

ADS Cataloguing Policy

Jenny O'Brien and Tim Evans
19/09/2021

Document Control Grid

Title:	ADS Cataloguing Policy
File name:	ads_cataloguing_policy_v2-0
Location:	Internal Drive
Status:	LIVE
Version:	2.0
Last updated:	19/09/2021
Created date:	27/08/2020
Review due:	01/09/2022
Authors:	Jenny O'Brien and Tim Evans
Maintained by:	Tim Evans
Required Action:	Circulate

Contents

Glossary (use if required)	4
Introduction	5
Background	5
Catalogue Requirements	6
User Requirements	6
Depositors	7
National and Local Authorities	7
The Heritage Sector	8
Academic Researchers	9
External Partners	9
The Unknown User	12
International Standards	12
Dublin Core Metadata Initiative	13
ISAD(G)	13
Spectrum	14
PREMIS	15
GEMINI	15
Internal Requirements	15
ADS Strategic Planning	16
Planning and Management of ADS Projects	16
Outreach	17
Research and Development	18
Internet Archaeology	18
Catalogue Records: Minimum Standards	18
The Inventory	19
Identifying the Resource	20
Item count	20
Resource Name	20
Resource Description	21
Creator	21

Dates	22
Archive Level	22
The Catalogue	23
Recording Provenance	25
Project Agents	25
Project Scope (Location)	26
Project Scope (Temporal)	26
Project Type	27
Project Prompt	27
Records and Collections Management	27
Access terms (licence agreements)	27
Depositor name and contact details	28
Copyright holder	28
Update frequency	28
Lineage	28
Resource language	28
Release status	29
Collection Type	29
Describing the Resource	29
Object Level Resource Types	29
Subject Classification	30
Geographic Subject Classification	30
Biographic Subject Classification	30
Chronological Subject Classification	31
Preservation Management	31
Data Types	31
File Formats	32
File Size	32
File Name	32
File Location	33
Object Relationships	33
Content Enhancement	33
Future Plans	33

Content Enhancement	34
Bibliographic Records	34
Addressing the Backlog	34
Managing Versions	35
Digital Files	35
Metadata Records	35
Identifying research interests	36
ADS Catalogue: Essential Fields List	36

Glossary

ADS	Archaeology Data Service
AHDS	Arts and Humanities Data Service
ARIADNE	
CIDOC-CRM	
DC	
DPC	Digital Preservation Coalition
FAIR	
GEMINI	
IA	Internet Archaeology
MEDIN	Marine Environmental Data and Information Network
Metadata	Descriptive information about data
MIDAS Heritage	
PREMIS	

Introduction

This document aims to set out a [standard approach](#) in describing resources within the ADS Archive. As well as being a central point of reference for the types of information to be included in cataloguing a resource, this document will also identify the reasons behind those policies. This document will also identify any external cataloguing standards and terminologies that should be used, when they should be used (for which record or metadata information types), and will specify minimum inventory- and catalogue-level description requirements for each resource type.

1. Background

The ADS is a CoreTrustSeal accredited data centre which has almost 25 years experience in digital preservation. Over this period the ADS has archived over (number) digital objects, which have all been described through depositor-created metadata which, though originally based on Dublin Core (DCMES) and the early requirements set by the now defunct AHDS, has developed organically as new requirements have been added. Although successful in terms of maintaining a certain (internally focussed) baseline level of metadata, a lack of clear Policy and Responsibility for Standards since the end of the AHDS, and a reliance on external engagement with metadata, has resulted in records of varying degrees of accuracy and quality.

There are clear advantages of having the content descriptions created by the Depositor; they are nominally the experts in their own data, and, most importantly, they understand the context in which it was created, and context is the ruling principle of all archaeological work. There is, however, a shift that has been taking place over the last few years, in which there has been increasing emphasis on data and metadata reuse and user engagement. Having built stable and effective digital [Preservation Procedures](#) and [Repository Operations](#), the ADS is now at the stage where we need to review the information we hold, and to think strategically about how and why such information is presented and how the contribution and impact of our collections can be improved. Informed by our recent responses to making our [\(meta\)data FAIR](#) This is both for the machine/technical approach through partnerships such as MEDIN and ARIADNE

where the accuracy of our metadata is reflected through enhanced Discoverability, but also for the users of our own website who represent a diverse community, and who in the modern-age expect an engaging and intuitive experience.

Having a catalogue which covers these needs, and with the flexibility to be both an interoperable baseline, and a tool for presenting data in new and interesting ways, thus engaging our Users, is now a Strategic Priority. It is also an area of Responsibility that is defined within our area of work, with a renewed emphasis on management of the catalogue to ensure accuracy, consistency, and to view our metadata as an active and dynamic resource.

2. Catalogue Requirements

This section aims to describe the various requirements placed on the ADS Collections Catalogue. The policies surrounding minimum standards for ADS resources will be based upon these requirements.

This section will be subject to regular review to ensure the function of the ADS Collections Catalogue remains true to user requirements and the policies based upon them will be updated accordingly to ensure that the catalogue remains fit for purpose.

This section will list the different forms of requirements that the catalogue needs to take into account. For a list of specific fields, or units of information, please refer to section 7.

2.1. User Requirements

This section will define the general requirements placed upon the catalogue by ADS users. The following users are included in these considerations:

- Depositors
- National and Local Authorities
- The Heritage Sector
- Academic researchers

- External Partners
- The 'unknown user' (of the website)

2.1.1. Depositors

The ADS depositors form a subset of the Heritage Sector. The ADS 'natural' Collections are collections of resources generally formed as a result of a distinct research project, commercial investigation, or publication on some aspect of the historic environment.

Depositors from commercial organisations may need to be able to find their work in order to prove they have fulfilled contracts. The ability to track their data after deposition with the ADS therefore requires project information and depositor information to be recorded and made available via the ADS Catalogue.

