Ingest Manual

Version 4.0

Created date:	2004
Last updated:	31 March 2018
Review Due:	August 2018 (unless significant change)
Authors:	Tim Evans and Ray Moore
Maintained by:	Digital Archivists
Previous version:	Ingest Manual 3.0

Introduction

The ingest process at the ADS is a two-stage process. First of all, the data is accessioned. This should be carried out as soon as possible after receipt of the *Submission Information Package* (SIP). The second stage of the process, preparing the data for archive and dissemination, is more time consuming and will be scheduled in to the ADS monthly work plan by the Collections Development Manager and the management team.

Accessioning the data

As a new SIP is received by the ADS it is a priority to get data stored safely on our systems and issue the depositor with a deposit receipt and email acknowledgement that the data has been received. Accessioning should be carried out at the earliest opportunity. At this stage the digital archivist will try to identify any potential problems with the data, so that where more documentation is required, or if files are corrupt, contain viruses or are inappropriate for deposition, the depositor or data producer can be informed as soon as possible. When required the depositor can be directed towards the *Guidelines for Depositors*[1], the list of *Preferred and Accepted File Formats*[2] and the Guidance *on the Selection of Material for Deposit and Archive*.[3]

An accession checklist is available to guide digital archivists through the accessioning process and should be used with reference to this *Ingest Manual*.[4] This checklist is replicated within the ADS' *Collection Management System* (CMS) and this must be completed to allow the archive to continue through the workflow. Only when all parts of the CMS checklist are completed can work begin on the creation of the *Archival Information Package* (AIP) and *Dissemination Information Package* (DIP).

[1] http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml
[2] http://archaeologydataservice.ac.uk/advice/FileFormatTable.xhtml
[3] http://archaeologydataservice.ac.uk/advice/selectionGuidance.xhtml
[4] http://archaeologydataservice.ac.uk/resources/attach/ADS_Accession_Checklist_2018.pdf

Since 2014 the ADS has also accepted the deposition of data through both ADS-easy[5] and OASIS Images[6] portals. These give depositors the ability to create metadata and upload data to an ADS holding area. In these instances the ingest process is largely the same, however the deposition of data online, is subject to a discrete accession process (incorporated below) and checklist.[7]

Data Transfer Session

Data is accepted in a variety of different file formats and can be delivered to the ADS in a number of ways.[8] Further information on what an archive should contain is available through the *Guidance on the Selection of Material for Deposit and Archive*[9] and discussion with the Collections Development Manager or Digital Archivists, with specific guidance on formats available in the list of preferred and accepted formats[10] and guidance on metadata requirement in the *Guidelines for Depositors*.[11]

Virus Check

Before transfer of the data to our servers, anti-virus software is run on all file within the SIP to ensure it is virus free. For deposits made by ADS-easy, virus checks are performed upon upload of the file(s) to that system, but an additional check is advised. If a virus is found, this must be removed either by disinfecting the relevant file, or requesting that the depositor provides a 'clean' copy. Further processing of the data files must not go ahead until the data is known to be free of viruses.

Manual checking will also highlight any files that are password protected. Depositors should be asked to re-submit copies of any files that are password protected.

Media and File Readability Check

If the SIP consists of only a small number of files these are opened to ensure they are readable and not corrupt. Unless batch processing is an option only a random sample of each file format present in a large archive will be checked at this stage. More thorough checks will be carried out in the data processing stage. The depositor should be contacted regarding any corrupt files and asked to re-submit.

Basic data validation may be carried out at the accessioning stage but it is recognised that it will not be possible to pick up on all potential issues with an SIP at this point. Fuller checks will be carried out at a later stage when the AIP and DIP are created. Checks will vary depending on the nature of the data and the file formats that have been submitted. Appendix 1 describes the sorts of checks that may be carried out here.

Check File Formats are Suitable for Deposit

The SIP must be checked to ensure that all files submitted are files that we can accept. Lists of suitable file formats are available in our *Guidelines for Depositors*[12] and should be adhered to by depositor unless special provisions have been agreed in advance.

