COLLABORATIVE EUROPEAN DIGITAL ARCHIVE INFRASTRUCTURE

Project Acronym: CENDARI
Project Grant No.: 284432
Theme: FP7-INFRASTRUCTURES-2011-1
Project Start Date: 01 February 2012
Project End Date: 31 January 2016

Deliverable No. :	D6.2
Title of Deliverable:	Guidelines for Applying the Schema
Date of Posting to Basecamp/Confluence for Partner Review:	December 2012
Date of Finalised Deliverable:	July 2013
Revision No.:	1
WP No.:	6
Lead Beneficiary:	King's College London (KCL)
Author (Name and email address):	Collaborative deliverable developed by the CENDARI WP6 team
Dissemination Level:	PU = public
Nature of Deliverable:	O = other
Abstract:	The first priority for the WP6 team was to create the institutional level metadata schema to enable WP5 to commence work on populating CENDARI with archival descriptions. This initial version was completed in December 2012. Subsequently, WP6 work on the metadata schema has concentrated on reviewing and deciding upon the necessary elements for the schema at collection and item level. Close consultation has taken place with the domain specialists to ensure it meets research needs. As such, although the schema maps to EAD, the archival standard used in many archives, it

	includes some additional elements to aid search and discovery in a historical context. We expect this approach to both allow for interoperability with archive portals and to satisfy the requirements of researchers. The schema was completed at the end of July 2013 There are three levels to the common metadata schema developed by the WP6 team: • Institutional-level • Collection-level • Item-level The guidelines for each level are presented as separate documents, which together make up this Deliverable. Document 1 – Introduction to the CENDARI Metadata Strategy Document 2 – EAG (CENDARI) Institutional Schema Guidelines Document 3 – CENDARI Collection Schema Guidelines Document 4 – CENDARI Item Schema Guidelines This deliverable should be read in conjunction with CENDARI Deliverable 6.1 – Common Metadata Schema: http://www.cendari.eu/public-project-deliverables/metadata/.

INFRA-2011-1-284432

Document 1

Introduction to the CENDARI Metadata Strategy

Introduction to the CENDARI Metadata Strategy

[Version 1.0 – 6/12/13]

1. Introduction

The Collaborative Digital Archive Research Infrastructure (CENDARI) was funded by the European Commission to create a humanities research infrastructure that integrates access to archives, connects knowledge, and supports the research process for two domains: First World War studies and medieval history. This research infrastructure is underpinned by a knowledge framework combines the development of an integrated metadata strategy with the development of dynamic domain ontologies. This document provides an overview of the integrated metadata strategy.

The CENDARI metadata strategy conceptualizes data about institutions and their archives at three levels:

- Institutional level
- Collection level
- Item level

The metadata strategy makes use of both new and existing metadata schemas. The description of holding institutions, are broadly met by a sub-set of a pre-existing schema, the Encoded Archival Guide (EAG). At the collection level the new CENDARI Collection Schema was created at the collection level due to the semantic limitations of the existing Encoded Archival Description (EAD). MODS (Metadata Object Description Schema was identified as meeting most of the CENDARI at an item level, although its standard element set is too generic for the more specific, particularly codicological, features necessary for many items in the medieval domain. However, MODS allows 'extension' elements to be embedded within its architecture, and additional elements are used from the TEI extension for the description of medieval manuscripts, as well as elements from the CENDARI Collection Schema.

The integrated metadata strategy is only one part of the CENDARI knowledge framework; the metadata is designed to be integrated with future CENDARI ontologies. The final section describes the nature of the integration of the metadata with these future ontologies, and the future impact of the ontologies on the guidelines.

2. Institutional level metadata – EAG (CENDARI)

Encoded Archival Guide (EAG) is a standard for encoding information about archive repositories, and EAG(CENDARI) a version of EAG designed to meet the needs of CENDARI for the encoding of information, and each cultural heritage institution represented in the CENDARI archive directory is represented by and an EAG(CENDARI) record.

EAG, a simple XML schema for encoding fully-interoperable descriptions of collection-holding institutions, was originally devised by the Spanish Ministry of Culture in 2002. It has been developed since then, first by the APEnet project and then by its sequel APEx at the same time as EAG(CENDARI) was being developed. Many of the suggest changes have since been included within the APEx version.

Most of the original amendments made to EAG by CENDARI which were designed to enable it to identify components more precisely than was originally possible. For example, a <person> element was introduced to allow the specific identification of individuals and temporal attributes were added to several elements to allow more precise chronological constraints to be incorporated into searching and browsing facilities. In many cases, these additions were incorporated from other schemas, most notably the Text Encoding Initiative, from which the personal, temporal and bibliographical components were mainly imported.

The most recent EAG (CENDARI) customization and the associated guidelines are available on the CENDARI web site: http://www.cendari.eu/metadata#institution.

The guidelines are designed to both provide an introduction to EAG(CENDARI) that is accessible to both encoders and the less technical readers, and to document the differences between between EAG(CENDARI), and the related standards. (i.e., EAG(Spain), EAG(APEnet), & ISDIAH).

3. Collection level – CENDARI Collection Schema

The CENDARI Collection Schema (CCS) was created to enable detailed information to be encoded about each of the collections housed by the associated cultural heritage institutions. The schema was designed to better meet the requirements of CENDARI users in two ways: extending the standard collection-level description metadata that would be found in encodings such as EAD; and overcoming the semantic limitations of highly descriptive elements.

There are three top-level elements included within CCS that have no direct analogue in EAD:

- **Lacunae** for encoding gaps within the collections.
- **Usage Impediments** for describing impediments users may experience in accessing or utilizing parts of the collection.
- **Collection Future** for making explicit future custodial plans for the collection.

Within EAD this information can only be recorded using <note> elements with a type attribute set to a suitable value. For example, a lacunae in EAD may be expressed as:

```
<note type="lacunae">
            <p>Years 1923-25 are missing as a result of mould
damage</p>
</note>
```

Within CCS the same lacuna could be recorded as:

```
<lacuna lang="en"
            type="missing component"
            typeURI="http://cendari.eu/id/lacunatypes/
missingcomponent"
            cause="mould"
            causeURI="http://cendari.eu/id/lacunacauses/mould"
            coverageID="cendari-sample-1-component1">
            <p>Years 1923-25 are missing as a result of mould
damage</p>
</lacuna>
```

CCS allows for an extensive range of attributes to be included so that a high level of semantic integration may be achieved with the CENDARI ontologies.

As CCS is more extensive than existing standards, this also means that more time is necessary for the encoding of the information according to the CENDARI Collection schema that the more established schemas. This is especially the case where the information is also ready available in an existing standard. Whereas certain established schemas have also had additional software built around them to aid in the creation of metadata records (e.g., ICA-AtoM for EAD).

When new metadata records are created it is recommended that they are created according to CSS, however a crosswalk is available to and from the popular EAD format. Existing or newly created records in EAD may be transformed into CCS records for enhancement with the additional details. Metadata encoded in this schema can readily generate EAD using an XSLT (eXtensible Stylesheet Language) transformation and so allow interoperability with pre-existing systems, although it would inevitably lose some of its granularity in the broader semantic space of most EAD elements.

The most recent version of CCS and the associated guidelines are available on the CENDARI web site: http://www.cendari.eu/metadata#collection.

The importance of a 'collection' can also vary considerably between research areas, and the guidelines provide recommendations for use within the two CENDARI research domains.

4. Item level metadata – MODS with TEI and CCS extension

The item-level description makes use of the pre-existing metadata standards: MODS. This was identified as being suitable for CENDARI requirements as it is both generic enough to be suitable for the wide range of resource that may be incorporated within CENDARI, but can also be extended where a high level of detail is necessary.