In order to carry out their investigations in the first place, a depositor may be required to undertake a desk-based assessment to investigate the results of any related previous investigations that may inform future work. In these cases the depositors become the users, potentially of their own data, so will need to find relevant resources effectively in the same way as other Heritage Sector users (see 3.1.3 below).

2.1.2. National and Local Authorities

This user includes Local Authority Archaeologists that require deposition as part of the Planning process, and also National bodies such as Historic England or Historic Environment Scotland who may request deposition with ADS under funding agreements including Scheduled Monument Consent or Listed Building Consent. In England such organisations use MIDAS Heritage (itself designed to complement SPECTRUM, GEMINI and the CIDOC-CRM). MIDAS is a common framework for Heritage Information Management developed by The Forum on Information Standards in Heritage (FISH) and published by Historic England. Where appropriate, the ADS Standard will incorporate or map to Level 1 Mandatory MIDAS elements. With this in mind, such organisations would expect at least a basic compliance in the ADS catalogue, and to identify projects based on:

- Investigative activity: the type of project undertaken.

- Map location: every project relates to somewhere, even if a research project it deals with empirical data or conceptual data at a national or continental level.
- Designation + Protection: if a project relates to a site or locale with a level of national protection (e.g. Scheduled Monument, Designated Wreck, World Heritage Site). This would include Name, Type, and Unique Identifier.
- Heritage Management Asset: for example the name of the area covered by the research project, a characterisation area, or type of landscape.
- Find and identify a project based on its association with an external identifier, such as a Monument Identifier (HER, Church of England HER, or Scheduled Monument).
- Archive + Bibliography: details of the associated physical archive and bibliography of works.
- Research + Analysis: the significance of the project as defined by a National or Regional Research Framework theme. At the time of writing the Research Frameworks overseen by Historic England and Historic Environment Scotland are being revisited. Space to record these within the ADS Catalogue to facilitate a degree of interoperability is paramount.
- Actor and role: to understand who undertook the project and in what capacity.
- Find and identify a project based on the relevant National inventory, e.g. [National Heritage List for England \(NHLE\)](#) and [Canmore](#).
- Date of the project activity (e.g. dates of study or the dates of fieldwork).
- Local Authority Archaeologists also need to find and identify a project based on its geographic location, normally based on Ordnance Survey Local Authority/District boundaries (e.g. City of York; County Durham).
- If a project relates to a Historic Event (e.g. Battle of Hastings).

2.1.3. The Heritage Sector

The 'Heritage Sector' forms the main user base of the ADS Collections. This large group encompasses Commercial Archaeologists, Geophysicists and earth scientists, Built Heritage Professionals, Maritime Archaeologists, Museum-based Archaeologists, Local Community Groups, Independent Researchers and

Undergraduate/Postgraduate Students. Typically this group would want to understand the relevance of the Collection to their thematic interest, primarily the type of project (Event/Activity), where it relates to (location of the project), and the subject (what was found or studied).

2.1.4. Academic Researchers

Academic Researchers encompass a range of users from the Higher Education sector. Commonly this is Lecturers, Postdoctoral Research Assistants, or Postgraduate Students undertaking a particular thematic study based on a research question relating to a monument form (e.g. settlement), traditional period range (Bronze Age) and a region (Southwest England). In recent years requirements have moved towards understanding of particular topics such as “rural settlement” or concepts such as “trade”, “warfare” that rely on a level of conceptual tagging beyond the monument and object thesauri historically used. In addition, there is also a type of Academic user that needs to understand the type of data within an archive, for example stratigraphic or structured data for addressing queries such as “stratigraphy from urban excavations.

2.1.5. External Partners

A growing number of [external services consume our metadata](#). For Objects this is primarily for documents such as unpublished reports and journal articles.

ARIADNE: The [ARIADNE Portal](#) brings together archaeological research datasets from ARIADNE partners from Europe and beyond. The ADS supplies all Archsearch Inventory datasets, all records of unpublished reports, and all records for Archives to ARIADNE. Data is mapped to the ARIADNE AO-Cat, an application profile of the CIDOC-CRM, and exported as XML. The AO-Cat requires as mandatory:

- Resource Language
- Resource Type (a working project vocabulary mapped to the Getty AAT e.g. Fieldwork Archive)
- Period Term using UK controlled vocabulary or a date range.

- Location: a single coordinate (centroid is acceptable) as either OSGB or decimal LL.

[MEDIN](#): As part of the federated Historic Environment DAC, ADS supplies metadata for maritime archives for inclusion in the MEDIN Data Portal. The ADS classification for 'maritime' is where the location of the project is, or the subject matter of the project describes:

- Archaeological sites or finds within the ocean;
- The coastline, including intertidal zone;
- Estuaries;
- Coastal settlements including ports.

The MEDIN Discovery Metadata Standard is a marine profile of the UK government Standard GEMINI2 and also complies with other international conventions such as INSPIRE and ISO19115. The [MEDIN Metadata standard](#) thus requires the following mandatory elements:

- Resource title.
- Resource abstract.
- Resource type: one occurrence based on ISO Scope list (dataset, series, service).
- Unique resource identifier.
- Resource Language: using controlled vocabulary, ISO 639-2.
- Topic category: has to use a term from ISO19115 Topic Categories. For ADS resources this should always be ['Society'](#).
- Keywords: have to use
 - ◆ An INSPIRE Keyword from the list of [INSPIRE Spatial Data themes](#). In-line with the [procedure of other Heritage agencies](#) this should be 'Protected sites'. As a constant value within exported metadata, this does not need to be part of the ADS cataloguing procedure.
 - ◆ A MEDIN keyword based on the [SeadataNet Parameter Discovery Vocabulary](#). This should always be 'Marine archaeology'. As a constant value within exported metadata, this does not need to be part of the ADS cataloguing procedure.
- Spatial Bounding Box: in decimal LL (WGS84).