---

[5] http://archaeologydataservice.ac.uk/easy/
[6] http://oasis.ac.uk/pages/wiki/Main
[7] http://archaeologydataservice.ac.uk/resources/attach/ADS_easy_Accession_Checklist_2018.pdf
[8] See the How to Deposit Data section of the ADS website - http://archaeologydataservice.ac.uk/deposit/How.xhtml
[9] http://archaeologydataservice.ac.uk/advice/selectionGuidance.xhtml
[10] http://archaeologydataservice.ac.uk/advice/FileFormatTable.xhtml
[11] http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml
[12] http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml, and specifically in the section on *Preferred and Accepted File Formats* - http://archaeologydataservice.ac.uk/advice/FileFormatTable.xhtml

Typically, problems will be mitigated during negotiations for deposition with the Collections Development Manager and/or Digital Archivists. However where a SIP contains file formats that the ADS are unable to deal with (for example, when no suitable migration path exists, or that we do not have access to any software that can read them), then the depositor must be contacted to see if they can provide the data in an alternative format. Where it is not possible to re-submit files in a suitable format, we may agree to accept this data, but archiving will be carried out on a 'best efforts' basis.

When an archive is deposited through ADS-easy these checks should be made during the upload process as depositors should only be allowed to upload certain formats specific to the data type. This is not the case with OASIS Images, however, here files and metadata are added to a zipped archive. In either case it is worth checking the file formats and reporting any problems to the depositor and ADS-easy Manager and the depositor.

Documentation and Integrity Check

The SIP should be checked against any accompanying documentation, such as file lists and metadata supplied by the depositor. The numbers of files, file names, formats and version should be checked here. The purpose of these initial checks is to highlight any missing files or major discrepancies between the delivered data and the documentation. Any discrepancies should be queried with the depositor.

All SIPs received by us should be accompanied by appropriate documentation and metadata. The *Guidelines for Depositors*[13] describe what documentation is required, and also contain links to metadata templates which can be downloaded and completed then sent to us to supply the level of metadata that is required.[14] Appropriate collection level documentation, which provides metadata on the deposit as a whole, should be completed.[15] At the same time each file requires discrete metadata appropriate to the type of data that the file contains.

Archives uploaded via the ADS-easy portal also require metadata, which can added manually through a series of online forms, or in some instances (raster images) uploaded using an appropriate spreadsheet template. Within ADS-easy additional metadata can also be uploaded through the online forms. In the case of OASIS Images[16], used specifically for uploading raster images associated with an OASIS record, the images and the completed file level metadata spreadsheet should be zipped together and uploaded through the OASIS system.[17] In this case the accompanying metadata will need to be checked for accuracy.

Requests for the submissions for any missing or additional metadata and documentation necessary for the processing and archiving of their data should be made directly to the depositor.

As we will be creating a web interface to allow access to the data, the SIP should also include some introductory text about the project and the digital archive which can be used on the front page/s of the archive. Where necessary a depositor can include additional metadata in the form of an overview. It is also useful to have one or two images to illustrate the archive where appropriate. The depositor can be contacted at this stage and asked to supply these if necessary.

---

[13] http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml
[14] For example, see the completed metadata form Text files - http://archaeologydataservice.ac.uk/resources/attach/ADS_text_metadata_example.pdf
[15] http://archaeologydataservice.ac.uk/advice/DatasetlevelMetadata.xhtml#Collection-level%20Metadata%20Requirements
[16] See Hardman, C (2014) OASIS: new image upload facility. OASIS Blog 22 Apr 2014 https://archaeologydataservice.ac.uk/blog/oasis/?p=41
[17] http://oasis.ac.uk/pages/wiki/Main

The digital archivist should highlight to the depositor any files within the SIP that appear not to belong to the depositor or be subject to their copyright (either the individual or organisation for which the depositor has signed the deposit licence). Similarly the digital archivist should check that any data meets current ethical guidelines and legal requirements as outlined in the *Policy and Guidance on the Deposition of Sensitive Digital Data*.[18] Should there be any unresolved copyright, ethical or legal concerns, the processing of the submission stops until relevant copyright clearances have been obtained. Any queries on the copyright of submitted material, or ethical and legal issues should go through the Collections Development Manager.

Record Details of SIP in Collection Management System

The date of receipt of data should be recorded in the CMS. This will allow the accessioning process to begin. All relevant details of the accession must be recorded in the CMS, with a completed data receipt (a list of all files deposited) and emails, alongside any further communications concerning the deposition, should be stored according to *Repository Operations* document.[19]

In the case of ADS-easy and OASIS Image submissions much of the technical and collection level metadata will be imported during the accession process.

