively. A raster rendering should be deposited for matrices produced using vector graphics tools (see 2.7.7). Additionally, a hardcopy plot (see Section 2.1.8.2) should also be deposited.

2.7.10.3 *Other software applications*

It is not recommended that matrices are produced in other software applications such as MS Word or MS Excel, however if matrices have been produced using these programmes, they must be converted to **PDF/A** for deposition. Converting to RTF, TXT or tab-delimited text (TAB) format will result in a loss of formatting to the matrix, rendering it illegible.

Note: if you are unable to create a PDF/A file for deposition, please deposit the matrix in its original format (i.e. *.doc or *.xls file).

2.7.11 XML

Definition: XML (eXtensible Markup Language) is a standard for creating markup languages that describe the structure of data. It provides a textual file format for representing data, a schema for describing data structure, and a mechanism for extending and annotating HTML with semantic information. XML is defined in the [World Wide Web Consortium's \(W3C\) XML 1.0 Specification](#).

[See also Section 2.7.4 (DOCX and ODT) and Section 2.7.5.1 (XLSX and ODS)]

Acceptable deposition formats:

- XML **and** Document Type Definition (DTD); **or**
- XML **and** XML Schema Definition Language (XSDL) **and** an appropriate style sheet (XSL)

Preparation Guidelines:

Well-formed, native XML should possess a valid DTD (Document Type Definition), or XSDL (XML Schema Definition Language) and a Stylesheet (XSL). The DTD explicitly defines the structure of the tags for an XML document; it defines the 'rules' for the document and must be **validated** against appropriate W3C standards. The Stylesheet (XSL) specifies how the data is presented in human readable form.

2.7.11.1 Metadata requirements for XML

- Comprehensive DTD or Schema;
- Ensure codes, abbreviations and so on are readily comprehensible.

2.7.12 Moving Images

Definition: A sequence of still images that create the impression of motion, often accompanied by synchronised audio. Includes: animations, movies, television programmes and so on.

Acceptable formats for deposit:

- Moving Picture Experts Group (MPEG-1 or MPEG-2).

If audio data is supplied separately: Waveform Audio (WAV) or Audio Interchange (AIF).

Preparation Guidelines:

Conversion pathway

- If 'born-digital': record using a standard DV camera and convert to MPEG-2;
- If digitised from analogue, MPEG-2 remains the preferred format, although MPEG-1 is also acceptable for low quality video with a picture size of 352 x 288, a maximum bit-rate of 1151929 bits / second and a single audio stream.

Format Conversion

- Locate appropriate codecs and conversion software capable of handling the ingest format and outputting MPEG-1 or MPEG-2 (see <http://www.divx-digest.com/>);
- The output should be the same quality as the original. No video resizing, channel merging or compression techniques should be employed;
- If audio is provided separately, migrate it to an acceptable format (see above).

2.7.12.1 Metadata requirements for moving images

Level 1: Minimum Requirements

General

- File name;
- File size (bytes);
- File format (and version if known);
- Software (and version if known) used;
- File length (hours, minutes and seconds);
- Relationships between files (if applicable; see section 2.7.2.3).

Movie:

- Name and version of codec;
- Video dimensions (in pixels);
- Frame rate (fps);
- Bit rate;

Audio

- Name and version of audio codec.

Level 2: Preferred

- Copy protection, if applicable;

- Purpose / use of video.

Audio

- Sample frequency;
- Bit rate;
- Channel information.

Level 3: Best Practice

Nothing further required.

2.7.13 Audio

Definition: Sound pressure waves that have been encoded in a digital form for processing, storage, or transmission. Audio is often used in archaeology during building recording work.

Acceptable formats for deposit:

- Waveform (WAV); **or**
- Audio Interchange (AIF).

Preparation Guidelines:

- Capture data at a sample rate which does not lead to the deterioration of the original sound;
- If data is captured in a format other than those set out under acceptable formats above, convert to Waveform or Audio Interchange Format. There are many free products available to perform this conversion (see <http://www.mp3-to-wav.net/>). The output should be configured at the same or higher bit rate and sample frequency as the original recording;
- Ensure that embedded metadata has not been lost during conversion or manually record this in a text file (*.txt).

2.7.13.1 Metadata requirements for audio files

Level 1: Minimum Requirements

- File name;
- File size (bytes);
- File format (and version if known);
- Software (and version if known) used;
- Length of recording;
- Name and version of audio codec;
- Purpose and use of the recording;
- Copyright clearances (where applicable; e.g., oral history);
- Relationships between files (if applicable; see section 2.7.2.3).

Level 2: Preferred

- Sample frequency;
- Bit rate;
- Channel information.

Level 3: Best Practice

- Transcriptions of interviews or similar (where applicable).