- Spatial reference system: important to note that this is as used in the Resource data (not metadata); derived from the [EPSG register of geodetic parameters](#).
- Temporal reference: The date of publication (i.e. the date at which the resource was made publicly available). This can be expressed as any form of date (i.e. YYYY or DD-MM-YYYY)
- Lineage: can use default text e.g. “This dataset was collected by XXXX and provided to the ADS for long term archive and management.”
- Limitations on public access: derived from the ISO 19115/TC 211 Geographic Information/Geomatics Metadata Standard. All ADS archives should be ‘otherRestrictions’ with sub-property ‘No restrictions to public access’.
- Conditions applying for access and use: a form of words which explains any constraints on the resource. ADS Archives can have a generic form of words, for example “Data is made freely available under a xxx licence”.
- Data format: [Controlled vocabulary](#).
- Frequency of update: [ISO frequency of update codelist](#).
- Conformity: a mandatory requirement for recording adherence to specific INSPIRE data specifications or MEDIN data guidelines. The ADS do not follow MEDIN specific guidelines for data, and thus this is not applicable. A suitable entry for ADS under adherence to INSPIRE, would be its use of locational terms for [Sea regions](#). This can be an automatically generated statement and does not need to be part of the ADS cataloguing procedure.
- Spatial representation type: using a controlled vocabulary, subset of ISO 19115 and does not need to be part of the ADS cataloguing procedure.
- Metadata specific entries which can be automatically generated and do not need to be part of ADS cataloguing procedure:
 - ◆ Metadata language: Controlled vocabulary, ISO 639-2.
 - ◆ Metadata standard name and version

[NERC](#): The NERC Data Catalogue Service provides access to all NERC funded datasets. Via the OAI-PMH Archives target, ADS shares metadata of all NERC project datasets it holds. The NERC data model specifically requires:

- [A NERC Topic](#): the recommended NERC vocabularies include ISO19115 Topic Categories. For ADS resources this should always be '[Society](#)'. This can be generated automatically in the CMS/exports and does not need to be part of the ADS cataloguing procedure.
- An INSPIRE Keyword from the list of [INSPIRE Spatial Data themes](#). In-line with the [procedure of other Heritage agencies](#) this should be 'Protected sites'. As a constant value within exported metadata, this does not need to be part of the ADS cataloguing procedure.

[DataCite](#). Metadata is supplied in the [DataCite Metadata format](#), via the DataCite API. The DataCite Metadata Schema is a list of core metadata properties chosen for an accurate and consistent identification of a resource for citation and retrieval purposes, along with recommended use instructions. Mandatory metadata includes:

- Identifier (with mandatory type sub-property)
- Creator
- Title (with optional type sub-properties)
- Publisher
- Publication Year
- ResourceType (with mandatory general type description subproperty)

2.1.6. The Unknown User

The ADS Collections will also aim to be discoverable and engaging to a general audience who may be browsing for resources. The ADS holds several resources relating to heritage work and discoveries of national importance, so the catalogue needs to ensure that the hierarchy, classifications, and discovery terms used are accessible to a broad audience and are capable of showcasing resources that may be of a wider interest than those more suited to certain groups, such as those listed above.

This will be achieved through the use of standards and vocabularies either with a broader scope or use than just heritage-specific standards, or, conversely,

through identifying data standards and vocabularies specific to certain industries that may have use for ADS resources.

In addition, through effective use of tags, hashtags, engagement with popular culture, and social media trend-watching and recording, the catalogue will facilitate conversations with user groups previously unconnected to the ADS Archive.

2.2. International Standards

2.2.1. Dublin Core Metadata Initiative

As stated above, the ADS Metadata Schema was originally built to Qualified Dublin Core standards.

Dublin Core is “A basic, domain-agnostic standard which can be easily understood and implemented, and as such is one of the best known and most widely used metadata standards.”

Along with other international standards, [DCMI metadata terms](#) will form the spine of the ADS Metadata Standard.

2.2.2. ISAD(G)

The [International Standard Archival Description \(General\) or ISAD \(G\)](#) is “an international standard which provides guidelines for creating the content of an archival description.”

As a standard aimed at the Archive Sector, and not necessarily created for the digital world, it is useful to incorporate into the ADS Standard to ensure the ADS Catalogue contains the key descriptive elements that would be familiar to traditional archival practice as well as digital data standards.

[“ISAD \(G\) has six mandatory fields at the collection level.](#)

- 3.1.1 Reference code
- 3.1.3 Title
- 3.2.1 Name of Creator
- 3.1.3 Dates of Creation
- 3.1.5 Extent of the Unit of Description
- 3.1.4 Level of description"

2.2.3. Spectrum

Spectrum was created for the museum sector as a Collections Management Standard and [suggests procedures and standards](#) that museums should aim to follow when cataloguing and managing their collections.

As the ADS is closely tied to the museum sector and often shares the results of heritage investigations (the ADS preserving and disseminating the digital outputs, and the Museums preserving and disseminating the physical outputs), the ADS Catalogue will look to Spectrum to inform its standards and procedures as well.

Spectrum shares many elements of Dublin Core and ISAD(G), but is particularly relevant to the ADS due to its object and collection focus, and the fact that items stored in a museum correlate so closely to the representations of those items and the events that produced them as stored at the ADS.

Spectrum's key inventory elements are:

Object number, Object name, Number of objects (if a group), Brief description (or image), Current location, Current owner (if not your museum), Recorder and date

Spectrum also includes elements based on the retrieval of objects from the field. As the ADS objects are also often the result of projects in 'the field', the project metadata submitted by depositors can also be mapped to these elements:

Field collection date (use a standard format).
Field collection event name (use a standard term source).
Field collection method (use a standard term source).
Field collection number.
Field collection place (use a standard form of name).
Field collector (use a standard form of name).
Geological complex name (use a standard term source).
Habitat (use a standard term source).
Stratigraphic unit name (use a standard term source).”

2.2.4. PREMIS

“The PREMIS Data Dictionary is a comprehensive, practical resource for implementing preservation metadata in digital preservation systems”

PREMIS forms the primary standard that the ADS adheres to in documenting the digital preservation process. PREMIS elements include: Objects, Events, Agents, and Rights, as well as file information such as Formats, Fixity, Integrity, and Environment.

