

# Specification for Subject Records

Final Version: 20100706

## Su--1: Scope / Purpose

The subject record is for indicating the topical content or format characteristic(s) of an item or set of materials being described. The subject record is not designed to record proper names that represent the topical content of materials; that is done via the name record.

The subject record is designed to support simple or compound, hierarchical subjects. It can be applied at any level of description for accessions, resources, and digital objects. And it supports exporting of subject headings to MARCXML, MODS, Dublin Core, VRA, and EAD.

The ability to designate the type of information conveyed by individual subject terms within a larger subject string composed of hierarchically related “simples” is reflective of the granularity of encoding supported by standards such as MARC and MODS. An intended benefit of the subject record design is to provide repositories richer options for organizing, exporting, searching, and displaying subject headings. Moreover this design will be conducive with the use of extension schemas envisioned for the EAD schema.

This record design will also help to enable more automated data flows, in that the granularity of the data imported will largely be retained and the need to replicate manually that granularity on export will diminish. This marks a substantial improvement on the data flow currently for the AT and will result in improved exports for data currently in Archon. Presently, the MARC subject heading “650 #0 Publishers and publishing |z New York (State) |x Manuscripts” is imported and stored as “Publishers and publishing—New York (State)—Manuscripts”. When exported back to a MARC data format, it is encoded as “650 #0 Publishers and publishing—New York (State)—Manuscripts”, an encoding that is not consistent with other subject headings in the MARC catalog and thus disruptive to searching and display protocols. To ameliorate the problem, the operator has to edit the exported MARC record, manually adding the subfield delimiters again to the subject heading. This problem is even more cumbersome and time consuming with a MODS data environment where the delimiters are field names and not two key stroke delimiters.

<subject>

```
<topic>Publishers and publishing</topic>
<geogname>New York (State)</geogname>
<genre>Manuscripts</genre>

</subject>
```

## Su--2: Record / template description

An operator will have the option to create a subject record directly, and then link it subsequently to a selected context record, or to create a subject record while working in a context record, in which case the subject record will be automatically linked to the open context record.

The subject record will include the following parts, arranged in a manner deemed appropriate by the user interface experts.

### 1) Required fields

- Subject term 1
- Subject term 1 type
- Subject term source
- Subject display form
- ExportsEnabled (Default= YES)

### 2) Optional fields

- Subject term 2
- Subject term 2 type
- Subject term 3
- Subject term 3 type
- Subject term 4
- Subject term 4 type
- Subject term 5
- Subject term 5 type

- Subject term 6
- Subject term 6 type
- Subject scope note
- Subject term identifier

### 3) Linked records

- Accession records
- Resource records
- Resource component records
- Digital object records
- Digital object component records

### 4) System Control Data (generated by application and not editable by operator)

- Record Created Timestamp
- Last Modified Timestamp
- Record Created Staff Name
- Last Modified Staff Name

The operator must provide all the required fields for a valid subject record to be created and stored. The required fields constitute the primary subject heading. All other fields in the subject record are either amplify or supplement the primary heading (subject term identifier and subject term scope note) or qualify the primary subject heading (subject term and subject term type 2-6)

For the optional fields, the operator must use Subject Term 2 before using Subject Term 3, and so forth. The number of the option subject term is used to sequence the term in a compound subject statement. That is, the value for Subject Term 2 will always occupy the second position (left to right) in a compound subject statement.

Consequently, it should not be permissible to enter data into Subject Term 3 until Subject Term 2 and Subject Term 2 Type have been complete, and so forth for the entire set of subject heading parts.

Subject Term Scope and Subject Term Identifier, two other optional fields, may be used according to the repository's policies.

The “Linked Records” fields simply a listing of all the records in an implemented system to which a particular subject heading is linked. This provides a quick, in template view, of how often a subject heading is used in a given system.

(The technical implementation team should carefully consider data storage options that will best support the generation of subject linking mechanisms in the staff and public interface. Archon current stores subjects in a parent/child relationship, and the data could also be stored in five fields within one row of a single table. In addition, the implementation team should examine how subjects are controlled in ICA-Atom/Qubit, and may wish to consider the possibility of using nested sets.)