Standardise File Names

Where necessary file and directories within the SIP should altered in accordance with the file naming policy outlined in the ADS internal wiki, but also documented in the *Repository Operations* document.[20]

For those archives submitted via ADS-easy and OASIS Images file names should be updated automatically, in line with the current guidance on file naming policy, at the point of upload or at ingest. However, it is worth checking file names to make sure that they match the current file naming policy.

Copy to Data Server

The SIP should be copied to our data server in the directory structure outlined in the within the *Repository Operations* document.[21]

In cases where data has been deposited using ADS-easy or OASIS Images data will automatically be copied to our data servers during the automated ingest process initiated in the CMS.

Create Checksums and Technical Metadata

Once files have been copied to the data server then file characterization software, DROID[22], can be run from within the CMS. This is used to populate the Object Management System (OMS) with technical metadata - physical location, filename, size, format, MIME type, PRONOM identifier and, most importantly, fixity value.[23]

---

[18] http://archaeologydataservice.ac.uk/advice/sensitiveDataPolicy.xhtml

[19] http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp

[20] http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp

[21] http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp

[22] http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/

[23] For further information see the *Preservation Policy* (http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#PresPol) and the *Repository Operations* document (http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp).

Issue and Store Licence in AIP Directory

A digital licence should be issued to the depositor through the CMS, although in some circumstances a physical licence can be printed out and sent to the depositor. It is important to make sure that the correct information (title and the details of the depositor information) are completed in the CMS prior to the issuing of a licence. If these are not known then the correct details should be requested from the depositor. A signed deposit licence should exist for every SIP.

N.B. *Work should not begin on creating an AIP until a signed deposit licence has been received.*

If a licence has been received by the ADS for the collection then this will be added to the collection record within the CMS. Typically, the digital licence will be returned in PDF form, but should be converted into a suitable preservation format (PDF/A) and stored according to the *Repository Operations* document. If a physical licence is returned then a scanned copy of it should be preserved according to the *Repository Operations* document. In either instance the licence should be stored in the /admin/ directory of the archive, but also added to the 'Negotiation' section of the CMS.[24]

Scan Paper Documentation

Most documentation arrives in a digital form, but on the occasion where only paper copies are provided and we are subsequently unable to get a digital version from the data producer, paper documentation will be scanned at an appropriate resolution.[25] This should be stored in the appropriate place within the AIP directory structure and where appropriate added to the 'Negotiations' sections of the CMS.

Acknowledge Receipt of Data

A receipt of the data received should be sent to the depositor. As outlined in the *Repository Operations* document a digital receipt should be generated from a subset of technical metadata produced by DROID (i.e. file name, location, file size and file type) stored in the OMS. This should be enclosed with the standardised email outlined in the ADS internal wiki. Any issues or queries concerning the deposition should be highlighted within the email to give the depositor an opportunity to address them. A copy of the receipt and the associated email should be stored in the '/admin/project_metadata/ directory of the AIP following the file naming strategy (i.e. email_deposit_yyyy_mm_dd.txt and deposit_receipt.xlsx).[26] Both the deposit receipt and email should also be attached to the CMS record (in the 'Negotiations' tab). Further details on this process are described in the *Repository Operations* document.[27]

Store Original Media

Once accessioning is complete, original media and associated paperwork are stored in the collections filing cabinets ordered by collection number. Both the media (CD/DVD etc.) and any accompanying paper documentation should be annotated with the collection number and the accession number as recorded in the CMS, plus the date data was received. Where the depositor has requested that original media is returned to them (this may happen if data has been delivered on a memory stick or portable hard drive) this should be returned to them

[24] http://archaeologydataservice.ac.uk/advice/RepositoryOperations.xhtml
[25] As described in the *Repository Operations* document - http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp
[26] In circumstances where there is more than one accession within the archive the deposit receipt should be superseded with the accession id (i.e.deposit_receipt-{accession_id}, e.g. deposit_receipt-1007337).
[27] http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp

only after the weekly backup to our deep storage facility has been carried out. In these cases a note should be made in the note section of the appropriate accession highlighting that the original media has not been retained.

In circumstances where data has been deposited electronically through ADS systems (ADS-easy or OASIS Images) or external file sharing services, these should not be deleted until the weekly backup to deep storage has been actioned.

Preparing the AIP and DIP for Archive and Dissemination

The process of normalising files and creating the AIP and DIP is carried out by a Digital Archivist and involves the steps described below.