Within CENDARI there are distinct differences in the requirements of the two research communities. Whereas the collection level is of primary importance to the World War 1 community of scholars, for the medievalists the item level is the primary focus of both research and archival documentation. In most instances the collection level will be sufficient for researchers into World War 1, but where item level records are required, MODS will be adequate for the wide range of resources that might appear, from bibliographic materials to video, images, and audio. In comparison as the item-level is the focus of medieval research it is important that the resources can be described to the fullest extent possible.

The standard MODS element set is too generic for the more specific, particularly codicological, features necessary for many items in the medieval domain. It is not capable, for instance, of recording such important features as provenance information, detailed physical descriptions, information on scripts or information on incipits, explicts and colophons. To alleviate problems of this type, MODS allows 'extension' elements to be embedded within its architecture: any XML marked-up data may be incorporated into a MODS record by inclusion in its <extension> element.

For the CENDARI project, a number of elements from the TEI are used as extensions in this way, these are taken from the subset of TEI designed for encoding detailed descriptions of manuscripts, particularly codicological and physical information. The item level also makes use of elements from CCS to include the potential for lacunae and an associated bibliography.

The most recent version of CCS and the associated guidelines are available on the CENDARI web site: http://www.cendari.eu/metadata#item.

The guidelines are particularly important at the item level due to the potential for redundancy between the element sets of MODS and TEI.

5. Integration with CENDARI ontologies

The ontologies are an integral part of the knowledge framework and are designed to integrate with the metadata records to facilitate the discovery and connecting of resources that may be dispersed amongst a wide range of cultural heritage institutions. There are a wide range of ontologies and controlled vocabularies that will be used; from small controlled vocabularies limited to one particular attribute (e.g., listing the causes for lacunae being missing) to extensive ontologies incorporating thousands of terms from multiple sources (e.g., CENDARI ontology of concepts incorporating LCSH, DDC).

One of the challenges of CENDARI is to enable the creation of CENDARI metadata at the same time as the ontologies are being developed. As such version 1.0 of the metadata guidelines includes a number of recommendations for the use of vocabularies that may be changed in future versions.

Document 2

EAG (CENDARI) Institutional Schema Guidelines

EAG (CENDARI) Institutional Schema Guidelines

Version 2.1 (05/02/2014)

Encoded Archival Guide (EAG) is a standard for encoding information about archive repositories, and this document describes EAG(CENDARI) a version of EAG designed to meet the needs of CENDARI. The original version of EAG was created by the Spanish Ministry of Culture in 2002 and is referred to within this document as EAG 0.2, although the documentation that is referred to is that created by APEnet[1]. During the period of developing EAG(CENDARI) APEx (the sequel to the APEnet project) also released a new version of EAG, referred to in this document as EAG 2012[2]. The International Council on Archives also released a related standard in 2008, the International Standard for Describing Institutions with Archival Holdings (ISDIAH). Rather than an encoding ISDIAH describes the information that needs to be encoded.

There are two aims for this documentation:
1) Provide an introduction to EAG(CENDARI) that is designed to be accessible to both encoders and the less technical readers.
2) Document the differences between EAG(CENDARI), and the related standards. (i.e., EAG 0.2, EAG 2012, & ISDIAH).

1. General structure of an EAG document

The EAG information model is intended to represent both the description of an institution with archival holdings and the metadata attached to such a description. This is reflected in the two main parts of an <EAG> root element, namely <eagheader> and <archguide>.

ex. 1.a

```
<eag xmlns="http://www.archivesportaleurope.eu/profiles/APEnet_EAG/"
audience="external">
      <control>
            …
      </control>
      <archguide>
            …
```

[1] http://www.apex-project.eu/index.php/outcomes/standards

[2] http://www.apenet.eu/index.php?option=com_content&view=article&id=50&Itemid=65#EAG

```
        </archguide>
        <relations>
                ...
        </relations>
</eag>
```

The root element <eag> owns two mandatory attributes.
The attribute @xmnls provides the linking to the EAG namespace location. The default value of @xmnls is www.archivesportaleurope.eu/profiles/APEnet_EAG/
The attribute @audience informs about the external or internal use of the document. Therefore the two possible values of @audience are external and internal. By default use the value external.

2. Control Informations (<control>)

The control element records all relevant information to identify the EAG document, names all relevant international standards, the language(s) used within the document, and the person(s) responsible for its maintenance.
It is organized as a series of subcomponents:

- <recordId> provides a unique identifier for the EAG document / archival description.
- <maintenanceAgency> describes the agency responsible for the maintenance of the EAG document.
- <maintenanceStatus> indicates the current maintenance state of the EAG document.
- <maintenanceHistory> gathers information about the maintenance history of the document and is organized as a series of maintenance events.
- <languageDeclarations> records the languages used in the document and is organized as a series of language declaration(s).
- <conventionDeclaration> lists all conventions and standards used within the EAG document.
- <localControl> indicates whether the description applies a minimal, partial or a full level of detail.
- <sources> lists the sources of evidence used in creating the archival description.

ex. 2.a

```
<control>
        <recordId>...</recordId>
        <maintenanceAgency>
                ...
        </maintenanceAgency>
        <maintenanceStatus>...</maintenanceStatus>
        <maintenanceHistory>
                ...
```

```
      </maintenanceHistory>
      <languageDeclarations>
            ...
      </languageDeclarations>
      <conventionDeclaration>
            ...
      </conventionDeclaration>
      <localControl>
            ...
      </localControl>
      <sources>
            ...
      </sources>
</control>
```

2.1 EAG identification number (<recordId>)

<recordId> records the identification number of the EAG document / archive description. It is automatically generated when a new EAG document is created on the CENDARI server. This should not be mixed up with any specific identifier attached to the archive itself. See <repositorid> below.

2.2 Maintenance agency (<maintenanceAgency>)

<maintenanceAgency> describes the agency responsible for the maintenance of the EAG document. It contains two pieces of information, namely the agency name and the agency code. In Cendari's case, the default content of <agencyName> is Cendari and the default content of <agencyCode> is EU-CENDARI.

ex. 2.2.a

```
<maintenanceAgency>
      <agencyCode>EU-CENDARI</agencyCode>
      <agencyName>Cendari</agencyName>
</maintenanceAgency>
```

2.3 Maintenance status (<maintenanceStatus>)

<maintenanceStatus> indicates the editing status of the document. There are eight possible values corresponding to the status of the document, namely new, cancelled, deleted, deletedReplaced, deletedMerged, deletedSplit, derived or revised.
By default use the value new.

2.4 Maintenance history (<maintenanceHistory>)

<maintenanceHistory> gathers information about the maintenance history of the archival description and consists of a series of <maintenanceEvent> (maintenance event) that record every change made, the date, the person responsible for the changes and the type of event. There are as many maintenance events as required.

ex. 2.4.a

```
<maintenanceHistory>
      <maintenanceEvent>
            <agent>
                  <persName>
                        <forename>Anna</forename>
                        <surname>Bohn</surname>
                  </persName>
                  <affiliation>
                        <orgName type="institution">FUB</orgName>
                  </affiliation>
            </agent>
            <agentType>human</agentType>
            <date normal="2012-11-14"/>
            <eventType>created</eventType>
      </maintenanceEvent>
</mainhist>
```

There are three types of maintenance events: creation, update, deletion. One must provide a date, a person responsible for the update and the types of event and agent.
There are two values for the agent's type (human or machine in case the document has been automatically created) and seven for the event's type (cancelled, created, deleted, derived, published, revised or updated).

Please note that the date is not given in the text of the element but inside the @normal attribute.