### **Su—3: Record uniqueness**

Duplicate subject records are not permitted. Duplication is calculated to occur if two subject records are identical in respect to:

- Subject Term 1
- Subject Term 1 Type
- Subject Term Source
- Subject Term Identifier

And where applicable:

- Subject Term 2
- Subject Term 2 Type
- Subject Term 3
- Subject Term 3 Type
- Subject Term 4
- Subject Term 4 Type
- Subject Term 5
- Subject Term 5 Type
- Subject Term 6
- Subject Term 6 Type

Creating and trying to save a duplicate subject record will result in the message “The subject record you are trying to exist already exists. You may not create a duplicate.”

#### **Su—4: Record linking / data relationships**

A subject record may be linked to many of the following context records:

- Accession records
- Resource records
- Resource component records
- Digital object records
- Digital object component records

As many subject records as necessary to describe the content / context of materials may be linked to one of the context records listed above.

A subject record may be linked only once to any of the context records listed above. That is, a subject record may not be linked twice to the same accession record, for example.

All context records to which a subject record is linked are to be displayed in the subject record by type of context record, i.e, accessions, resources, and digital objects.

Currently, this subject record specification makes no provision for linking subject records to records, although that may be a useful future development.

#### **Su—5: Removing subject record links / Deleting subject records**

An operator should be able to remove the link from a context record to a subject record. This simply severs the relationship between the context record and the subject record within the system. However, an operator will not have the ability to remove the link from a subject record to a certain context record. Removing of links to subject records may only be done from the context record.

An operator will be able to delete a subject record, in the event of which all links to context records from the deleted subject record will be removed. The user will be provided a warning: “Warning: deleting [name of subject] will sever all links to resource, resource component, accession, digital object, and digital object component records. Do you wish to proceed?”

#### **Su—6: Business rules**

1. Subject records are indicating the content / context of archival resources, including accessions and digital objects. They are used to promote topical (exclusive of proper names) and format access to the materials described, and they can be exported as controlled access terms to finding aids, subject headings in catalogs, or subject guides.
2. A valid subject record must have values recorded for subject term 1, subject term type 1, and subject term source.
3. Optional qualifying terms or supplemental information (scope note or reference ID) may be provided.
4. Each unique subject heading may be represented by only one unique subject record. Uniqueness is established by comparing for two records the values of subject term 1, subject term type 1, subject term source, subject term identifier, and all other qualifying subject terms used in the subject records. If all points of comparison are identical, one record is deemed to be a duplicate of a previous record and will be disallowed by the system.
5. A given subject record may be linked once to a given context record but to many different context records.
6. Subject records may be created and applied at a later time. There is no requirement that a subject record must be linked to be stored in the system.
7. Successful application of a subject records results in 1) the subject heading be listed in the context record as a descriptor for the materials described in the context record, and 2) the linked context record being listed in the subject record
8. An operator may remove a link by opening a context record and exercising the option to remove the link to a linked, or listed, subject record. Completion of the task will result in the subject record no longer being listed in the context record.
9. An operator may delete a subject record, in which case the record is removed from the system and all links from the subject record to all context records are lost.
10. Each new and edited subject record must include the following administrative information:
  - a. Record created timestamp
  - b. Last modified timestamp
  - c. ID of the staff user who created the subject record

- d. ID of the staff user who last modified the subject record
11. The Last Modified time stamp and staff user ID must be updated when the subject record is updated.

**Su—7: Required task sequence**

1. Select option to create a new subject record within the subject record module or within an open context record
2. Record a value for subject term 1
3. Indicate the type for subject term 1
4. Indicate the source for subject term 1
5. Save the subject record

**Su--8: Optional task sequence**

1. Add qualifying subject terms and subject term types as appropriate and according rules for constructing terms within a certain vocabulary. Up to five qualifying subject terms and types may be added to the subject record.
2. Add a note explaining the scope and use of the heading delineated in the subject record
3. Add an identifier to the subject heading, presumably the identifier of the heading in an external source
4. Change the Export Enabled field to 'yes' or 'no', as applicable.

**Su--9: User intentions / Application response sequence**

<b>User intention (Required fields in <i>italics</i>)</b>	<b>Application response / action</b>
<i>Select option to <b>create</b> a subject record either in the subject record module or within an open context record</i>	
	Displays a subject record with assigned repository default values.
<i>Record value for subject term 1</i>	

<i>Select type for subject term type 1</i>	
<i>Select source for subject term</i>	
<i>Save the subject record</i>	
	Application will indicate the record cannot be saved and display the reason, e.g. the subject record is not valid, in which case a list of required fields not completed will be displayed, or subject record is not unique, in which case the system identifier for the duplicate record will be displayed
	Or, application will indicate the record has been saved. A newly saved valid and unique subject record created in the subject record module will appear in the list of subject records in the systems. If it was created within a context record, the subject record will be listed in the context record as a linked subject record, as well as appearing in the list of all subject records in the system.