2.2.5. GEMINI

“[GEMINI](#)” is the UK geographic metadata standard. It provides guidance on how to publish geographic metadata in a way that conforms to relevant ISO standards and the UK’s implementation of the European INSPIRE regulations.

The mandatory [metadata elements](#) are used in the MEDIN Metadata standard, and therefore a de-facto requirement of the ADS catalogue.

2.3. Internal Requirements

The ADS Catalogue, as a management and administrative tool, will also be used for the following purposes by the ADS:

- Planning
- Management

- Outreach
- Research and Development
- Internet Archaeology

It is therefore essential that the Catalogue contains enough information to facilitate the needs of the teams responsible for setting priorities, managing records, social media campaigns, and submitting funding applications, ensuring that they can find the records, collections, and statistics they need for their tasks.

2.3.1. ADS Strategic Planning

ADS Management Team (Deputy Director and Collections Development Manager) need the Catalogue to inform our understanding of the longer term trends in deposition, particularly within commercial work:

- Understand current trends in Deposition by project type, creator, creator type (e.g. Academic or Community Group) funder type, local authority, date deposited, date of work; to begin using this data to model the typical annual trends of ingest from commercial work and understand its impact on staffing requirements.
- To use details such as creator type to inform Collection Development priorities and campaigns, and use any identified lacunae to inform partnership building with relevant sector bodies and champions.
- Understand why an Archive was deposited e.g. under instruction from Local Authority, National Body, Funding Body, or Community Research, and in order to better understand the distribution of archive work between projects classified as 'External Consultancy', and further inform the development of SLAs with partners such as AHRC, MEDIN and NERC.
- Highlight specific Archive types that require particular workflows for metadata sharing and reporting to project partners: e.g. all Maritime projects (MEDIN), all Natural Environment Projects (NERC), all Journal collections (Keepers Registry).

2.3.2. Planning and Management of ADS Projects

ADS requires the Catalogue to help identify, prioritise and plan internal development projects and/or external partnerships, that cover but are not limited to:

- Data Migrations - to be able to quantify data and file types, so to assess the impact of file migrations for purposes of staffing, resource, and external provision of data.
- Data Sharing - to accurately and consistently understand which Archives need to be provided to external partners (see above), lacunae, and where appropriate plan and schedule projects to update metadata as external requirements/standards evolve. Also to be able to link to the equivalent record where held in another portal or aggregator.
- Update frequency - to understand which collections are due to be updated, so as to plan work across a year. In addition, knowing at a record level which collections are due to be updated is needed to build better tools for checksum maintenance, and automated backups and off-site storage.
- Technical delivery - to quantify and assess the different methods of technical delivery of each archive, for example use of web templates, use of WMS/Oracle.

2.3.3. Outreach

ADS requires the Catalogue to help identify archives which will help promote re-use of resources by all parts of the user communities:

- Allow staff to find an Archive or Object based on a specific monument or artefact type.
- To allow staff to find an Archive or Object that relates to a particular historical event (e.g. Battle of Trafalgar)
- To allow staff to find an Archive or Object that relates to a particular seasonal event, celebration or activity (e.g. Christmas, Diwali, Passover, Beltane, FA Cup final).

- To allow staff to find an Archive or Object that relates to a named individual as subject (e.g. Leonard Wooley, Charles Darwin, Mary Wollstonecraft)
- To allow staff to find an Archive or Object based around broad themes used by GLAM partners (e.g. Witchcraft, Folklore, Immigration/migration).
- To identify the unusual! For example archives that contain, or objects that depict, amusing/light hearted pictures or quotes.
- To identify archives or objects that contain or depict maps.
- To identify archives or objects that contain or depict sensitive material (e.g. human remains) that should not be used for social media, or else should follow specific guidelines for display.
- To identify archives which relate to specific History Key stages used in UK education.

2.3.4. Research and Development

Research and Development needs are primarily based on understanding the research value of an archive, and how the data within could be used. This goes beyond just knowing the data type, and with a focus on human assessment of the potential for data for (primarily), machine based operations or studies. For example:

- A general re-use value based on an assessment of the types of data within.
- To indicate the presence of 2d archaeological drawings that could be digitised (including finds drawings especially pottery)
- To indicate the presence of text documents that could be used for NLP or topic modelling - this would not include documents such as project designs, context sheets and thus require a level of human intervention to identify 'useful' content.
- To indicate a project dataset that contains LOD terminologies.
- To differentiate types of structured data (databases), for example if a stratigraphic database, finds inventory, mixture etc.

2.3.5. Internet Archaeology

Internet Archaeology articles are treated as collections in the ADS CMS, albeit collections that consist of HTML articles, images and on occasion additional data such as 3d models. The Editor needs to be able to query Collections based on the kind of object type, in order to manage content within more technical interfaces.

3. Catalogue Records: Minimum Standards

The catalogue will comprise inventory records enhanced with further metadata describing the provenance and subject matter of each resource.

3.1. The Inventory

Traditionally, the ADS metadata records have been built on an implementation of Dublin Core (DCMES) with DCMI qualifiers used selectively. This policy is the result of an ongoing metadata review and, having reviewed various description standards, the ADS identified 7 fields that are required to make a basic inventory record applicable to all ADS objects and resources.

The following fields have been selected to ensure that the inventory provides an effective 'at-a-glance' indication of what each resource contains. New Collections and Collection Updates should not be released for dissemination online unless the minimum inventory standards outlined here have been met. The fields below fulfil the requirements of the General International Standard Archival Description (ISAD [G]), as well as the essential fields required by the National Archives Discovery catalogue and, also complies to SPECTRUM standards. The inventory record for every resource (whether object, or collection) will contain the following:

- A Unique Identifier
- The number of items that comprise the resource

- A name
- A brief description of the resource
- Agencies involved in the creation of the resource
- Temporal Information (Covering dates or start date and end date)
- Level of Description (Series, Collection, Fond, Object etc.)