Cendari customization: to encode more precisely information about a person, the <agent> element has been customised mirroring the TEI <person> element and its <persName> and <affiliation> children elements were introduced.
To simplify the encoding, it was decided to indicate just the institution the person belongs to and not the full description of his/her function. The granularity of the data is variable: if needed you can also indicate a department or even a laboratory (using the @type attribute). If the full description of the function needs to be encoded, there exists an <occupation> element.

2.5 Language declaration (<languageDeclarations>)

<languageDeclarations> lists the language(s) used in the archival description. It includes as many <languageDeclaration> elements as there are languages used in the document. Each <languageDeclaration> is divided in a <language> and a <script> to provide precise information on the language.

ex. 2.5.a

```
<languageDeclarations>
      <languageDeclaration>
            <language languageCode="en">English</language>
            <script scriptCode="Latn">Latin</script>
      </languageDeclaration>
      <languageDeclaration>
            <language languageCode="ru">Russian</language>
            <script scriptCode="Cyrl">Cyrillic</script>
      </languageDeclaration>
</languageDeclarations>
```

The @languageCode attribute displays the language code according to ISO 639-1 (see Annex 3) while the @scriptCode attribute displays the script code according to ISO 15924 (see Annex 2). Those two attributes are mandatory when the <language> and <script> elements occur.

2.6 Convention declaration (<conventionDeclaration>)

<conventionDeclaration> lists the standards used within the document. In a <abbreviation> subelement the code of the standard is recorded and its full name goes in a <citation> element.

ex. 2.6.a

```
<conventionDeclaration>
      <abbreviation>ISDIAH</abbreviation>
      <citation>International Standard for Describing Institutions with
Archival Holdings</citation>
</conventionDeclaration>
<conventionDeclaration>
      <abbreviation>EAG</abbreviation>
      <citation>EAG (Encoding Archival Guide) 2012</citation>
</conventionDeclaration>
<conventionDeclaration>
      <abbreviation>iso3166</abbreviation>
      <citation>Codes for the representation of names of countries and
```

```
their subdivisions - Part 1: Country codes</citation>
</conventionDeclaration>
<conventionDeclaration>
      <abbreviation>iso639-1</abbreviation>
      <citation>Codes for the representation of names of languages — Part
1: Alpha-2 code</citation>
</conventionDeclaration>
<conventionDeclaration>
      <abbreviation>iso15924</abbreviation>
      <citation>Codes for the representation of names of scripts</citation>
</conventionDeclaration>
<conventionDeclaration>
      <abbreviation>iso15511</abbreviation>
      <citation>International Standard Identifier for Libraries and Related
Organisations</citation>
</conventionDeclaration>
<conventionDeclaration>
      <abbreviation>iso8601</abbreviation>
      <citation>Data elements and interchange formats - Information
interchange - Representation of dates and times</citation>
</conventionDeclaration>
<conventionDeclaration>
      <abbreviation>UN/LOCODE</abbreviation>
      <citation>United Nations Code for Trade and Transport
Locations</citation>
</conventionDeclaration>
```

In all EAG documents issued by Cendari, the following standards have been used:

- ISDIAH and EAG are the two standards used to elaborate the archival description.
- The country codes used in the archival description are attributed following ISO 3166 alpha 2 standard. The list of codes is to be found in Annex 1.
- The language codes used in the archival description are attributed following ISO 639-1 standard. A selection of most used codes is to be found in Annex 3.

CENDARI customization: in CENDARI we suggest to remain compliant with the most general recommendations of W3C concerning language values and in particular language codes to be used in @langcode (BCP 47). This will ensure a higher operability between our data and other types of data expressed in other formats in a context where we are likely to be working with a wide variety of such formats in CENDARI. We have therefore chosen to use ISO 639-1 rather than ISO 639-2b.

- The script codes used in the archival description are attributed following ISO 15924 standard. A selection of most used codes is to be found in Annex 2.
- For repository identifier, there is no default standard used by Cendari. You must insert here the standard you used to attribute a code to @repositorycode in <repositorid>. In a context where institutions with archival holdings have several identifiers and we want to record them all, it is recommended to list the broadest number of PIDs an archive may have. There might therefore be several standards

listed: i.a. iso15511 (for ISIL), national standards, Cendari.

- All dates in the archival description must be encoded following ISO 8601 rules. A definition of ISO 8601 standard is to be found in Annex 4.
- The city codes (if used to identify the city in the description part of the document) are attributed following UN/LOCODE standard, which you can find here (http://www.unece.org/cefact/locode/service/location.html).

2.7 Local control (<localControl>)

<localControl> indicates whether the description applies a minimal, partial or a full level of detail.
CENDARI customization: we decided to use the <localControl> element to classify the institutions with archival holdings in two categories (first World War or Medieval manuscripts) depending on the type of content they hold.
The Annex 5, paragraph "Theme" lists the only two possible values of the <term> element: WW1 and MM.

ex. 2.7.a

```
<localControl>
      <term>WW1</term>
</localControl>
```

2.8 Sources (<sources>)

<sources> lists the sources of evidence used in creating the archival description and consists of a series of <source>. Each <source> element contains a <sourceEntry> child, which is the (optional) textual identification of this particular source of evidence.
CENDARI customization: a @xml:id attribute can be attached to the <source> element so that, in the descriptive part of the EAG document, it can be referred to when needed by means of a @source attribute (e.g., indicate that the authorized form of name comes from the one source, while the archive's history was from another).

ex. 2.8.a

```
<eag>
   <control>
      […]
         <sources>
            <source xml:id="source1" href="http://www.warszawa.ap.gov.pl/">
                  <sourceEntry>Webpage of the Polish State
Archives</sourceEntry>
            </source>
         </sources>
   </control>
```

```
   <archguide>
      <identity>
      […]
      </identity>
      <desc>
      […]
         <repositorhist xml:lang="pl" source="#source1">
            <p>Oddział mławski archiwum powołany został 21 lipca 1950 roku
[…]</p>
            <p>Pierwsza zmiana terytorialnego zasięgu działania nastąpiła w
roku 1956, kiedy przyłączono powołany wtedy powiat żuromiński.</p>
         </repositorhist>
         […]
      </desc>
   </archguide>
</eag>
```

3. Archival guide (<archguide>)

The archival guide groups all information used to identify and describe a repository. It is divided into two parts, <identity> and <desc>. The <identity> section provides detailed information to identify the described repository, such as its name(s) and identification number(s). The <desc> section holds describing information about the repository, such as its location, history, contact person, etc.

ex. 3.a

```
<archguide>
      <identity>
            […]
      </identity>
      <desc>
            […]
      </desc>
</archguide>
```

3.1 Identity area (<identity>)

<identity> gathers information to identify the described repository, the different forms of name it has and refers to its identification number.

ex. 3.1.a

```
<identity>
      <repositorid countrycode="RS" type="CENDARI">cndr_xqVC3U7N
</repositorid>
```

```
      <autform xml:lang="rs">Народна библиотека Србије</autform>
      <parform xml:lang="en">National Library of Serbia</parform>
      <repositoryType>national</repositoryType>
</identity>
```

<repositorid> records the repository's identifier.
One of CENDARI customizations was to make the <repositorid> element repeatable and introduce an @type attribute to indicate what standard or identification system is referred to, as many of the institutions described within Cendari don't possess a ISIL identifier (which, according to EAG 2012, is supposed to be recorded here while other identifiers can be recorded in <otherRepositorId>).

The element <repositorid> possesses two attributes. @countrycode is mandatory, @repositorycode has been deprecated in favour of the element content (see below).
The attribute @countrycode records the name of the country the repository is located in according to ISO 3166-1 (see Annex 1 for the list of country codes).
The attribute @repositorycode records the code identifying the repository according to national or international standards. Record minimally the internal Cendari identifier of the repository. When existing, please record its ISIL identifier or its national identifier (according to Archon, etc.)
Cendari customization: Unlike the EAG 2012 schema, <repositorid> is not an empty element and can contain text. It is recommended to use the element content rather than the attribute @repositorycode to record the repository identifier in order to avoid redirection.