<b>User intention (Required fields in <i>italics</i>)</b>	<b>Application response / action</b>
Select subject record to edit	
	Opens the selected subject record will all previously recorded data
<i>Save the edited subject record</i>	
	Application will indicate the record cannot be saved and display the reason, e.g. the subject record is no longer valid, in which case a list of required fields not completed will be displayed, or subject record is no longer unique, in which case the system identifier for the duplicate record will be

	displayed
	Or, application indicates updated record has been saved.

<b>User intention (Required fields in <i>italics</i>)</b>	<b>Application response / action</b>
Select subject record(s) to delete	
<i>Select option to delete selected subject record(s)</i>	
	Application responds: "Are you sure you want to delete "N" subject record(s)?" Yes or No
Select No option	
	Application reverts to state before selecting delete option
Select Yes option	
	Application indicates "N" records have been deleted
	Application responds within two seconds
	Application refreshes list of subject records, now absent those deleted.

**Su--10: Subject data elements**

Element	Definition	Type	Default Values	Required
subjectDisplayForm	<p>The form of all the subject component values displayed as a single string in the sequence designated by the operator, e.g. Archery--Korea--20<sup>th</sup> century.</p> <p>Parts of the display subjectDisplayForm will be delimited by double hyphens with no intervening spaces (Archery--Korea)</p> <p>The subject DisplayForm will be capable of being edited only indirectly, by editing the values in the subject record.</p>	Automated display		Y
subjectTerm1		String		Y
subjectTermType1		Nonconfig. lookup	Cultural context Function Geographic Genre / form	Y

			Occupation Style / period Technique Temporal Topical Uniform title	
subjectTerm2		String		N
subjectTermType2		Nonconfig. lookup	Genre/Form Geographic Temporal Topical	Y, if SubjectTerm2 is used
subjectTerm3		String		N
subjectTermType3		Nonconfig. lookup	Genre/Form Geographic Temporal	Y, if SubjectTerm3 is used

			Topical	
subjectTerm4		String		N
subjectTermType4		Nonconfig. lookup	Genre/Form Geographic Temporal Topical	Y, if SubjectTerm4 is used
subjectTerm5		String		N
subjectTermType5		Nonconfig. lookup	Genre/Form Geographic Temporal Topical	Y, if SubjectTerm5 is used
subjectTerm6		String		N
subjectTermType6		Nonconfig. lookup	Genre/Form Geographic	Y, if SubjectTerm6 is used

			Temporal Topical	
subjectTermSource	The vocabulary from which an established term is use, e.g., Art and Architecture Thesaurus (aat)	Config. Lookup	Art and Architecture Thesaurus (aat)  Genre Terms: A Thesaurus for Use in Rare Books and Special Collections (rbgenr)  Getty Thesaurus of Geographic Names (tgn)  Library of Congress Subject Headings (lcs)h  Local sources (local)  Medical Subject Headings (mesh)  Thesaurus for Graphic Materials (gmgpc)	

subjectTermScopeNote	A note defining the term and explaining its use within the system	Text		N
subjectTermID	The unique identifier for a subject term	String		N
ExportEnabled	A value that determines whether the subject term should be included when MARCXML, EAD, MODS or other output is generated.	Y/N		Y
subjectRecordCreatedTimestamp		Timestamp	Automatically generated when valid record is saved	Yes
subjectRecordLastModifiedTimestamp		Timestamp	Automatically updated when record is modified	Yes

subjectRecordCreatedStaff			Automatically insert username of operator from user tables	Yes
subjectRecordLastModifiedStaff			Automatically insert username of operator from user tables	Yes

DRAFT

## Su--11: Imports

Subject records will be created from the import of data records when those data record contain subject headings. These will include accession records, EADs and MARCXML records for resources, and digital object records.

In addition, we should provide for the import of simple (one part) subject lists from a simple csv file containing the following fields: Term1, Type, Source, Scope, ExportEnabled. Import of multipart term lists will be supported from an xml file. The following XSD scaffold describes how the xml file should be structured.