This information is stored in the ADS Collections Management System and Object Management System databases for collections and individual objects respectively. From these two systems, the ADS can identify and locate every digital object within our care, including those we create ourselves.

Where inventory information is missing, or insufficient, the creation of this metadata will be categorized as 'High Priority' for Object-level resources, and 'Extremely High Priority' for Collection-level resources for Archivist and Cataloguing project planning.

3.1.1. Identifying the Resource

(DCMI: Identifier "An unambiguous reference to the resource within a given context.", ISAD(G):Reference code(s), Spectrum:Object number)

The ADS Management Systems (CMS and OMS) records each resource according to a unique identifier: a Collection ID, for groups of objects such as series, deposited collections, and curated collections, and an Object ID for the smallest level of resource, a single report, a shapefile etc. These identifiers link the resources to all of their associated metadata, whether that metadata be stored separately as digital files, or as a record within the management systems.

For access and citation purposes, each resource will also be given a persistent identifier in the form of a DOI.

3.1.2. Item count

(ISAD(G):3.1.5 Extent of the Unit of Description, Spectrum:Number of objects)

The number of objects within a collection may be counted via a query to the database. The number of objects available for download is displayed for each collection on the 'Metadata' page as part of the user interface, this is divided according to data type.

The number of individual files that form any given representation of an object, for example, the number of files that comprise a shapefiles, which may be made available as a ZIP file, is recorded by the association of all of the various representations of a object with a single object ID, so the original number of files within a disseminated ZIP file can be counted using the Object Management System, as long as those relationships are recorded.

3.1.3. Resource Name

(DCMI: Title "A name given to the resource." ISAD(G):3.1.3 Title, Spectrum:Title)

The name of each resource should either be a formal title or a supplied title describing the resource content in a consistent and concise manner.

From the title of any given collection, the user should be able to surmise the likely general origin and subject of the collection as well as the likely resources that may be available within that collection.

For example: 'Images from a Building Survey of Home Farm, Leicestershire 2020'

Where a collection comprises multiple data types and/or events, the title should give a broad indication of this.

For example, 'Site and Post-Excavation data from multiple fieldwork investigations at Home Farm, Leicestershire 2019-2020'.

3.1.4. Resource Description

(DCMI:Description "An account of the resource.", ISAD(G) 3.3.1 Scope and Content)

The Description of the resource should expand upon the title, giving further information aimed at the user and information manager.

For formal documents, this should take the form of an abstract. For collections and objects, this should take the form of a content and scope note.

For example, for a site photograph, this might be in the form of a descriptive caption such as: 'North-facing shot of ditch [123] post-excavation with scale' - the type of information generally found in site photo records.

3.1.5. Creator

(DCMI:Creator "An entity primarily responsible for making the resource.", ISAD(G): 3.2.1 Name of Creator, Spectrum: Object production organisation, Object production person)

Where possible, the creator of any given resource should be recorded. This may be at an individual level, or an organisational level. For collections, for example, fieldwork archives, where an organisation is responsible for oversight of data created by multiple individuals over the course of a project, the organisation should be given as creator. Where individual authors, or researchers need to be credited, an ORCID ID must be used to uniquely identify the individual.

For collection-level resources, the creator would be the creators of the archive collection as a whole, where the depositing organisation has inherited a collection from another organisation, and prepares an archive for deposit, the new organisation would be the joint creators of the archive along with the original organisation who undertook the project.

3.1.6. Dates

(DCMI:Date Created "Date of creation of the resource", ISAD(G): 3.1.3 Dates of Creation, Spectrum:Date - earliest/single and Date - latest)

Every resource should have a creation date. This is particularly important with heritage data as interpretations change as research advances, so researchers need to know when a resource was created in order to assess if it has been superseded, or to evaluate a resource's merits against preceding or subsequent research.

For formal documents, the creation date is generally equivalent to the publication date. A resource may be created over a period of time such as a journal run, or a research database compiled over several seasons or years of work. In these instances both the earliest date (the date the database was first created, for example), and the latest date (the final entry was completed) should be recorded. This should be recorded as years (YYYY format).

3.1.7. Archive Level

(ISAD(G):3.1.4 Level of description, Spectrum: Record type).

For the ADS purposes the Archive will be arranged into the following levels:

- Fonds (a broad 'parent collection' of resources pertaining to a single organisation);
- Series (generally a broader term than 'collection' this refers to collection-level resources that have been 'accumulated and used together for a specific purpose, during a distinct period of time, and the records in a series are usually arranged in a particular order.' For bibliographic resources, such as journals, a series would be a run of journal volumes from a single source. A series may also reflect a common activity or type of record such as 'Infrastructure Projects', or 'Academic Research Collections');
- Sub-series (a sub-division of a series to group resources by a particular timespan, theme, or name for example, 'Historic England Research Reports Series' may be a sub-series of the broader 'Historic England Publications Series'. Narrower groupings of Collection, or 'File' level resources may be classified as 'Sub-sub-series' or Sub-sub-sub-series if necessary);

- Volume (Applicable only to bibliographic publication Collections, comprise a group of articles, for example, in a journal run. Volume is usually associated with a year of publication. A volume, as a collection of objects, equates to a 'File' level of description.);
- Collections (curated 'artificial' or deposited 'natural' collections of volume- or object-level resources surrounding a common theme or deriving from a common project or activity. ADS Collections, where non-bibliographic, generally equate to the 'File' level of archive description.);
- Objects (Individual resources, the smallest level of resource recorded. For example, an individual article, a digital image, or a spreadsheet. ADS Objects equate to the 'Item' level of archive description).

The level of description is recognisable by the form of identifier assigned to the resource and by its relationships to the other levels of description. The ADS intends to record these levels more explicitly as a result of a forthcoming project to review and align bibliographic metadata to standards and to the existing OMS.