<autform> records the authorized form of the repository's name (= official name) in its original language.
It is important to note that the element <autform> is repeatable to allow two (or more) authorized forms of name, when for example a country has several official languages and several official denominations of the repository co-exist (note: the @xml:lang attribute is mandatory).
CENDARI customization: An @source attribute has been introduced in the context of CENDARI to refer to the source from which the authorized form of name originates. When used, it has to refer to a source listed in the header part of the document. Same applies to all sub-elements belonging to <identity> and <archguide>.

<parform> records the parallel form(s) of the repository's name. They "indicate the various forms in which the authorised form of name occurs in other languages or script form(s)" (ISDIAH 5.1.3). In this case, use the @xml:lang attribute to specify the

language.
<parform> is repeatable and can support an @source attribute to specify where the form of the name originates.

<nonpreform> records the non-preferred form(s) of the repository's name. “This could include other forms of the same name, acronyms, other institutional names, or changes of name over time.”(ISDIAH 5.1.4)
CENDARI customization: Temporal attributes (like @when, @notBefore, @from, etc.) have been introduced so that the period of time during which the name was used can be specified. See http://www.tei-c.org/release/doc/tei-p5-doc/en/html//ref-att.datable.w3c.html for further description.

ex. 3.1.b

```
<identity>
      <autform xml:lang="rs">Архив Србије</autform>
      <parform xml:lang="en">National Archives of Serbia</parform>
      <nonpreform notBefore="1898" notAfter="1918">Државна Архива Краљевине
Србије</nonpreform>
      <nonpreform notBefore="1918" notAfter="1945">Државна
Архива</nonpreform>
      <nonpreform notBefore="1945" notAfter="1948">Државна Архива
Федеративне Народне Републике Југославије</nonpreform>
      <nonpreform notBefore="1948" notAfter="1958">Државна Архива Народне
Републике Србије</nonpreform>
      <nonpreform notBefore="1958" notAfter="1969">Државни Архив НР
Србије</nonpreform>
</identity>
```

<repositoryType> describes the type of archival institution. Cendari has chosen to use APEx typology (the list of the 12 different types of institutions with archival holdings can be found in Annex 5). Please note that it is possible to select more than one type where applicable.

3.2 Description area (<desc>)

<desc> gathers informative description about the repository, its location, administrative hierarchy and the services it offers. In case of an archival institution splitted in an HQ and several branches, it is possible to describe the different branches using several <repository> elements.
When several repositories of the same institution are described, please record their name and role (headquarter, branch, interim archive) in the corresponding <repositoryName> and <repositoryRole>.

ex. 3.2.a

```
<desc>
      <repositories>
            <repository>
                  <location>
                        <country>Serbia</country>
                              <municipalityPostalcode>Београд 11000
                        </municipalityPostalcode>
                        <street>Скерлићева 1</street>
                  </location>
                  <telephone>to be completed</telephone>
                  <email href="nbs@nb.rs">nbs@nb.rs</email>
                  <access question="yes"/>
                  <repositorhist xml:lang="en">
                        <p>The establishment of the Library is
                        connected to an historical  event in February
                        1832, when Dimitrije Davidovic, publisher of
                        the first Serbian newspaper, sent a letter
                        about setting up a Library to Prince Milos
                        Obrenovic. The same year in November, Prince
                        Milos ordered the obligatory deposit of
                        copies to the National Library. In the
                        foundation of the Library there was no
                        Library establishing act, no permanent
                        location nor official name. The Library grew
                        up stage-by-stage, from several institutions
                        and private libraries.</p>
                  […]
                  </repositorhist>
            </desc>
```

First there is a group of subelements recording the **contact details** of the archive.

<location> records the location(s) of the archival institution (street, number, postal code and city, state, etc.). It contains an address. It is repeatable to better serve the purpose of ISDIAH 5.2.1: "Indicate any other addresses which can be used to contact the archival institution (for example correspondence, administrative, etc.)."

The element <location> bears an @type attribute permitting to specify what type of address is recorded (postal, visitors, etc.). Two @latitute and @longitude attributes are also available and record the geographical coordinates of the location.

CENDARI customization: the two elements <municipalityPostalcode> and <country> bears a @key attribute used to refer to codes from UN/LOCODE (for city) and from ISO 3166-1 (for country). It permits to avoid ambiguity in city names (in case of homonymy) and country names (in case of multilingualism).
See Annex 1 for the list of country codes and the UN/LOCODE webpage

(http://www.unece.org/cefact/locode/service/location.html) for the list of municipalities (browse by country).

ex. 3.2.b

```
<country key="FR">France</country>
```

<telephone>, **<fax>**, **<email>** and **<webpage>** gather additional contact details, which can be used to contact and/or communicate with the repository.
In <email> and <webpage>, it is recommended to include the email address or URL twice, both as text and as value of the attribute @href, for a better use in browsers.

ex. 3.2.c

```
<email href="nbs@nb.rs">nbs@nb.rs</email>
```

Then comes **<access>**, an element, which provides "information about the conditions, requirements and procedures for access to, and use of institutional services" (ISDIAH 5.4.2).
The value of the attribute @question indicates whether the public has free access to the archive ("yes") or whether there are conditions governing the access ("no").
The subelement <restaccess> can be added and provides detailed information about the "access policies, including any restrictions and/or regulations for the use of materials and facilities. Information about registration, appointments, readers' tickets, letters of introduction, admission fees, etc. [can be recorded here]. Where appropriate, make reference to the relevant legislation." (ISDIAH 5.4.2)
The subelement **<termsOfUse>** can be supplied to incorporate information (and links) to the terms of use, archival law or fee structure.

ex. 3.2.d

```
<access question="yes">
      <restaccess>ID card needed</restaccess>
      <termsofUse
href="http://www.bundesarchiv.de/benutzung/voraussetzungen/
allgemeinehinweise/index.html.en">General information on access and
use</termsOfUse>
</access>
```

A series of elements is then encountered that contributes to a more in-depth description of the repository.

<services> describes the services available to users at the repository. Its child **<searchroom>** provides a description of the search room and facilities available.
Cendari customization: the purpose of Cendari not being to provide an in depth description of the archival institutions, it was decided to keep only elements of direct use for the researcher. That's why only the <contact> element (recording the contact details of the repository officer(s) in charge) is used.

<contact> can consist of an email, a phone number and when needed, the name of the person in charge of the repository. If the phone number and email are generic, it is not necessary to specify the name of the officer.
Cendari customization: to better fit the corresponding ISDIAH field, it was decided to widen the scope of the original EAG element. Not only the repository director but "(all) the officers in charge" (ISDIAH 5.2.4) should be recorded here.
Cendari customization: for this purpose, the <nameEntry> element has been customised mirroring the TEI <person> element and its <persName> and <occupation> children elements were introduced (similarly to <respevent>, see above). It allows to encode more precisely information about a person. If several officers are in charge of the search room, the <nameEntry> element can be repeated as many times as needed.

ex. 3.2.e

```
<services>
      <searchroom>
            <contact>
                  <email>angelika.menne-haritz@...</email>
                  <nameEntry>
                        <persName>
                              <surname>Menne-Haritz</surname>
                              <firstname>Angelika</firstname>
                        <persName>
                        <occupation>Leiterin der Stiftung</part>
                  </nameEntry>
            </contact>
      </searchroom>
</services>
```