Source Element	Where Stored/AT Target	Data Type	Rule
<subject>			
<term1>	subjectTerm1	string	
<termType1>	subjectTermType1	string	Types are : Cultural context Function Genre / form Geographic Occupation Style / period Technique Temporal Topical Uniform title
<term2>	subjectTerm2	string	
<termType2>	subjectTermType2	string	Types are : Genre/Form, Geographic, Temporal, and Topical
<term3>	subjectTerm3	string	
<termType3>	subjectTermType3	string	Types are : Genre/Form, Geographic, Temporal, and Topical
<term4>	subjectTerm4	string	
<termType4>	subjectTermType4	string	Types are : Genre/Form, Geographic, Temporal, and

			Topical
<term5>	subjectTerm5	string	
<termType5>	subjectTermType5	string	Types are : Genre/Form, Geographic, Temporal, and Topical
<term6>	subjectTerm6	string	
<termType6>	subjectTermType6	string	Types are : Genre/Form, Geographic, Temporal, and Topical
<subjectSource>	subjectSource	string	Default "ingest"
<subjectID>	subjectID	string	
</subject>			

**Example of a subject record in this format:**

```

<subject>
  <subjectTerm1>Railroads</subjectTerm1>
  <subjectTermType1>Topical</subjectTermType1>
  <subjectTerm2>Mexico</subjectTerm2>
  <subjectTermType2>Geographic</subjectTermType2>
  <subjectSource>Library of Congress Subject Headings
  </subjectSource>
  <subjectID>LC0989789</subjectID>
</subject>

```

Import fails for any record if subjectTerm1, subjectTermType1, or subjectSource are null.

## Su--12: Exports

Export of subject records will result in subject headings in Dublin Core, EAD, MARCXML, MODS, and VRA data formats.

Proposed Subject Types	EAD*	MARC - 6XX Fields	MARC - Subfields	MODS	DC*	VRA*
<i>Cultural context</i>	<subject>	650 Topical term		<topic>	<subject>	culturalContext
<i>Function</i>	<function>	657 Function		<topic> ??	<subject>	subject (term type = descriptiveTopic)
<i>Genre/Form</i>	<genreform>	655 Genre/Form	Form (#v) ????	<genre>	<subject>	worktype
<i>Geographic</i>	<geogname>	651 Geographic Name	Geographic (#z)	<geographic>	<coverage>	subject (term type = geographicPlace)
<i>Occupation</i>	<occupation>	656 Occupation		<occupation>	<subject>	subject (term type = descriptiveTopic)
<i>Style/Period</i>	<subject>	650 Topical term		<topic>	<subject>	stylePeriod
<i>Technique</i>	<genreform>	655 Genre/Form		<genre>	<subject>	technique
<i>Temporal</i>	<subject>	648 Chronological Term	Chronological (#y)	<temporal>	<coverage>	stylePeriod

<i>Topical</i>	<subject>	650 Topical Term	Topical term (#x)	<topic>	<subject>	subject (term type = descriptiveTopic)
<i>Uniform Title</i>	<title>	630 Uniform Title		<titleInfo>	<subject>	subject (term type = descriptiveTopic)

\*Do not support hierarchical expressions. Consequently, on export, hierarchically sequenced values will be “flattened” into a single, primary string.

### Examples of a subject record exported:

#### EAD

Compound subjects will be flattened for export to EAD as it does not support granular encoding.

```
<subject authority="lcsch">Publishers and publishing--New York (State)--Manuscripts</subject>
```

#### MARC

```
650 0 Publishers and publishing |z New York (State)|x Manuscripts
```

#### MODS

```
<subject authority="lcsch">
  <topic>Publishers and publishing</topic>
  <geographic>New York (State)</geographic>
  <genre>Manuscripts</genre>
</subject>
```

#### DC

Strings will be flattened for export to DC as it does not support granular encoding.

<subject>Publishers and publishing--New York (State)--Manuscripts</subject>

### VRA

Compound subjects will be flattened for export to VRA

<subject type="descriptiveTopic" vocab="lcs" refid="LC0989789">Publishers and publishing--New York (State)--Manuscripts</subject>

DRAFT

1. A list of selected subject records, sortable by heading, type, and source.
2. A list of subject headings with linked accession, resource, and / or digital object record. Scopable to context record type and sortable by subject heading and subject heading type. In effect, a subject guide to a repository's materials.

DRAFT