3.2. The Catalogue

The Catalogue records build upon the basic information held in the Inventory records and each record, at its most basic, is required to describe:

- Information about the Provenance, including:
 - ◆ Who created it: the agencies responsible for the project and the project archive data;
 - ◆ Where was it created: where did the project that resulted in the resource take place;
 - ◆ The type of project
- Administrative and Collection Management information, including:
 - ◆ Access terms (licence agreements)
 - ◆ Depositor name and contact details
 - ◆ Copyright holder
 - ◆ Update frequency
 - ◆ Lineage

- ◆ Resource language
- ◆ Metadata language
- ◆ Release status
- ◆ Collection Type
- Descriptive Information about the resource itself, including
 - ◆ Type of resource
 - ◆ Project prompt
 - ◆ Subject Classification
 - ◆ Location-as-concept
 - ◆ Person-as-concept
 - ◆ Chronological Period-as-concept (with dates where specified)
- Preservation and Digital File Management information, including:
 - ◆ Data type classification;
 - ◆ Preservation Actions;
 - ◆ File formats (extension, version + mime type);
 - ◆ Checksum (actual checksum, checksum type e.g SHA-1, date of checksum creation)
 - ◆ File size;
 - ◆ File Names;
 - ◆ File Location;
 - ◆ Object relationships.

The ADS requires the depositors (ideally the Creators) to provide the information about the provenance of the resources, as well as basic descriptive information that gives the archivists and future users an indication as to the content and subject matter of the resources they are depositing. This ensures that the context is represented accurately, which will facilitate any subsequent enhancement by the archivists.

The bespoke ADS ingest systems check for the presence of all mandatory fields required to form a minimal catalogue record (which includes the essential inventory information). Quality assurance will be undertaken by the Digital Archivists prior to ingest, according to current procedures. Where metadata has not been submitted, or does not meet minimum standards in terms of content,

the submission will be declined with comments to the depositor as to what is required before the archive is accessioned into the ADS Archive.

Enhanced descriptions, tags, administrative information, and any mapping to ontologies and controlled vocabularies will be undertaken by the ADS Digital Archivists, following the Cataloguing and Records Management Procedures (forthcoming).

3.2.1. Recording Provenance

Project Agents

(ADS:ProjectAgent)

Provenance and trust have been highlighted as important elements in how researchers assess a resource's reuse potential or significance. The agencies involved in a project may be important factors in weighing the value of a resource, so recording this information in the ADS Catalogue is essential. It is also important to show the journey of the data from the point of project planning - what decisions were made that affected the data's creation, who made those decisions, and who can be contacted for further information.

As a minimum, the catalogue should contain the names of people or organisations responsible for the project, the funder type using the [OASIS Funder Type vocabulary](#) (i.e. not the funder name), and a point of contact for users of the data, with electronic contact details only. The depositors should provide this information, primarily via the OASIS heritage event recording system.

Project Scope (Location)

(ADS:ProjectLocation)

Heritage data tends to focus on place, with data deriving from a specific project undertaken at a specific location, or locations. As such the project location is of greater significance to the ADS Catalogue than the location of the creation of the data itself, and is distinct from the location-as-concept subject classification described below. At present, this field does not map to an established standard.

The catalogue will contain the location of the project from which the data derives, including the following levels: country, region, county, district/council, and site name. The depositors should provide this information, primarily via the OASIS event recording system.

The scope of the project will be defined via a bounding box or polygon which is a requirement of GEMINI.

The location of the project will be enhanced, by the Archivists, by mapping the most specific location area possible to the [Getty Thesaurus of Geographic Names](#).

Project Scope (Temporal)

(ADS:ProjectPeriod)

The scope of the project in terms of the historic, or prehistoric, time periods covered by the investigation or research. This may also include time period terms where finds or features from heritage investigations have been identified as belonging to a certain time period, so are incorporated into the interpretation of the chronological origin of the site, or research area. The chronological periods in these cases will be linked to the identified monument or artefact types.

This information will be recorded by the depositor, primarily via the OASIS heritage activity recording system.

As this element is intended to provide information about the scope of the project, it is distinct from the subject or focus of the resources within the archive.

Project Type

(ADS:ProjectType)

Depositors and users of the ADS Archive classify heritage investigations and research by the type of work undertaken. Because the need for this form of classification is widespread in the ADS Archive's designated community, every project deposited at the ADS will have one or more 'Event Types' recorded in the Catalogue.

While classifications can vary, the Forum on Information standards in Heritage (FISH) has developed the FISH Archaeological Event Type Thesaurus, the OASIS heritage activity recording system uses this wordlist, so this information will primarily be created by the Depositors via the OASIS system.

Project Prompt

Understanding the reason why a project was undertaken is a mandatory requirement for strategic planning of Collections policy.

The Archivist should use a single term from the [FISH Reason for Investigation](#) vocabulary.

3.2.2. Records and Collections Management

Access terms (licence agreements)

(DCMI:accessRights, ISAD(G) 3.4.1 Conditions Governing Access)

The Catalogue will record the type of licence agreement signed by the Depositor at the point of data submission, a single licence agreement covers all objects within a deposited (natural) collection and the data for each deposited collection will be related to this agreement through the hierarchical relationship between the OMS and the CMS.

Depositor name and contact details

The depositor name and electronic contact details are mandatory in order for the Archivist to have a point of contact to discuss any issues with the deposited Collection.

This information will be provided by the depositor at the point of submission, particularly via the primary ADS ingest system (currently ADS-Easy).

Copyright holder

(DCMI:rightsHolder)

The copyright holder of the Resource is mandatory in order to fulfill the requirements of the ADS Licence agreement, and to display with the resource (meta)data.

This information will be provided by the depositor at the point of submission, particularly via the primary ADS ingest system (currently ADS-Easy).

Update frequency

(DCMI:accrualPeriodicity)

The update frequency is mandatory, and is required by MEDIN/NERC and for internal management of collections. [ISO frequency of update codelist](#)

Lineage

(DCMI:provenance)

A literal statement explaining who created the data, and who is currently hosting and maintaining it. This is applicable at the Collection level only, and should follow a set form of words as set out in the Cataloguing Procedure.

Resource language

(DCMI:language)

Resource language is mandatory, and should use a literal, and controlled vocabulary ISO 639-2.