<repositorhist> details the repository history and contains one or several <descriptiveNote>, which is structured in a series of paragraphs <p>. “This element may include information on dates of establishment, changes of names, changes of legislative mandates, or of any other sources of authority for the institution with archival holdings.” (ISDIAH 5.3.1)

Cendari customisation: If this type of information is available in several languages, use the @xml:lang attribute on <descriptiveNote> to indicate the language used (do not use it on <p>) and duplicate the element (= use two <descriptiveNote> after another).

ex. 3.2.f

```
<repositorhist>
      <descriptiveNote xml:lang="en" source="#source2">
            <p>The establishment of the Library is connected to an
            historical event...</p>
            <p>During World War I, the Library building and
      collections...</p>
```

```
            […]
      </descriptiveNote>
      <descriptiveNote xml:lang="fr" source="#source2">
            <p>La création de la Bibliothèque est liée à un évènement
            historique...</p>
            <p>Pendant la première guerre mondiale, le bâtiment et les
            collections de la Bibliothèque...</p>
            […]
      </descriptiveNote>
</repositorhist>
```

<holdings> provides an overview of the holdings of the archival institution. Record in this element "a short description of the holdings of the archival institution, describing how and when they were formed." (ISDIAH 5.3.6) The element <holdings> is structured in a series of paragraphs <p>.
Cendari customization: the <holdings> element was introduced by Apex in its revision EAG 2012 and adopted by Cendari. Similarly to <repositorhist>, the attribute @xml:lang is used on <descriptiveNote> (not on <p>), which can be duplicated if the information is available in several languages (= use two <descriptiveNote> after another).

ex. 3.2.g

```
<holdings>
      <descriptiveNote xml:lang="de" source="#source1">
            <p>Das Landesarchiv Berlin verwahrt als zentrales Staatsarchiv
       des
            Landes Berlin die schriftliche Überlieferung von Berliner
       Behörden
            und Einrichtungen. Es ist zuständig für die obersten
            Landesbehörden (Senatskanzlei und Senatsverwaltungen), für die
            Bezirke, die Gerichte und sonstigen öffentlichen Stellen des
            Landes sowie für die Mittel- und Unterbehörden des Bundes, die
            ihren Dienstsitz in Berlin haben. Gleichzeitig nimmt das
            Landesarchiv die Funktion eines Parlamentsarchivs wahr.</p>
       </descriptiveNote>
</holdings>
```

4. Relations (<relations>)

<relations> is a wrapper for grouping one or more relations. Cendari is particularly interested in gathering here the bibliographical information related to the archival institution.

<resourceRelation> describes a resource related to the institution and in Cendari's bibliographical case, provides "a general overview of the published and/or unpublished finding aids and guides prepared by the institution with archival holdings and any other

relevant publications." (ISDIAH 5.3.8)
The attribute @resourceRelationType specifies the relation (its values are *creatorOf*, *subjectOf* or *other* - in case of a bibliography, use *creatorOf*). The attribute @xml:lang can be used to document the language of the document. The attribute @href, as shown in the example below, can be used for linking an online holdings guide.
The bibliographical information (when not structured) is located in a **<descriptiveNote>** (similarly structured as in <repositorhist> and <holdings>).
Cendari customization: to better encode bibliographical elements which are structured, it was decided to allow the **<bibl>** element from the TEI vocabulary and allow annotated plain text. It is therefore possible to record its author (<author>), editor, (<editor>) publisher (<publisher>), publication date (<date>), publication place (<pubPlace>) and title (<title>).
Cendari customization: to better encode bibliographical elements which are structured, it was decided to allow the **<biblItem>** element, which is a container for the MODs element set (http://www.loc.gov/standards/mods/v3/mods-3-0-outline.html) to describe the documents. from the MODS vocabulary. It is used for describing any bibliographic items associated with a collection or component.

ex. 3.2.h

```
<relations>
      <resourceRelation resrourceRelationType="creatorOf" xml:lang="de"
href="http://www.landesarchiv-berlin.de/php- bestand">
            <descriptiveNote>
                  <p>Beständeübersicht</p>
            </descriptiveNote>
      <resourceRelation>
</relation>
```

ex. 3.2.i

```
<relations>
      <resourceRelation resourceRelationType="creatorOf" xml:lang="de">
            <objectXMLWrap><bibl><title>Tektonik des Geheimen Staatsarchivs
            PK.</title> Bearbeitet von <author>Rita Klauschenz</author>,
            <author>Sven Kriese</author> und <author>Mathis
            Leibetseder</author>, herausgegeben von <editor>Jürgen
            Kloosterhuis</editor>. <pubPlace>Berlin</pubPlace>
            <date>2011</date>(Veröffentlichungen aus den Archiven
            Preußischer Kulturbesitz, Arbeitsbericht
            12).</bibl></objectXMLWrap>
      <resourceRelation>
</relation>

<relations>
      <resourceRelation resourceRelationType="creatorOf" xml:lang="de">
            <biblItem>
                  <mods:modsCollection>
                        <mods:mods>
                              <mods:titleInfo>
```

```
                                    <mods:title>Tektonik des Geheimen
                                    Staatsarchivs PK</mods:title>
                              </mods:titleInfo>
                              <mods:originInfo>
                                    <mods:publisher>Archiv Preußischer
                                    Kulturbesitz, Arbeitsbericht
                                    12</mods:publisher>
                                    <mods:place>Berlin</mods:place>
                              </mods:originInfo>
                        </mods:mods>
                  </mods:modsCollection>
            </biblItem>
      </resourceRelation>
</relations>
```

<notes> stands for maintenance notes. They “document additional information relating to the creation of and maintenance of the description.” (ISDIAH 5.6.9)

CENDARI customization: There are a number of elements that have not been included within EAG(CENDARI) subset that were included in the description area of EAG 0.2:
ISDIAH 5.3.4 - Administrative structure.
ISDIAH 5.3.6 - Building(s).
ISDIAH 5.4.1 - Opening Times.
ISDIAH 5.4.3 – Accessibility.
ISDIAH 5.5.1 – Research Services.
ISDIAH 5.5.2 – Reproduction Services.

These elements were not considered core to the CENDARI research environment, and as such it was decided time should not be spent collecting the data. Especially as in some cases (e.g., opening times, research services, and reproduction services) the data may be subject to frequent changes.

ANNEX 1: Country Name - ISO 3166-1-alpha-2 code

AFGHANISTAN	AF
ÅLAND ISLANDS	AX
ALBANIA	AL
ALGERIA	DZ
AMERICAN SAMOA	AS
ANDORRA	AD
ANGOLA	AO
ANGUILLA	AI
ANTARCTICA	AQ
ANTIGUA AND BARBUDA	AG
ARGENTINA	AR
ARMENIA	AM
ARUBA	AW
AUSTRALIA	AU
AUSTRIA	AT
AZERBAIJAN	AZ
BAHAMAS	BS
BAHRAIN	BH
BANGLADESH	BD
BARBADOS	BB
BELARUS	BY
BELGIUM	BE
BELIZE	BZ
BENIN	BJ
BERMUDA	BM
BHUTAN	BT
BOLIVIA, PLURINATIONAL STATE OF	BO
BONAIRE, SINT EUSTATIUS AND SABA	BQ
BOSNIA AND HERZEGOVINA	BA
BOTSWANA	BW
BOUVET ISLAND	BV
BRAZIL	BR
BRITISH INDIAN OCEAN TERRITORY	IO
BRUNEI DARUSSALAM	BN