2.7.14 SPSS

Although we are prepared to accept SPSS (Statistical Package for the Social Sciences) portable files (*.por) for deposit, currently we do not have the necessary software to perform any validation or processing actions on the data. Consequently, we are unable to offer any guarantees regarding the longevity of these files. We encourage depositors to export data from SPSS files as tab-delimited text and internal metadata as plain / rich text. Regardless of deposition format, the depositor is advised to ensure that the metadata conforms to the guidelines set out in Section 2.7.5.3 (Databases and Spreadsheets).

2.7.15 File naming conventions

Wherever possible, the following file naming conventions should be employed:

- Include the site code (digital photographs are exempt if this is not practicable);
- Use alphanumerical characters only;
- Punctuation and 'special' characters should be avoided;
- White space should be replaced by an underscore ('_');
- All file names should be in lowercase characters (apart from the site code);
- Descriptive file names should be used **whenever possible**.

If a file has been deposited in two different file formats, the name should be the same for both - the file extension will distinguish between them (e.g. 'Figure1.tif' and 'Figure2.jpg'). The same convention should be applied when multiple deposit formats have been derived from a single source file (e.g. *.lst and *.bmp exports from ArchEd or BASP).

See section 2.2.3 of these Standards for the naming of digital photographs.

2.7.1.16 Glossary

TERM	DEFINITION
Algorithm	A mathematical function that is used to encrypt and decrypt information.
Application	A software program designed to perform a specific task.
ASCII	American Standard Code for Information Interchange; assigns a number to each key on the keyboard. Does not include special formatting features.
Atomised Data	Data that has been broken down into its smallest constituent unit to facilitate analysis. For example, atomised address data is represented by the fields "House Number", "Street", "Town", "County" and "Postcode", rather than the single field "Address".
Attribute Data	Data that relate to a specific, precisely defined location, linked in the GIS to spatial data that define the location.
Bit Depth	The number of bits used to represent each pixel in an image.
Bit Rate	The rate at which bits are transmitted over a communication path. Typically expressed in bits per second (bps).
Bit Sequence	A series of binary digits.
Cardinality	With relational databases, the cardinality of a relationship describes how many of the records in a parent table can be directly related to records in a child table. Terms such as 'one-to-one', 'one-to-many', 'many-to-one', or 'many-to-many' are applied when describing cardinality.
Category-Level Metadata	Metadata pertaining to the category classification of MAP(2).
CD-R	Optical disc to which data can be written. Can only be recorded once and cannot be altered.
CD-ROM	Compact Disc Read-Only-Memory
Checksum	An alphanumeric error detection algorithm that comprises a count of the number of bits in a transmission unit and is included with the unit allowing the receiver to check whether the same number of bits arrived.
Codec	Short for compressor/decompressor. A codec is any technology for compressing and decompressing data.
Colour Model	An abstract mathematical model describing the way colours can be represented as tuples (one row of a database table) of numbers.
Colour Space	A three-dimensional representation of colours that can be contained within a specific colour model.
Command Line	A computer systems main prompt. Commands can be typed in an MS-DOS prompt or in the Run dialog box in the Program Manager of Windows.
Component-Level Metadata	Metadata pertaining to the component classification of MAP(2).
Compression	A method of electronically reducing the number of bits required to store or transmit data. Lossy compression sacrifices data to achieve affect. Lossless compression allows the original data to be recreated as it was.

Coverage	A digital version of a map forming the basic unit of vector data storage in ARC/INFO. A coverage stores geographic features as primary features (such as arcs, nodes, polygons, and label points) and secondary features (such as tics, map extent, links, and annotation).
DAT Tapes	A digital magnetic tape format used for computer backup; the latest DAT storage format is DDS (digital data storage).
Data	Shorthand for a digital object.
Data Set	All the digital objects in a data collection (i.e. the 'archive').
Data-type	The characteristic of columns that defines what types of data values they can hold.
Delimiter	The "divider" character between separate fields in database / spreadsheet records. Indicates a field boundary within a text stream. The comma and tab are commonly used delimiters.
Deposit Format	Preferred or mandatory file formats for submission to the Museum of London.
Deposition Forms	Administrative forms supplied by the Museum of London at strategic intervals in the archive creation and deposition process.
Digital Object	A bit sequence that represents a meaningful unit of information; associated with a data structure that provides information on how to correctly interpret the bits.
Digital Preservation	The process of ensuring that a digital object is accessible over the long term.
Digital Repository	An organisation that has responsibility for the long-term curation and dissemination of digital resources.
Digital Surrogate	A digital copy of physical material.
Directory	A folder that contains a group of files and other directories.
Directory Tree	Hierarchical structure for organizing files and directories.
Documentation	Information about the provenance, contents and structure of a digital resource, and the terms and conditions that apply to its use. There are different forms of documentation ranging from unstructured contextual documentation to structured file-level metadata.
DVD-R	Optical disc capable of having data written onto it once. Much larger capacity than CD-R.
Embedded	A Digital object inserted into another as an integral part of a surrounding whole. E.g. an image file or table inserted into a text document.
Entity Relationship Diagram (ERD)	A diagram that represents the tables and the relationships between them in a database.
Export	An action that saves data in a format that another program can read.
Field	A column in a database or spreadsheet that contains a particular unit of information.
File Extension	The final part of a filename that denotes the file format; usually three or four letters separated from the file name by a period.
File Format	The specification of how the bits stored in a file should be interpreted.
File-Level Metadata	Metadata pertaining to individual files. Roughly equivalent to preservation metadata.
Fixity Information	See Checksums