Though a certain amount of checking and validation of the delivered data will have occurred at the accessioning stage, it is often the case that previously undetected issues may arise once we start preparing the AIP and DIP for archiving and dissemination.

An electronic checklist is available through the CMS and is intended to guide staff through the archiving process[28], and should be used with reference to this *Ingest Manual*. Only when all parts of the CMS checklist are completed can work on the AIP be signed off.

Consistency Checks

Basic checks of data against accompanying documentation should already have taken place as part of the accessioning process, but once work on the data begins in earnest, there may be further checks that need to be carried out.

The nature and complexity of consistency checks depends on the data type under scrutiny. Databases, GIS and spreadsheets would usually require more checking than Raster image files for example. Examples of the sorts of checks that may be carried out can be found in Appendix 1 of this manual.

Where problems with any of the data are highlighted, the depositor should be contacted, informed of the problem and encouraged to re-submit. In order to maintain the integrity and authenticity of the data, we would always prefer the depositor to re-submit data rather than carrying out edits in-house. Digital archivists should consult the *Data Procedures* guidelines when editing and updating archived data for further guidance.[29]

Selecting Preservation and Dissemination File Formats

With reference to the ADS Data Procedures[30] and the contents of the archive, as documented in CMS and OMS, a suitable preservation and dissemination pathway should be decided upon by the Digital Archive for each element. Consideration should be paid to preservation and dissemination formats for each file, with particular attention given to the original format version. In carrying out this step it is important to consider the significant properties of each of the files and ensure that these will be preserved within the new formats.

It may be that the formats and format version of the files were delivered in are suitable for preservation or dissemination with no changes. It may be that the same format is suitable for both preservation and dissemination, or it may be that two different versions of the file need to be created for the purposes of preservation and dissemination.

Develop a Conversion Plan

Once decisions have been made on appropriate file formats for preservation and dissemination, a conversion plan may need to be established. If the conversion into a

[28] A static version of this checklist is available - http://archaeologydataservice.ac.uk/resources/attach/ADS_Procedure_Checklist_2018.pdf
[29] These are available in the ADS internal wiki, although static versions of these procedures are available via the Preservation Policy, Appendix - http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#PresPol
[30] See above.

preservation or dissemination format is a complex procedure, map out each of the steps to be taken and check that we have the software and expertise to carry out this work. Some conversions may be straightforward single-step operations which can be done as batch processes on large numbers of files. Other conversions may involve multi-stage processes using various different software packages.

Convert the Files

Following instructions contained within the *Data Procedures* documents, copy and convert the deposited files as necessary until every file supplied is in the correct format and location. Files intended for preservation or dissemination should be stored in the correct place within the directory structure as defined in the *Repository Operations* document.[31]

Note that all work on files should be carried out on local copies of the data, not on the actual SIP that has been accessioned on to the data server.

Validate File Conversion

Ensure that the conversion of the digital resource has been successful and that the significant properties of each file remain unchanged. Though it is not practical to check each individual file, it is important to check a sample size set at the discretion of the archivist and relevant to the size and complexity of the overall dataset against the original SIP to ensure that significant properties are maintained.

Record AIP and DIP in OMS

The digital archivist should use the DROID-based interface within the CMS to record the technical details of all files created for the AIP and DIP in the OMS. Once technical metadata has been added to the OMS then the 'Match Objects' process (accessed through the CMS) can be used to link related files into notional 'objects'. Typically matches can be made programmatically using one of the automated 'computer matching' settings. Once complete these relationships should be checked as errors can occasionally occur. In circumstances where the automated process is not used or where it has been unable to 'match objects' then the digital archivist should use the interface within the CMS to manually relate files into objects. It is worth highlighting that in instances where deposit is particularly large[32] then the CMS interface will be unable to display the files included in the archive, consequently the 'match objects' process must be run outside the CMS, or matches may be added manually to the database.

As part of the file characterisation each file is also assigned a particular data type, typically based upon the file extension. While, by and large, the file extension proves an accurate reference to the data type, in some circumstances the automated system will have problems identifying the correct data type. With this in mind digital archivists must check that the correct data type has been assigned to the data where necessary changes can be made using the 'update object table' interface within the CMS. Particular attention should also be paid to those files containing collection level or file level metadata which can be deposited in a variety of formats, but which should be categorized as 'Documentation'. Similarly files which have an administrative function within the archive, and which typically are created in variety of formats, should be have an 'Admin' data type.[33]

---

[31] http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp

[32] This upper limit is currently 30000 files.