Release status

(DCMI: available)

Project actions such as the first contact regarding deposition of data, provision of a quote, deposition of data, completed accessions, completed preservation procedures, completed final checks, and release online¹ are all recorded within the CMS.

When the appropriate ingest, preservation, cataloguing, and checking procedures have been completed, the Archivist will mark the collection as 'Ready to Release' or set an embargo date as appropriate. The Catalogue can therefore be searched to see the status of any given collection, or project.

Collection Type

Primarily for internal management purposes, and interface management, the ADS has developed a classification of its collections, aimed at providing statistics as to how many collections are disseminated on different areas of the ADS Website, or its sister site Internet Archaeology. The Collection Type also includes terms as defined by AO-CAT to ensure the record meets the requirements of the Ariadne plus portal.

This classification is a mandatory field in the CMS.

¹ <http://purl.org/dc/terms/available>

3.2.3. Describing the Resource

Object Level Resource Types

Resource Type is a mandatory requirement used to identify different types of content represented within an individual object for management, user access and searching.

The Archivists are the primary creators of this metadata and the terms will derive from the English Heritage and National Trust Resource Description where possible, though further terms may be defined by the Archivists where an appropriate vocabulary term does not exist. A single object may comprise multiple resource types. Definitions and use of these terms within the ADS will be clearly described in the forthcoming Cataloguing Procedure.

Subject Classification

(DCMI: Subject)

All resources within the ADS Archive, whatever their archival level, will be catalogued with subject terms taken, in the first instance from the [FISH Heritage Subjects & Themes thesaurus](#).

Additional subject keywords may be added as discovery aids, these should be taken from linked open data resources where possible, and should include in particular the heritage-specific FISH vocabularies which are integrated into the CMS cataloguing interface, as well as broader subject thesauri such as [The Library of Congress Subject Headings](#). Where possible, concept URI's will be stored in the ADS CMS along with the relevant literal strings. Any non-linked keywords or phrases will be periodically mapped to linked open data terms where possible.

Geographic Subject Classification

(DCMI: Subject)

Spatial terms should only be used to describe resources where the subject of that resource centres around a specific location. These terms should not be used to describe where a resource was created.

Place names will be taken from the [Getty Thesaurus of Geographic Names](#)

Biographic Subject Classification

(DCMI: Subject, foaf:focus)

Where the resource focuses on a specific person or organisation, the person's name will be added as a subject term. Ideally this would include either an ISNI ID or a WikiData ID.

This field is mandatory only when a resource has an individual or organisation as the primary focus of the resource, such as an obituary, or biographical journal article, for example, or a photograph depicting an individual or organisation, or representation of a historic character.

Chronological Subject Classification

(DCMI: Subject)

Where the resource focuses on a specific chronological period, this time period will be recorded in the catalogue. This may be distinct from the project period as recorded in the Provenance metadata, as the individual resources deposited may have a different focus than the overall interpretation resulting from the project as a whole.

It may be that the content of the resource does not focus on a specific period of time as the primary subject; this field is therefore mandatory only in cases where a distinct time period is the focus of the resource, or where a definitive interpretation has been documented for the subject of the resource. For example, if an image portrays a ditch, and that ditch has been interpreted as originating in the Roman period in the accompanying documentation submitted by the depositor, then the term 'Roman' should be added as a subject term in the Catalogue entry for that image.

The terms themselves should derive from linked data vocabularies, such as the [FISH Periods List](#) where possible.

3.2.4. Preservation Management

Data Types

Every digital object will be given a 'data type'. Data type can be multiple for certain objects and use should follow Cataloguing Procedure. For ADS purposes, these terms classify objects according to their documentation, management, and preservation needs. Each object may only be assigned a single data type classification and the terminology used for this purpose is defined by the ADS, based upon DCMI terms where possible. Heritage data comes in a vast range of formats and is used for a variety of purposes, so the DCMI Type list is not sufficient to cover all ADS Data Types, and therefore requires additional types to be added in order to create clear categories for metadata classification.

Preservation Actions

The lifecycle of every digital object (i.e. the actions/events performed on the digital object) will be recorded using [Preservation Events Controlled Vocabulary](#). At the time of writing the following actions will be documented:

- Deaccession
- Deletion
- filename change
- Modification
- Migration
- Normalization
- Redaction

It is acknowledged that further work has to be done on widening the list of actions to include events such as appraisal, ingest, and metadata modification. The nascent Preservation Action Registries (PAR) set of 'Preservation Actions' will also be reviewed as an alternative option.

File Formats

File format will be generated using the [National Archives' DROID tool](#). An automated procedure should record the:

- Mime type
- File format literal name
- File format version

File Size

File size (bytes) should be recorded, and generated by an automated ingest tool.

File Name

File name should be recorded and matching the accessioned version of the file (i.e. matching Repository Operations - Ingest Procedure). This should be generated by an automated ingest tool.

File Location

The file location should be recorded, and generated by an automated tool within the ADS CMS.

Object Relationships

Relationships between Objects within the ADS archive should use the [PREMIS Relationship Type Collection](#).

4. Content Enhancement

Much of the documentation needed as part of the minimum cataloguing requirements should be created and submitted by the depositors as the creators of the 'natural' collections.

Where recording and ingest systems (OASIS and ADS-Easy) allow, this metadata will be in the form of linked data and will provide quantitative control to ensure that minimum standards are met in terms of the required fields being populated by the depositor, and to provide quality assurance where possible.

The Archivists' role in terms of cataloguing descriptive metadata, therefore, is to provide quality assurance where the ingest systems do not, and to enhance the submitted metadata by mapping keywords to established thesauri where possible, and appropriate, and to use their expertise as archivists and heritage professionals to add descriptive search terms where a reuse case demands.

The Digital Archivist (Collections and Records) will also identify cataloguing needs in legacy data, where older collections fail to meet current standards. These needs will be translated into cataloguing projects and may inform future strategic aims.

