4.1 - Archive

When DDI talks of archive it is referring to specific activities that take place for the purpose of organizing and preserving the metadata and related data files. Archive activities are performed by the individual or organization responsible for maintaining and preserving the metadata. This could be the individual researcher, a data production organization, or a formal archive dedicated to the preservation of metadata and data. It focuses on the information related to management, preservation, and access.

Archive is a module which contains two types of information. The first is information specific to the instance of the metadata in a specific location. This “ArchiveSpecific” information may have no relevance outside of a specific location. For example, information on the physical storage location of the metadata parts, how they fit into a local collection, or what related materials the archive may contain. In addition to the “ArchiveSpecific” information, the Archive module contains a record of events that are significant in the life of the metadata and houses the OrganizationScheme which describes the organizations and individuals related to the metadata. OrganizationScheme and its contents are described in “Organizations, Individuals, and Relations” (xx.xx.xx). An OrganizationScheme may be included in-line or by reference. This allows for the maintenance of a master OrganizationScheme as a ResourcePackage and the inclusion of the relevant sections in individual Archive modules.

The Archive section describes the overall structure of the module, describes the ArchiveSpecific information and provides details on AccessType which is used to describe both default and specific access restrictions at the level of the archive, the collection, or an item.

Archive

Extension base: r:MaintainableType

ArchiveModuleName (0..n)

r:Label (0..n)

r:Description (0..1)

ArchiveSpecific (0..n)

CHOICE (0..n)

OrganizationScheme

r:OrganizationSchemeReference

END CHOICE

r:LifecycleInformation (0..1)

LifecycleEvent (0..n)

Extension base: IdentifiableType

Label (0..n)

EventType (0..1)

Date (0..1)

AgencyOrganizationReference (0..n)

Description (0..1)

Relationship (0..n)

r:OtherMaterial (0..n)

The Archive contains an ArchiveModuleName, Label, and Description to support its maintenance within a registry system. The purpose of the module is to organize the information about the location, management, and important events surrounding the metadata to which it belongs. ArchiveSpecific information is discussed below and the details of the OrganizationScheme are found in “Organizations, Individuals, and Relations” (xx.xx) . A key feature in Archive is the LifeCycleInformation object which contains an unsorted stack of LifecycleEvents. These events are identifiable and may have Note or OtherMaterials attached to them. These should all be contained within the Archive module so they do not risk becoming disassociated with the related object.

A LifecycleEvent is any event that the user decides has value to the archive or the end user. Capturing data and metadata processing information related to ingest and management of metadata within an archive or similar organization is a common use of this section. The LifecycleEvent contains a label for the event, specifies the type of event being captured (supports the use of a controlled vocabulary for this object), specifies the date of the event, the organizations or individuals involved by reference to their description in an OrganizationScheme, and describes the event. The event may be attached to one or more identifiable objects within the metadata by use of Relationship. This is the same structure as used in Note(xxx.xx) and OtherMaterial (xxx.xx)

ArchiveSpecific

ArchiveOrganizationReference (0..1)

Item (0..n)

r:Citation (0..1)

LocationInArchive (0..n)

CallNumber (0..1)

r:URI (0..1)

ItemFormat (0..1)

Media (0..1)

StudyClass (0..1)

r:Description (0..1)

ClassType (0..1)

Access (0..1)

OriginalArchiveOrganizationReference (0..n)

AvailabilityStatus (0..1)

DataFileQuantity (0..1)

CollectionCompleteness (0..1)

Item (0..n)

Collection (0..n)

r:Citation (0..1)

LocationInArchive (0..n)

CallNumber (0..1)

r:URI (0..1)

ItemQuantity (0..1)

StudyClass (0..1)

r:Description (0..1)

ClassType (0..1)

DefaultAccess (0..1)

OriginalArchiveOrganizationReference (0..n)

AvailabilityStatus (0..1)

DataFileQuantity (0..1)

CollectionCompleteness (0..1)