[33] A full list of the discrete data types and a general discussion of data types can be found in the Repository Operations, Appendix 5 http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp.

Where these objects have a relationship with other files and objects within the archive these should be documented in the relationships table within the OMS. At the same time each of these relationships should be given a 'relationship type' that reflects the nature of the relationship. This is particularly important where the archive includes metadata submitted in supplemental files or documentation. There is currently no interface that allows the creation or maintenance of the these relationships so additions and changes must be made directly into the OMS tables.[34]

Record Conversion and Editing Processes Undertaken

Once the conversion plan has been successfully implemented, the 'Processes' section of the CMS should be updated to document the details of these operations. In many circumstances this can achieved through the 'Generate Processes' function within the CMS, this automatically generates a list of actions carried out on files and the results of matching file-based representations. Again this automatic process makes assumptions about the nature of conversion, so when used, any processes should be checked for errors and where necessary updated or additional information added. In some circumstances it may be necessary to add or edit processes manually; these can be carried out through the interface within the CMS. Each process documents the nature of the conversation undertaken carried (type), the source and destination formats, with the start and completion dates. A full or shorthand list of inputted and outputted files is included. Information about the hardware, operating system and software used is added, alongside the person who carried out the process. Any problems or additional information about the process is documented in the comments section.

Interface creation

Once the preservation work has been completed digital archivists should create an appropriate archive interface for each dataset. Specific guidance on this is available from the ADS internal wiki and through consultation with the Collections Development Manager. A series of templates and exemplars are available to facilitate this process. Each interface should be subject the appropriate checks in terms of accessibility, validation and compatibility as outlined in the internal wiki and checklist.[35]

Add and/or Update Collection and File Level Metadata

All collections deposited with ADS should be accompanied with appropriate collection and file level metadata. In instances where the SIP is submitted digitally through ADS-easy the collection level metadata will be transferred to the CMS during accession. In instances where data is submitted on physical media, through OASIS Images, or transferred digitally using external file sharing platforms, any collection level metadata should be added manually to the CMS. Particular attention should be paid to recording those individuals or organisations responsible for the dataset, but should also include the addition of appropriate subject and coverage terminologies.[36] A list of the required collection metadata fields is outlined in the ADS template[37], within the ADS-easy interface[38] and is also documented in the ADS internal wiki.

---

[34] A full discussion of the ADS' implementation of PREMIS can be found in the Repository Operations, Appendix 6 http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp.

[35] A static version of this checklist is available - http://archaeologydataservice.ac.uk/resources/attach/ADS_Procedure_Checklist_2018.pdf

[36] Where appropriate these should use the thesauri and locational terminologies outlined in the Repository Operations documentation. It is worth highlighting that in instances where this data has been transferred from ADS-easy that the terms themselves are transferred but these will need to updated with the related thesauri terms.

[37] http://archaeologydataservice.ac.uk/advice/DatasetlevelMetadata.xhtml#Collection-level%20Metadata%20Requirements

Where the archive has been deposited using ADS-easy any data type specific metadata will be added to the OMS during the semi-automated accession into the CMS. In instances where data has been submitted within supplemental files, typically following the standardized templates available from the *Guidelines for Depositors*[39], this metadata should be added to appropriate tables within the OMS. For those archives which contain (raster) images, particularly those archives submitted through the OASIS Images system, a loader has been created which loads the data directly into the OMS, this is accessible through the CMS. For other data types metadata will need to be added directly to the database through a series of insert statements.

Submit AIP for Checking

Once preparation of an archive is complete then the digital archivist must decide whether the AIP should be submitted for checking by another Digital Archivist. Historically, **all** archives were submitted for checking, but as workload has increased it has become impossible to check all AIP's on completion, consequently it is left up to the digital archivists who worked on the archive to decide whether AIP should be submitted for checking. In many cases AIP's are relatively simple, for example, the archive may contain a small number of raster images and a spreadsheet of metadata, as a result it may not require additional checking. However, where the archive is large, particularly complex or where the digital archivist is unsure whether preservation and dissemination work has been completed correctly they may be submitted for AIP checking. If checks are required or warranted then a request should be sent to the Collections Development Manager through the checklists in the CMS. The AIP check can then be allocated to digital archivists.