Foreign Key	A column in a table that does not uniquely identify rows in that table, but is used as a link to matching columns in other tables.
Foreign Material	Externally produced material that has been incorporated into another digital object.
Formatting	1) Page-layout commands and font specifications for a document. 2) The process of initializing a new disk so that it can store information.
Formula	A rule that describes the relationship of two or more variables. An equation stating the rule.
Frame Rate	Frequency in which video frames are displayed on a monitor, typically described in frames-per-second (fps).
FTP	File Transfer Protocol. The language used for file transfer from machine to machine across the internet.
Functionality	A measure of the capability of an item of software or hardware.
Geo-Reference	The relationship between page coordinates on a planar map and real-world coordinates.
Information Content	The meaningful information stored as bits in a digital object and made understandable through the digital object's associated data structure.
Ingest	The process by which a digital file and documentation (i.e. metadata) is accepted and loaded into a digital store for subsequent retrieval.
INSCRIPTION	Wordlists maintained or recommended by the Forum on Information Standards in Heritage (FISH).
Interoperability	The ability of different types of computers, networks, operating systems and applications to work in unison without prior communication.
ISO	International Standards Organizations. Working with the United Nations the ISO do not create standards but provide a mechanism for validating a proposed standard.
Line (record) Breaks	A special character that forces a new line on the page without creating a new paragraph. In the context of format migration, unwanted line breaks cause the record concerned to be erroneously spread over a number of rows, resulting in data being placed in inappropriate fields.
Machine Readable	Information encoded in a form that can be read by a computer and interpreted by the computer's hardware / software.
Macro	A mini-program which will execute a series of commands in series.
MD5	A secure hashing function that converts an arbitrarily long data stream into a digest of fixed size.
MDA	The Museum Documentation Association
Metadata	Data that is used to describe other data. A structured form of documentation that exhibits consistency.
Moral Rights	Applicable to any work in which copyright subsists. The rights remain with the creator, irrespective of the ownership of copyright. Concerned with the right of attribution, the right to not have authorship falsely attributed and the right of integrity.
Native Format	Proprietary file format of a given application which is not intended to be opened by any other application.

Non-Proprietary	Software application that is not owned or controlled solely by one company or institution.
OS	1) Ordnance Survey; the national mapping agency of Great Britain. 2) Operating system; program(s) that manage the operations of a computer.
Package	A SIP sub-directory that used to contain a number of discrete files that need to be integrated for the effective recovery of a digital items data and intellectual content.
Preservation Metadata	Records the information required to manage the preservation of digital resources. Loosely equivalent to the file-level metadata discussed in these standards.
Primary Key	A column in a table whose values uniquely identify the rows in the table.
Project-Level Metadata	The metadata associated with the project as a whole. Loosely equivalent to resource discovery metadata.
Proprietary	Privately developed and owned technology.
Query	An object that requests information from a database and creates a dataset of the requested information.
Record	A single item stored in a database or spreadsheet. The equivalent of a single row.
Rectangular Object	A flat file containing data allocated to row and columns such as spreadsheets and database tables.
Rectification	The process by which an image or grid is converted from image coordinates to real-world coordinates.
Render(ed/ing)	To transform digital information in the form received from a repository into a display on a computer screen.
Resolution	The number of pixels per square inch on a computer-generated display.
Resource Discovery Metadata	The catalogue and index record component of a study. Roughly equivalent to project-level metadata in this instance.
Row	A single, implicitly structured data item in a table / spreadsheet (see record).
Save As	The option to save a file under a different name, format, or in a different directory.
Schema	A description of structure and rules.
Standards	Technical specifications that lay down characteristics of a product with a view to engendering interoperability.
Submission	The act of depositing data at the Museum of London.
Text Encoding	A method of representing text as binary values.
Text Qualifiers	Characters (usually “ or ') used to anchor the content in each field. These characters are used to indicate that everything in between them should be interpreted by software as it appears.
UNICODE	A standard for international character encoding. Supports characters that are 2 bytes wide, rather than the 1 byte, allowing it to include 65,536 characters rather than the 256 available to 1-byte systems.
Validation	A process for confirming the correctness or soundness of migrated data.
Variables	See Field
Version	A major edition of a software program / file format.