Item (0..n)

Collection (0..n)

DefaultAccess (0..n)

r:FundingInformation (0..n)

r:Budget (0..n)

r:QualityStatementReference (0..n)

r:Coverage (0..1)

ArchiveSpecific focuses on the place of the metadata within the collections of the organization which maintains the original or a local depository copy. The first object in ArchiveSpecific is a reference to the description of the ArchiveOrganization itself. This description within an OrganizationScheme should contain all of the relevant contact information for the Archive Organization, any special identifiers such as its DDI Agency identification, information on how it handles versioning, and any other information that would be useful to anyone needing to interact with the agency. Item and Collection, both of which can internally nested objects of the same type, provide flexibility for the archive to describe its holdings. At the level of ArchiveSpecific you can describe the general or DefaultAccess rules and processes for the Archive (these may be superseded by rules for specific items or collections), FundingInformation for archival processing or collection and Budget information. QualityStatements, such as adherence to OAIS or various certification programs, are included here by reference. A statement of the overall coverage of the archive may be described in spatial, temporal, and topical terms.

Item and Collection are similar in structure. Archives deal with addressable byte streams, files, representations (objects collectively representing an intellectual entity), as well as collections and sub-collections of those objects. An archive should be internally consistent in how it organizes this information about the objects it contains, but DDI does not dictate the organization structure. For example, a Representation may be described as an Item containing several sub-items each representing a file and/or byte stream. Others may define a Representation a form of Collection with an Item used to define individual files and byte streams.

Both Item and Collection capture the following information about a digital object: a citation for the object, it’s physical location in the archive such as a remote or offline storage location, a CallNumber (internal identification number), the URI, the StudyClass which is used to represent any processing or other non-topical means of classification within the archive, a reference to the original archive organization (where did it come from), availability status, quantity of related data files, a statement concerning the completeness of the collection, and a recursive object (Item or Collection) used to create nested hierarchies.

The unique objects in Item refer to the physical manifestation of a single item including the ItemFormat, Media, and Access restrictions/permissions specific to the Item. The unique objects in Collection reflect its description of a set of objects including, ItemQuantity (a check sum), and DefaultAccess providing restrictions/permissions for the set (these can be overridden at the item level).

AccessType

Extension base: r:IdentifiableType

AccessTypeName (0..n)

r:Label (0..n)

r:Description (0..1)

ConfidentialityStatement (0..1)

AccessPermission (0..n)

FormNumber (0..1)

r:URI (0..1)

Statement (0..1)

@isRequired (optional)

Restrictions (0..1)

CitationRequirement (0..1)

DepositRequirement (0..1)

AccessConditions (0..1)

Disclaimer (0..1)

AccessRestrictionDate (0..1)

ContactOrganizationReference (0..n)

AccessType is used by both Access and DefaultAccess to describe the details of any permissions or restrictions related to a specific object or a set of objects. In general, default access specifications at the archive or collection level are applied to their contained objects unless overridden by conflicting Access information lower down. For example, an archive may have a policy of open access to all users, but a specific collection may require registration, or an individual object may be private and inaccessible to any users outside of the research project that created it.

AccessType is identifiable so that external documents may be associated with it. It contains a Name, Label, and Description to assist in managing multiple access statements. Note that while the term “restriction” is used the content may be captured as descriptions of specific restrictions or permissions, whichever is clearer. AccessRestrictionDate provides a timeframe for the access description. ContactOrganizationReference provides a reference to the organization or individual to contact for further information regarding access. If an organization or sub-organization is referenced, changes in personnel or contact numbers can be captured in the OrganizationScheme. The following descriptive content are all of type StructuredStringType supporting both language replication and structured content for clarity: ConfidentialityStatement, Restrictions, CitationRequirement, DepositRequirement, AccessConditions, and Disclaimer. AccessPermission is a reference to a specific form, either physical or digital, used to arrange for access to an item. It contains a form number, a URI, a Statement relaying the use and purpose of the access form (InternationalStringType), and a Boolean attribute indicating whether completion of the form is required for access.