BULGARIA	BG
BURKINA FASO	BF
BURUNDI	BI
CAMBODIA	KH
CAMEROON	CM
CANADA	CA
CAPE VERDE	CV
CAYMAN ISLANDS	KY
CENTRAL AFRICAN REPUBLIC	CF
CHAD	TD
CHILE	CL
CHINA	CN
CHRISTMAS ISLAND	CX
COCOS (KEELING) ISLANDS	CC
COLOMBIA	CO
COMOROS	KM
CONGO	CG
CONGO, THE DEMOCRATIC REPUBLIC OF THE	CD
COOK ISLANDS	CK
COSTA RICA	CR
CÔTE D'IVOIRE	CI
CROATIA	HR
CUBA	CU
CURAÇAO	CW
CYPRUS	CY
CZECH REPUBLIC	CZ
DENMARK	DK
DJIBOUTI	DJ
DOMINICA	DM
DOMINICAN REPUBLIC	DO
ECUADOR	EC
EGYPT	EG
EL SALVADOR	SV
EQUATORIAL GUINEA	GQ
ERITREA	ER
ESTONIA	EE
ETHIOPIA	ET
FALKLAND ISLANDS (MALVINAS)	FK
FAROE ISLANDS	FO

FIJI	FJ
FINLAND	FI
FRANCE	FR
FRENCH GUIANA	GF
FRENCH POLYNESIA	PF
FRENCH SOUTHERN TERRITORIES	TF
GABON	GA
GAMBIA	GM
GEORGIA	GE
GERMANY	DE
GHANA	GH
GIBRALTAR	GI
GREECE	GR
GREENLAND	GL
GRENADA	GD
GUADELOUPE	GP
GUAM	GU
GUATEMALA	GT
GUERNSEY	GG
GUINEA	GN
GUINEA-BISSAU	GW
GUYANA	GY
HAITI	HT
HEARD ISLAND AND MCDONALD ISLANDS	HM
HOLY SEE (VATICAN CITY STATE)	VA
HONDURAS	HN
HONG KONG	HK
HUNGARY	HU
ICELAND	IS
INDIA	IN
INDONESIA	ID
IRAN, ISLAMIC REPUBLIC OF	IR
IRAQ	IQ
IRELAND	IE
ISLE OF MAN	IM
ISRAEL	IL
ITALY	IT
JAMAICA	JM
JAPAN	JP

JERSEY	JE
JORDAN	JO
KAZAKHSTAN	KZ
KENYA	KE
KIRIBATI	KI
KOREA, DEMOCRATIC PEOPLE'S REPUBLIC OF	KP
KOREA, REPUBLIC OF	KR
KUWAIT	KW
KYRGYZSTAN	KG
LAO PEOPLE'S DEMOCRATIC REPUBLIC	LA
LATVIA	LV
LEBANON	LB
LESOTHO	LS
LIBERIA	LR
LIBYA	LY
LIECHTENSTEIN	LI
LITHUANIA	LT
LUXEMBOURG	LU
MACAO	MO
MACEDONIA, THE FORMER YUGOSLAV REPUBLIC OF	MK
MADAGASCAR	MG
MALAWI	MW
MALAYSIA	MY
MALDIVES	MV
MALI	ML
MALTA	MT
MARSHALL ISLANDS	MH
MARTINIQUE	MQ
MAURITANIA	MR
MAURITIUS	MU
MAYOTTE	YT
MEXICO	MX
MICRONESIA, FEDERATED STATES OF	FM
MOLDOVA, REPUBLIC OF	MD
MONACO	MC
MONGOLIA	MN
MONTENEGRO	ME
MONTSERRAT	MS
MOROCCO	MA

MOZAMBIQUE	MZ
MYANMAR	MM
NAMIBIA	NA
NAURU	NR
NEPAL	NP
NETHERLANDS	NL
NEW CALEDONIA	NC
NEW ZEALAND	NZ
NICARAGUA	NI
NIGER	NE
NIGERIA	NG
NIUE	NU
NORFOLK ISLAND	NF
NORTHERN MARIANA ISLANDS	MP
NORWAY	NO
OMAN	OM
PAKISTAN	PK
PALAU	PW
PALESTINIAN TERRITORY, OCCUPIED	PS
PANAMA	PA
PAPUA NEW GUINEA	PG
PARAGUAY	PY
PERU	PE
PHILIPPINES	PH
PITCAIRN	PN
POLAND	PL
PORTUGAL	PT
PUERTO RICO	PR
QATAR	QA
RÉUNION	RE
ROMANIA	RO
RUSSIAN FEDERATION	RU
RWANDA	RW
SAINT BARTHÉLEMY	BL
SAINT HELENA, ASCENSION AND TRISTAN DA CUNHA	SH
SAINT KITTS AND NEVIS	KN
SAINT LUCIA	LC
SAINT MARTIN (FRENCH PART)	MF
SAINT PIERRE AND MIQUELON	PM

SAINT VINCENT AND THE GRENADINES	VC
SAMOA	WS
SAN MARINO	SM
SAO TOME AND PRINCIPE	ST
SAUDI ARABIA	SA
SENEGAL	SN
SERBIA	RS
SEYCHELLES	SC
SIERRA LEONE	SL
SINGAPORE	SG
SINT MAARTEN (DUTCH PART)	SX
SLOVAKIA	SK
SLOVENIA	SI
SOLOMON ISLANDS	SB
SOMALIA	SO
SOUTH AFRICA	ZA
SOUTH GEORGIA AND THE SOUTH SANDWICH ISLANDS	GS
SOUTH SUDAN	SS
SPAIN	ES
SRI LANKA	LK
SUDAN	SD
SURINAME	SR
SVALBARD AND JAN MAYEN	SJ
SWAZILAND	SZ
SWEDEN	SE
SWITZERLAND	CH
SYRIAN ARAB REPUBLIC	SY
TAIWAN, PROVINCE OF CHINA	TW
TAJIKISTAN	TJ
TANZANIA, UNITED REPUBLIC OF	TZ
THAILAND	TH
TIMOR-LESTE	TL
TOGO	TG
TOKELAU	TK
TONGA	TO
TRINIDAD AND TOBAGO	TT
TUNISIA	TN
TURKEY	TR
TURKMENISTAN	TM

TURKS AND CAICOS ISLANDS	TC
TUVALU	TV
UGANDA	UG
UKRAINE	UA
UNITED ARAB EMIRATES	AE
UNITED KINGDOM	GB
UNITED STATES	US
UNITED STATES MINOR OUTLYING ISLANDS	UM
URUGUAY	UY
UZBEKISTAN	UZ
VANUATU	VU
VENEZUELA, BOLIVARIAN REPUBLIC OF	VE
VIET NAM	VN
VIRGIN ISLANDS, BRITISH	VG
VIRGIN ISLANDS, U.S.	VI
WALLIS AND FUTUNA	WF
WESTERN SAHARA	EH
YEMEN	YE
ZAMBIA	ZM
ZIMBABWE	ZW

ANNEX 2 - (selection of) ISO 15924 script code

Name	Code	Number
Arabic	Arab	160
Cypriot	Cprt	403
Cyrillic	Cyrl	220
Cyrillic (Old Church Slavonic variant)	Cyrs	221
Greek	Grek	200
Han (simplified variant)	Hans	501
Han (traditional variant)	Hant	502
Hebrew	Hebr	125
Japanese (alias for Han + Hiragana + Katakana)	Jpan	413
Korean (alias for Hangul + Han)	Kore	287
Latin (Fraktur variant)	Latf	217
Latin (Gaelic variant)	Latg	216
Latin	Latn	215

In case you need a script code, which is not in this (restricted) list, you can find the complete list of script codes here: http://unicode.org/iso15924/iso15924-codes.html