5. Future Plans

5.1. Content Enhancement

Content enhancement is currently built into the CMS interface, with integrated vocabulary look-ups and classification drop-down lists, making content enhancement at a Collection-level straightforward and simple, and as such, this is standard procedure at Collection-level.

The current Curatorial Strategy includes plans to review the OMS and build an Object-level interface that will work in a similar way to facilitate content enhancement at a more granular level for the smallest level of resources within the ADS Archive. (see also 6.3 Addressing the Backlog).

Specific areas of content enhancement will include the capability to add tags for resources to allow for user engagement and informal labelling and classification beyond the largely sector-specific cataloguing elements currently used. In

addition hashtags will be added so that, during the cataloguing process, archivists can identify resources that may be useful to future social media campaigns. The facility to do this already exists within the ADS Library internal record editing forms; the aim being to introduce this into the CMS and OMS interfaces for consistency.

5.2. Bibliographic Records

The ADS is currently undertaking a review of the bibliographic resources with the ADS Archive. This review involves looking at what needs to be documented and transforming the Library metadata from an externally inherited structure to integrate with the standards and structure in place for the rest of the ADS Archive (the CMS and OMS), as well as mapping the fields we choose to take forward to qualified [Dublin Core and Marc21](#).

5.3. Addressing the Backlog

By introducing reuse values, and an audit page to identify gaps in the ADS Inventory and Catalogue, backlog Cataloguing Projects will be identified. These projects can then be incorporated into the Curatorial Strategy Action Plan to bring all ADS Resources up to current standards. It is envisaged that any such projects would focus first on Collection-level resources, then work towards having the same standards across all Object-level resources, placing emphasis on those collections with a greater potential reuse value.

5.4. Managing Versions

Over time, the ongoing preservation and curation of the data within the ADS Archive and its associated metadata and catalogue entry may change as requirements change. It is important to be able to view the data and metadata as they were originally deposited, and to identify when and what kind of changes have been made throughout the life of any given resource.

5.4.1. Digital Files

A file may be replaced, removed, or amended for the following reasons:

- It has been identified as a duplicate of another file in the ADS Archive.
- The depositor has requested that the file no longer be a part of the archive.
- The depositor has updated the file (for example, replacing a draft version of a report with a final version, or a more accessible version).
- The file has been found to contain sensitive information.
- The file contains inaccurate information.

Where a file has been removed or amended, this will be reflected in the catalogue, so that the catalogue contains not only a picture of the resources currently within the ADS Archive, but also those that may have changed or been removed from the Archive entirely. This will ensure that a complete picture of a resource's lifecycle at the ADS is preserved.

5.4.2. Metadata Records

The metadata associated with a resource will be subject to change as records are enhanced, restructured, or amended to reflect changes in policy, procedure, and user requirements.

Future versions of the ADS Catalogue structure should allow for timestamping and the identification of any curatorial activity so that internal and external users of the Catalogue can view the original record as created by the depositors, as well as any subsequent versions up to the latest record information, which should be the most compliant with the current policy at the time of viewing.

5.5. Identifying research interests

It is envisaged that the information stored about individuals and organisations will evolve as the ADS works to engage users and depositors in order to support ongoing curation. As part of this process, and in response to Internet Archaeology requirements, the ADS will aim to associate individuals with discovery elements so that new resources can be surfaced that match specific

user requirements, social media interaction can reflect documented user interests and, in the case of Internet Archaeology, this will be useful in identifying appropriate peer reviewers for new articles.

6. ADS Catalogue: Essential Fields List

The following section provides an overview of the ADS Catalogue standard

Term = the internal name of the term we need

Level = The level this is applied to: o(object) and/or r(resource)

Source = Captured (from template, OASIS, ADS-EASY) | Curated (by ADS staff) | Created (by automated ADS tools)

Standard element = Mapping to our preferred Data Standards or required external Standards

Vocabulary = Link to vocabulary if being used

Term	Level	Source	Vocabulary/Scheme used
Identifier	o,r	Created	ADS internal / DOI
ExternalAssetId	r	Captured Curated	
Title	o,r	Captured Curated	N/A
Description	o,r	Captured Curated	N/A
Creator	o,r	Captured Curated	ORCID
Level	o,r	Curated	ADS internal vocabulary
Publisher	o,r	Captured Curated	ORCID
MetadataPublisher	o,r	Created	ORCID
MetadataContact	o,r	Created	ORCID
MetadataVersion	o,r	Created	N/A

MetadataDate	o,r	Created	N/A
MetadataLanguage	o,r	Created	ISO 639-2
ExternalAssetID	r	Captured Curated	ADS internal vocabulary
ActivityBoundingBox	r	Captured Curated	N/A
ActivityCoordinate	r	Captured Curated	N/A
ActivityLocation	o,r	Captured Curated	Getty Thesaurus of Geographic Names
ActivityCreationDates	o,r	Captured	N/A
ActivityType	r	Captured Curated	FISH Archaeological Event Type Thesaurus
Prompt	r	Captured Curated	FISH Reason for Investigation
AccessConditions	r	Created	ADS internal vocabulary
Depositor	r	Captured	ORCID
CopyrightHolder	o,r	Captured	ORCID
UpdateFrequency	r	Curated	ISO Frequency Code List
Lineage	r	Created	N/A
Language	o,r	Captured	ISO 639-2
ReleaseStatus	r	Curated	ADS internal vocabulary
CollectionType	r	Curated	Internal ADS list
Subject	o,r	Captured Curated	FISH Heritage Subjects & Themes thesaurus LCSH
PersonAsConcept	o,r	Captured Curated	Examples include Getty ULAN or WikiData
PeriodAsConcept	o,r	Captured Curated	Heritage Data
LocationAsConcept	o,r	Captured Curated	Getty Thesaurus of Geographic Names
DataType	o	Created	PRONOM

		Curated	
FileFormat	o	Created	PRONOM
FileSize	o	Created	N/A
FileName	o	Captured	N/A
FileLocation	o	Created	N/A
Relations	o	Created Curated	PREMIS