The following example contains all the major sections noted above including the ArchiveSpecific information, a short set of Items within a collection, three LifecycleEvents, and a related OtherMaterial brief citation. Access information is provided for the Archive (MPC\_NHGIS) as DefaultAccess.

Example:

<a:Archive isMaintainable="true" typeOfIdentifier="Caonical" scopeOfUniqueness="Agency">

<r:URN>urn:ddi:us.mpc:Arch\_1:1</r:URN>

<a:ArchiveModuleName context="IPEDS"><r:String xml:lang="en-US">MPC\_NHGIS\_HIST1900</r:String></a:ArchiveModuleName>

<r:Label><r:Content xml:lang="en-US">MPC NHGIS Archive Record for HIST1900</r:Content></r:Label>

<r:Description><r:Content xml:lang="en-US">Information specific to the HIST1900 data files and documentation in the NHGIS project</r:Content></r:Description>

<a:ArchiveSpecific>

<a:ArchiveOrganizationReference isReference="true" isExternal="true" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.mpc:OrgS\_MPCwide:1.0</r:URN>

<r:TypeOfObject>OrganizationScheme</r:TypeOfObject>

</a:ArchiveOrganizationReference>

<a:Collection>

<r:Citation>

<r:Title>NHGIS Historical 1900</r:Title>

</r:Citation>

<a:ItemQuantity>2</a:ItemQuantity>

<a:AvailabilityStatus>On-Line</a:AvailabilityStatus>

<a:DataFileFileQuantity>4</a:DataFileQuantity>

<a:CollectionCompleteness>Original data dictionary, DDI documentation, US level data file, State level data file, and County level data file.</a:CollectionCompleteness>

<a:Item>

<r:Citation>

<r:Title>us1990cnty</r:Title>

</r:Citation>

<a:LocationInArchive>aggdata/Translation/HIST/us1990cnty.sps</a:LocationInArchive>

<a:ItemFormat>ASCII text</a:ItemFormt>

<a:ItemQuantity>1</a:ItemQuantity>

<a:AvailabilityStatus>archive</a:AvailabilityStatus>

<a:DataFileFileQuantity>1</a:DataFileQuantity>

<a:CollectionCompleteness>SPSS setup file and related data file.</a:CollectionCompleteness>

<a:Item>

<a:Item>

<r:Citation>

<r:Title>Census of Population for National, State, and County Levels - 1900: NHGIS documentation</r:Title>

</r:Citation>

<a:LocationInArchive>NHGIS/metadata/DDI/HIST1900-cnty.xml</a:LocationInArchive>

<a:ItemFormat>xml</a:ItemFormt>

<a:ItemQuantity>1</a:ItemQuantity>

<a:AvailabilityStatus>NHGISsubversion</a:AvailabilityStatus>

<a:DataFileFileQuantity>3</a:DataFileQuantity>

<a:CollectionCompleteness>DDI documentation, US level data file, State level data file, and County level data file.</a:CollectionCompleteness>

<a:Item>

<a:Collection>

<a:DefaultAccess isIdentifiable="true" typeOfIdentifier="Caonical" scopeOfUniqueness="Agency">

<r:URN>urn:ddi:us.mpc:NHGIS\_ACCESS\_1:1</r:URN>

<a:AccessName context="IPEDS"><r:String xml:lang="en-US">MPC\_NHGIS\_ACCESS</r:String></a:AccessName>

<r:Label><r:Content xml:lang="en-US">NHGIS Access Requirements</r:Content></r:Label>

<r:Description><r:Content xml:lang="en-US">Describes access and usage requirements for NHGIS</r:Content></r:Description>

<a:ConfidentialityStatement><r:Content xml:lang="en-US">NHGIS contains public data obtained through the U.S. Census Bureau.</r:Content></a:ConfidentialityStatement>