ANNEX 3 - (selection of) ISO 639-1 language code

Albanian	sq
Arabic	ar
Belarusian	be
Bosnian	bs
Bulgarian	bg
Chinese	zh
Croatian	hr
Czech	cs
Danish	da
Dutch	nl
English	en
Estonian	et
Finnish	fi
French	fr
German	de
Greek	el
Hebrew	he
Hungarian	hu
Irish	ga
Italian	it
Japanese	ja
Korean	ko
Latin	la
Latvian	lv
Luxembourgish	lb
Lithuanian	lt
Macedonian	mk
Maltese	mt
Norwegian	no
Polish	pl
Portugese	pt
Romanian	ro
Russian	ru
Serbian	sr
Slovak	sk
Slovenian	sl

Spanish	es
Swedish	sv
Turkish	tr
Ukrainian	uk
Welsh	cy

In case you need a language code, which is not in this (restricted) list, you can find the complete list of language codes here: http://www.loc.gov/standards/iso639-2/php/English_list.php

ANNEX 4 - A profile of ISO 8601 referenced by the W3C HTML recommendation for date and time formats

(http://www.w3.org/TR/NOTE-datetime)

This document defines a profile of ISO 8601, the International Standard for the representation of dates and times. ISO 8601 describes a large number of date/time formats. To reduce the scope for error and the complexity of software, it is useful to restrict the supported formats to a small number. This profile defines a few date/time formats, likely to satisfy most requirements.

Different standards may need different levels of granularity in the date and time, so this profile defines six levels. Standards that reference this profile should specify one or more of these granularities. If a given standard allows more than one granularity, it should specify the meaning of the dates and times with reduced precision, for example, the result of comparing two dates with different precisions.
The formats are as follows. Exactly the components shown here must be present, with exactly this punctuation. Note that the "T" appears literally in the string, to indicate the beginning of the time element, as specified in ISO 8601.

Year:
YYYY (eg 1997)
Year and month:
YYYY-MM (eg 1997-07)
Complete date:
YYYY-MM-DD (eg 1997-07-16)
Complete date plus hours and minutes:
YYYY-MM-DDThh:mmTZD (eg 1997-07-16T19:20+01:00)
Complete date plus hours, minutes and seconds:
YYYY-MM-DDThh:mm:ssTZD (eg 1997-07-16T19:20:30+01:00)
Complete date plus hours, minutes, seconds and a decimal fraction of a second
YYYY-MM-DDThh:mm:ss.sTZD (eg 1997-07-16T19:20:30.45+01:00)

where:

YYYY = four-digit year
MM = two-digit month (01=January, etc.)
DD = two-digit day of month (01 through 31)
hh = two digits of hour (00 through 23) (am/pm NOT allowed)
mm = two digits of minute (00 through 59)
ss = two digits of second (00 through 59)
s = one or more digits representing a decimal fraction of a second
TZD = time zone designator (Z or +hh:mm or -hh:mm)

ANNEX 5 - CENDARI controlled vocabulary lists

Typology

National archives
Regional archives
County/local authority archives
Municipal archives
Specialised governmental archives
Private persons and family archives
Church and religious archives
Business archives
University and research archives
Media archives
Archives of political parties, of popular/labour movement and other non-governmental organisations, associations, agencies and foundations
Specialised non-governmental archives and archives of other cultural (heritage) institutions.

Theme

WW1
MM

Document 3

CENDARI Collection Schema Guidelines

CENDARI Collection Schema Guidelines

[version 1.0 – 17/12/13]

1. Introduction

The CENDARI Collection Schema (CCS) was created to enable detailed information to be encoded about each of the collections housed by the associated cultural heritage institutions. Within the CENDARI metadata strategy collection is conceptualized as being positioned between the institution and the item. In most cases each collection will be associated with one institution that is responsible for the collection, and each collection record may also be associated with any number of item records providing detailed descriptions of items within the collection.

CCS was designed to better meet the requirements of CENDARI users than existing standards in two ways: extending the standard collection-level description metadata that would be found in encodings such as EAD; and overcoming the semantic limitations of highly descriptive elements.

The schema is written in XML (eXtensible Markup Language), a widely-used standard for encoding and interchanging metadata. It aims to provide a structure to allow the most important components of collection information to be collocated and linked up as necessary. The schema defines 16 top-level components and a mechanisms for linking these together using XML identifiers: in addition every component may be identified by an Universal Resource Identifier (URI) by which it may be linked to external resources (such as the controlled vocabularies and ontologies).

2. Format of this document

The rest of this document consists of two main parts. The first part provides a breakdown of the different components, which is followed by details of the cross walk with the existing collection level metadata schema Encoded Archival Description (EAD).

2.1 Description of Elements

There are 16 top level components within CCS each of which is discussed in turn along with any sub-elements, associated attributes, and guidelines on use. Within CCS only two top

level elements are mandatory: <collectionDescIdentifiers>, which in turn must have at least one <collectionDescIdentifier>; and <recordInformation>, which must have a <recordCreation>. Such a minimum record may act as a place-holder for a collection of associated item records where no collection level records exist, although in most instances it is recommended additional elements are included.

Each element, and associated sub-elements and attributes, in the documentation show whether or not they are mandatory through the minimum and maximum number of instances shown in brackets. These take one of four forms:

- (0/1) –The item is not mandatory, but one instance is allowable.
- (0/*) – The item is not mandatory, and the item may appear multiple times.
- (1/1) – The item is mandatory, and can appear only once.
- (1/*) – The item is mandatory, and it can appear multiple times.

2.2 CENDARI Guidelines

Many of the elements, sub-elements, and attributes, whilst not mandatory are nonetheless recommended for use when creating collection level records for CENDARI, and recommendations are made in the guidelines.

CCS has been developed for two research domains: First World War studies and medieval history. The two communities have very different requirements for a collection level schema due to the relative importance of the collection level to research in the two domains. In broad terms the collection level is of primary importance to the World War 1 community of scholars, whereas for the medievalists the item level is the primary focus of both research and archival documentation. The extensiveness of metadata records should reflect the different user requirements, and it is expected that the collection metadata records aimed at the World War 1 community of scholars will in most instances be more extensive than those aimed at the medievalists. Nevertheless some medieval collection records may require more extensive metadata that some World War 1 collection records, although even extensive metadata records will not necessarily make use of the full potential of CSS. As such the guidelines refer to *CSS Basic* and *CSS Full* records as appropriate.

A *CSS Basic* is a minimal collection level record that is sufficient for the identification of collections that are relevant to their research. A *CSS Full* is a collection level record that makes to the full use of CSS.

2.3 Crosswalks

When new metadata records are created at the collection level for CENDARI it is recommended that they are created according to CSS, however a crosswalk has been provide to and from the existing Encoded Archival Description (EAD) format. Existing or

newly created records in EAD may be transformed into CCS records for enhancement with the additional details, and metadata encoded in CSS can readily generate EAD using the associated XSLT (eXtensible Stylesheet Language) transformation on the CENDARI metadata page (http://www.cendari.eu/metadata#institution). This allows for interoperability with pre-existing systems, although it would inevitably lose some of its granularity in the broader semantic space of most EAD elements.

The final part of the document provides a table outlining the crosswalk from CCS to EAD.