W3C	The World Wide Web Consortium, the governing body for web standards. (http://www.w3.org/).
White Space	Characters that serve as token separators; space, tab and new-line characters.
Worksheet	A work area comprised of rows and columns.

Appendix: Metadata document example

Note: this example only includes Level 1 metadata for the most frequently deposited data types. If you require assistance with metadata for other less common data types (Sections 2.7.8 – 2.7.13) or help with Level 2 or 3 metadata, please contact the LAARC Records Officer (Digital). Also, this example only includes master files (see Section 2.7.2). The inclusion of metadata for native files and any optional files is strongly encouraged, but not compulsory.

NON11

=====

Portable Document Format

=====

The five PDF/A-1a files were created in MS Word 2003 (v.11) using Adobe PDF Maker 10.0.

NON11_dba.pdf	19,968 bytes.
NON11_site_abstract.pdf	8,247 bytes
NON11_deposit_survival.pdf	3, 008 bytes
NON11_evaluation.pdf	235,520 bytes
ABC_Archaeology_CAD_layer_conventions.pdf	6,429 bytes

The following PDF/A-1b file was created in AutoCAD 2009 using Adobe PDF Maker 10.0

"NON11_site_location_plan.pdf" 294, 305 bytes

All PDF/A files conform to the PDF/A Standard (ISO 19005-1) and have been verified using Adobe Acrobat Professional V.10.

Spreadsheets/ Databases

=====

The four tab-delimited (UNICODE) files were created in MS Excel 2003 (v.11).

FILE_NAME	CONTENTS	FILE_SIZE (BYTES)	No_ROWS	No_COLUMN
NON11_find_inv.tab	Finds Inventory	3024	92	10
NON11_imge_reg.tab	Image register	2459	61	11
NON11_context_reg.tab	Context Register	9,520	96	9
NON11_biblio.tab	Bibliography	406	3	13

All files conform to the structure set out in the Museum of London Standards for Deposition (2009 revision) and established Museum of London coding schema, but 'NON11_context_reg.tab' also contains three additional fields:

Field 3 (Label "Length")=Length, defined in Metres or full extent of trench

Field 4 (Label "Width")=Width, defined in Metres or full extent of trench

Field 5 (Label "Depth")=Depth, defined in Metres or full extent of trench

Vector Graphics

=====

The one DXF file was created in AutoCAD 2009 (v. 17.2):

NON11_site_location_plan.dxf 291, 300 bytes

Caption: 1:200 Site location plan for Archaeology House, 123 Noname Street, EC2, City of London.

Drafting conventions: the plan was produced using the standard ABC Archaeology

layer-naming convention - see deposited documentation:
'ABC_Archaeology_CAD_layer_conventions.pdf'.

Raster Images and Digital Photographs

=====

The following 10 TIFF (v.6) files were created in *CorelDRAW* 2006 (version X3):

NON11_dba_fig1.tif/ eval_rpt_fig1.tif 3,438,642 bytes
Caption: Site location plan for NON11

NON11_dba_fig2.tif 935,192 bytes
Caption: Plan of proximate sites to NON11

NON11_dba_fig3.tif 8,147,768 bytes
Caption: Proposed development plan

NON11_dba_cover_fig.tif 2,985,320 bytes
Caption: Desk-based assessment report cover page

NON11_eval_rpt_fig2.tif 4,292,618 bytes
Caption: Detailed site and trench layout

NON11_eval_rpt_fig3.tif 2,754,321 bytes
Caption: Trench plans and sections

NON11_eval_rpt_cover_fig.tif 2,843,900 bytes
Caption: Evaluation report cover page

9 digital photographs were deposited as TIFF (v.6) files:

Date of capture: 22/02/1980

NON11_1.tif 2,437,481 bytes
Caption: ID shot

NON11_2.tif 3,987,030 bytes
Caption: General shot Trench 1: north view

NON11_3.tif 3,298,315 bytes
Caption: General shot Trench 1: south view

NON11_6.tif 2,985,482 bytes
Caption: Sample section Trench 1 - west facing

NON11_7.tif 3,319,753 bytes
Caption: Sample section Trench 1 - east facing

NON11_8.tif 3,674,259 bytes
Caption: General shot Trench 2: east view

NON11_8.tif 3,031,654 bytes
Caption: General shot Trench 2: west view

NON11_9.tif 2,902,783 bytes
Caption: Sample section Trench 2- north facing

NON11_10.tif 2,876,430 bytes
Caption: Sample section Trench 2 - south facing

Additional Level 2 metadata is embedded in all 10 digital images as follows:

- Camera Make/Model
- Colour Representation
- Resolution
- Bit Depth