<a:Restrictions><r:Content xml:lang="en-US">The National Historical Geographic Information System (NHGIS) provides, free of charge, aggregate census data and GIS-compatible boundary files for the United States between 1790 and 2010.</r:Content></a:Restrictions>

<a:AccessPermission isRequired="true">

<r:URI>https://data2.nhgis.org/users/login</r:URI>

</a:AccessPermission>

<a:CitationRequirement><r:Content xml:lang="en-US">All persons are granted a limited license to use this documentation and the accompanying data, subject to the following condition:<xhtml:ul><xhtml:li>Publications and research reports based on the database must cite it appropriately. The citation should include the following: Minnesota Population Center. National Historical Geographic Information System: Version 2.0. Minneapolis, MN: University of Minnesota 2011.</xhtml:li><xhtml:li>If possible, citations should also include the URL for the NHGIS site: http://www.nhgis.org</xhtml:ul><xhtml:br/>In addition, we request that users send us a copy of any publications, research reports, or educational material making use of the data or documentation. Printed material should be sent to:<xhtml:br/>NHGIS<xhtml:br/> Minnesota Population Center<xhtml:br/> University of Minnesota<xhtml:br/>50 Willey Hall<xhtml:br/>225 19th Ave S<xhtml:br/>Minneapolis, MN 55455</r:Content></a:CitationRequirement>

<a:AccessConditions><r:Content xml:lang="en-US">Preregistration</r:Content></a:AccessConditions>

<a:AccessRestrictionDate><r:StartDate>2001-05-01</r:StartDate></a:AccessRestrictionDate>

<a:ContactOrganizationReference isReference="true" isExternal="true" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.mpc:OrgS\_MPCwide:1.0</r:URN>

<r:TypeOfObject>OrganizationScheme</r:TypeOfObject>

</a:ContactOrganizationReference>

</a:DefaultAccess>

</a:ArchiveSpecific>

<r:OrganizationSchemeReference isReference="true" isExternal="true" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.mpc:OrgS\_MPCwide:1.0</r:URN>

<r:TypeOfObject>OrganizationScheme</r:TypeOfObject>

</r:OrganizationSchemeReference>

<r:LifecycleInformation>

<r:LifecycleEvent isIdentifiable="true" typeOfIdentifier="Caonical" scopeOfUniqueness="Agency">

<r:URN>urn:ddi:us.mpc:NHGIS\_Event\_1:1</r:URN>

<r:Label><r:Content xml:lang="en-US">HIST1900-cnty aquisition</r:Content></r:Label>

<r:EventType>Aquisition</r:EventType>

<r:Date><r:SimpleDate>2001-03-10</r:SimpleDate></r:Date>

<r:AgencyOrganizationReference isReference="true" isExternal="true" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.mpc:OrgS\_MPC\_NHGIS:1.0</r:URN>

<r:TypeOfObject>Organization</r:TypeOfObject>

</r:AgencyOrganizationReference>

<r:Description><r:Content xml:lang="en-US">The 1900 History file was aquired as part of a collection of 1900-1950 data entered from printed documents through NGHIS subcontract.</r:Content></r:Description>

<r:Relationship>

<r:RelatedToReference isReference="true"true isExternal="false" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.mpc:NHGIS\_HIST1900-cnty:1</r:URN>

<r:TypeOfObject>StudyUnit</r:TypeOfObject>

</r:RelatedToReference>

<r:RelationshipDescription><r:Content xml:lang="en">DDI Documentation</r:Content></r:RelationshipDescription>

</r:Relationsip>

<r:LifecycleEvent>

<r:LifecycleEvent isIdentifiable="true" typeOfIdentifier="Caonical" scopeOfUniqueness="Agency">

<r:URN>urn:ddi:us.mpc:NHGIS\_Event\_1:1</r:URN>

<r:Label><r:Content xml:lang="en-US">HIST1900-cnty creation of DDI metadata document</r:Content></r:Label>