That said, AIP checking is an important part of the archiving process ensuring that the recommended directory structure has been maintained, recommended preservation file formats have been used and required metadata and checksums are present. This double checking of an AIP ensures that all staff are working to a consistent standard and following the archival procedures as laid out in this *Ingest Manual*, the *Repository Operations* document[40] and *Data Procedures*.[41] Where AIP checking is warranted an AIP checklist, which provides a rough guidelines on the checks to be made, should be completed within the CMS.[42] This checklist should be used in conjunction with the 'AIP checking' documentation.[43]

Once AIP checking is carried out and any problems or issues with the AIP are highlighted and corrected by the digital archivist responsible for the initial work, the ingest process is then complete. The AIP will remain unchanged until such a time as further data is accessioned in to it, or preservation or dissemination files need to be migrated into newer or more suitable formats.

Interfaces which have been created during the archiving process should also be checked for problems and errors, by other ADS staff and the depositor.

N.B. An archive cannot be signed off and publicly released until the final AIP checks have been completed and signed off by the Collections Development Manager.

---

[38] http://archaeologydataservice.ac.uk/easy/
[39] http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml
[40] http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp
[41] These are available in the ADS internal wiki, although static versions of these procedures are available via the Preservation Policy, Appendix - http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#PresPol
[42] A static version of this checklist is available - http://archaeologydataservice.ac.uk/resources/attach/ADS_AIP_Checklist_2018.pdf
[43] Available in the ADS internal wiki.

Archive release

Once work on the AIP check has been completed and signed off by the Collections Development Manager the archive is ready for release. Any issues raised by colleagues and the depositor should similarly be addressed prior to release. The release process is documented in the ADS internal wiki, with a shorthand version available in the Procedures checklist.[44]

[44] A static version of this checklist is available - http://archaeologydataservice.ac.uk/resources/attach/ADS_Procedure_Checklist_2018.pdf

Appendix 1

Data Validation and Consistency Checks

Here are some examples of the types of checks which may be carried out as a part of the ingest process. This list is derived from the AHDS Archive Ingest Procedure Framework: HS Preservation Procedures Manual, working draft 1.3 prepared by Raivo Ruusalepp, Estonian Business Archives Ltd, December 2002/January 2003:

- Check that digital resources and their items adhere to the relevant formal definitions of their structure (e.g., an XML document conforms to its XML schema, a relational database conforms to its SQL schema, an image conforms to its stated image format – dpi, colour depth, compression, etc.).
- Image compression algorithm, dimensions, orientation, resolution, colour space, etc. correspond to the values stated in documentation.
- Digital audio compression algorithm, length of the recording, sampling frequency, bit rate, etc. correspond to the values stated in documentation.
- Digital video compression algorithm, length/duration of the recording, codec structure, frame rate, sound format, etc. correspond to the values stated in documentation.
- Linkages and dependencies between items within a particular type of digital resource should be checked for correctness (e.g., in a database, foreign keys having a matching primary key; in a spreadsheet, formulas refer to correct cells, etc.).
- Linkages and dependencies to other digital resources are correct (e.g., hyperlinks point to a currently valid URL, details of published works in a bibliography are correct, etc.).
- Items within a digital resource adhere to the relevant definition (e.g., a numeric field in a database contains a number, text strings do not exceed a stated maximum length, etc.).
- Items within a digital resource contain ‘sensible’ values that do not contradict relevant logical assumptions (e.g., age of a person should not be less than 0) and subject/resource type specific concerns.
- Documents (word processor files) should be checked for changes or errors in footnotes, tables of contents, links, auto-fields and formatting that may hinder the later use of the data resource.
- GIS, CAD and virtual reality data resources may require domain- or research area specific consistency checks to be applied (e.g., scale of different layers in a GIS, level of precision and sufficiency of coordinates in a CAD and VR data, etc.).
- Simple data types (numbers, text strings, dates, etc.) are not truncated, restricted in range, formatted or otherwise defined in a potentially confusing or ambiguous way (e.g., dates contain four digits for the century, date format, memo fields in a database do not contain embedded end-of-lines, etc.).
- Coded data must be checked that the data have been consistently assigned the documented code.

- Any codes that are used in data must be used consistently and according to the specified coding rules.
- Standardised data has been standardised consistently and according to specified rules or a recognised schema for the standardisation.
- Exceptions to particular standards, coding schemes, formats, etc. are documented and justified in the documentation for the data collection.