<r:EventType>DDICreation</r:EventType>

<r:Date><r:SimpleDate>2001-04-01</r:SimpleDate></r:Date>

<r:AgencyOrganizationReference isReference="true" isExternal="true" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.mpc:OrgS\_MPC\_NHGIS:1.0</r:URN>

<r:TypeOfObject>Organization</r:TypeOfObject>

</r:AgencyOrganizationReference>

<r:Description><r:Content xml:lang="en-US">Data dictionary information plus additional content from original print documents and materials regarding the 1990 U.S. Census in IPUMS were complied in DDI and verified.</r:Content></r:Description>

<r:Relationship>

<r:RelatedToReference isReference="true"true isExternal="false" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.mpc:NHGIS\_HIST1900-cnty:1</r:URN>

<r:TypeOfObject>StudyUnit</r:TypeOfObject>

</r:RelatedToReference>

<r:RelationshipDescription><r:Content xml:lang="en">DDI Documentation</r:Content></r:RelationshipDescription>

</r:Relationsip>

<r:LifecycleEvent>

<r:LifecycleEvent isIdentifiable="true" typeOfIdentifier="Caonical" scopeOfUniqueness="Agency">

<r:URN>urn:ddi:us.mpc:NHGIS\_Event\_1:1</r:URN>

<r:Label><r:Content xml:lang="en-US">HIST1900-cnty integration into 2001 NHGIS release</r:Content></r:Label>

<r:EventType>NHGISIntegration</r:EventType>

<r:Date><r:SimpleDate>2001-05-01</r:SimpleDate></r:Date>

<r:AgencyOrganizationReference isReference="true" isExternal="true" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.mpc:OrgS\_MPC\_NHGIS:1.0</r:URN>

<r:TypeOfObject>Organization</r:TypeOfObject>

</r:AgencyOrganizationReference>

<r:Description><r:Content xml:lang="en-US">The HIST1990-cnty metadata and data files were successfully integrated into the NHGIS system for the May 2001 release.</r:Content></r:Description>

<r:Relationship>

<r:RelatedToReference isReference="true"true isExternal="false" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.mpc:NHGIS\_HIST1900-cnty:1</r:URN>

<r:TypeOfObject>StudyUnit</r:TypeOfObject>

</r:RelatedToReference>

<r:RelationshipDescription><r:Content xml:lang="en">DDI Documentation</r:Content></r:RelationshipDescription>

</r:Relationsip>

<r:LifecycleEvent>

</r:LifecycleInformation>

<r:OtherMaterial isIdentifiable="true" typeOfIdentifier="Caonical" scopeOfUniqueness="Agency" xml:lang="en">

<r:URN>urn:ddi:us.mpc:NHGIS\_Event\_1:1</r:URN>

<r:TypeOfMaterial>Print.Book</r:TypeOfMaterial>

<r:Description><r:Content xml:lang="en-US">Data source</r:Content></r:Description>

<r:Citation>

<r:Title>Twelfth Census of the United States Taken in the Year 1990, Census Reports Volume I - Population Part I</r:Title>

<r:Creator>

<r:CreatorName affiliation="Interior Department">United States Census Office</r:CreatorName>

</r:Creator>

<r:Publisher>

<r:PublisherName>Washington: United States Census Office</r:PublisherName>

</r:Publisher>

<r:PublicationDate>

<r:SimpleDate>1901</r:SimpleDate>

</r:PublicationDate>

</r:Citation>

<r:Relationship>

<r:RelatedToReference isReference="true"true isExternal="false" lateBound="false" typeOfIdentifier="Canonical">

<r:URN>urn:ddi:us.mpc:HIST1900-cnty\_Item2:1</r:URN>

<r:TypeOfObject>Item</r:TypeOfObject>

</r:RelatedToReference>

<r:RelationshipDescription><r:Content xml:lang="en">Data Source</r:Content></r:RelationshipDescription>

</r:Relationsip>

</r:OtherMaterial>

</a:Archive>