3. Top-level components

The 16 top-level components of a collection-level record are as follows:-

	Component name	Definition	Example
1.	Identifier for the collection description <collectionDescIdentifiers>	An identifier for the collection-level description itself, using any recognised format (eg. URI)	http://cendari.eu/id/collection-description/cendari-sample-1-master
2.	Title for the collection-level description <collectionDescTitles>	A title for the collection-level description itself	Cendari Sample Collection 1 - Master Record
3.	Holding institutions <holdingInstitutions>	Details of the archive or other organisation which hosts or administers the collection	European Imaginary Archive
4.	Date <dates>	Any date associated with the archive	1922-01-01
5.	Lacunae <lacunae>	Details of any material missing from the archive	Years 1923-25 are missing as a result of being eaten by mice
6.	Subject coverage <subjectCoverage>	Subject terms or a prose description of the subject coverage of a collection	Middle Ages This collection is mainly centred on materials covering..
7.	Languages <languages>	The languages present in the collections held by the archive	German
8.	Rights Information <rightsInformation>	Intellectual property information relating to the collection	This collection is open to registered users of the archive
9.	Georgraphic information <geogInformation>	Geographic terms associated with the collection	Germany

	Component name	Definition	Example
10.	Source (provenance information) <sourceInformation>	Information on the provenance of the collection, including events in its history	John Smith donated the collection to the archive in 1922
11.	Contents <contents>	A container for information on the collection as a whole and its components	
12.	Relations <relations>	Any relationship between the collection and other entities (eg institutions)	
13.	Usage impediment <usageImpediments>	Any factor inhibiting use of the collection	Approx. 75% of texts illegibility owing to mice damage
14.	Collection future <collectionFuture>	Information on the likely future availability of the collection, or future plans for it	The collection will be maintained indefinitely at the European Imaginary Archive
15.	Bibliography <bibliography>	A set of bibliographic references to literature related to the collection	
16.	Record information <recordInformation>	Information on the metadata record itself, including details of its creation and changes made to it	

These component names are intentionally generic, but may be rendered more specific in the context of a given collection by the use of four attributes which are available for most elements in the schema (shown with examples which may be used to refine a <title> element:

Attribute name	Definition	Example
type	A human-readable indication of the type of component being described (a more precise specification of the component than the generic term itself)	title
typeURI	A URI for a controlled, regularised form of the type, as defined in a controlled vocabulary	http://purl.org/dc/elements/1.1/title
reg	A regularised form for the name of the unit	Cendari Sample Collection 1

Attribute name	Definition	Example
regURI	A URI for the regularised form for the name of the unit, as defined in name authority files or taxonomies, or by content rule	http://cendari.eu/id/collection/title/sample-1

Two further attributes are available for most elements, the first to establish a link to the parts of the collection to which it applies (for instance the parts of the collection to which a <language> element relate), the second to record the language in which the element content is given:

Attribute name	Definition	Example
coverageID	An XML ID referencing a <collection> or <component1..6> element in <contents>	ID of the <collection> or <component1..6> element in the top-level <contents> section to which this element refers
lang	A code (in ISO639-1) format	en

Dependencies

The internal linking mechanisms of CCS require that some elements and attributes are present if others are to function.

The following elements must be present in a CCS record:-

<collectionDescIdentifiers>

<recordInformation>

The following attributes are required:-

<collectionDescIdentifier>: type

<date>: type

<provenance>: role

<contents>/<collection>/<identifier>: type (and <component1>..<component6>

<collection> and <component1>..<component6>: ID

<relation>: type

EAG record and CENDARI profile of record creator should also be created.

3.1 Identifier for the collection description (1/1)

The <collectionDescIdentifiers> (1/1) element is a container for multiple <collectionDescIdentifier> (1/*) elements which record an identifier for the collection-level description itself, using any recognised format (eg. URI).

```
<collectionDescIdentifier
      type="cendari-collection-description"
      typeURI="http://cendari.eu/id/identifier-type/collection-
      description">
      http://cendari.eu/id/collection-description/cendari-sample-1-
mastercendari-sample-1- master
</collectionDescIdentifier>
```

The *type* (1/1) and *typeURI* (0/1) attributes may be used to specify the type of identifier recorded

Identifiers for the collection itself rather than the collection record should be stored in the <identifier> under the <contents><collection> elements (see section 3.11.2).

Guidelines

<collectionDescIdentifiers> is a mandatory element that must contain at least one <collectionDescIdentifier>. A unique CENDARI collection description ID should be created on the ingestion of the record into CENDARI, and a temporary accession number may be used before then. As in the above example, the <collectionDescIdentifier> element for CENDARI should include the two attributes:

```
      type="cendari-collection-description"
      typeURI="http://cendari.eu/id/identifier-type/collection-description"
```

Where alternative identifiers for the collection description are available, for example when a collection description identifier is available from the associated cultural heritage institution, additional <collectionDescIdentifier> elements should be added.

3.2 Title for the collection-level description (0/1)

The <collectionDescTitles> (0/1) element is a container for multiple <collectionDescTitle> (1/*) elements which record a title for the collection-level description itself

```
<collectionDescTitle
```

```
      lang="en"
      type="title"
      typeURI="http://purl.org/dc/elements/1.1/title"
      reg="Cendari Sample Collection 1"
      regURI="http://cendari.eu/id/collection/title/sample-1">
            Cendari Sample Collection 1 - Master Record
</collectionDescTitle>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of title recorded.

The *reg* (0/1) and reg*URI* (0/1) attributes may be used to specify a normalized form of the title, and a URI for this normalized form.

The *lang* (0/1) attribute records the language in which the title is recorded.

Guidelines

The title for the collection-level description is a title for the metadata record, not a title for the collection itself, and is not a recommendation for *CCS Basic.*

In *CCS Full* it is recommended that <collectionDescTitle> should be of type title (http://purl.org/dc/elements/1.1/title).

3.3 Holding institutions (0/1)

The <holdingInstitutions> (0/1) element is a container for multiple <holdingInstitution> (1/*) elements which record details of archive or other organisation which hosts or administers the collection

```
<holdingInstitution
      coverageID="cendari-sample-1-collection"
      type="primary holding institution"
      typeURI="http://cendari.eu/id/holdinginstitutiontype/primary"
      reg="European Imaginary Archive"
      regURI="http://cendari.eu/id/holdinginstitution/europeanimaginaryarch
      ive"
      lang="en"
      institutionDescLink="http://cendari.eu/eag/EIA.xml"
      logo="http://eia.org/logo.gif">
            European Imaginary Archive
</holdingInstitution>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of holding institution recorded.

The *reg* (0/1) and reg*URI* (0/1) attributes may be used to specify a normalized form of the name of the holding institution, and a URI for this normalized form.

The *lang* (0/1) attribute records the language in which the name of the holding institution is recorded.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Two further attributes may be used with this element:-

Attribute name	Definition	Example
institutionDescLink (0/1)	A reference (usually a URL) to an external item describing the institution.	http://cendari.eu/eag/EIA.xml
logo (0/1)	A reference to an image file containing the logo of the holding institution.	http://eia.org/logo.gif

Guidelines

CCS Basic recommends the inclusion of <holdingInstitutions> and the subsequently mandatory <holdingInstitution> if an EAG(CENDARI) record has not already been created that can be linked to through the relations element (see 3.12).

3.4 Dates associated with the archive (0/1)

The <dates> (0/1) element is a container for multiple <date> (1/*) elements which record details of any dates associated with the archive.

```
<dates>
      <date
            typeURI="http://cendari.eu/id/dates/archivecreation"
            type="archive creation"
            reg="1922-01-01"
            coverageID="cendari-sample-1-collection"
            sourceURI="http://cendari.eu/id/sourcetypes/findingaid"
            source="finding aid"
            certaintyURI="http://cendari.eu/id/dates/certainty/certain"
            certainty="certain"
            calendar="gregorian"
      />
```

</dates>

The *type* (1/1) and *typeURI* (0/1) attributes may be used to specify the type of date recorded.

The *reg* (0/1) attributes may be used to specify a normalized form of the date, this will follow the profile of ISO 8601 defined by W3C (http://www.w3.org/TR/NOTE-datetime). The original form of the date, and the calendar, are kept where appropriate to guard against incorrect conversion of dates.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Five additional attributes may also be used with the element:

Attribute name	**Definition**	**Example**
source (0/1)		finding aid
sourceURI (0/1)		http://cendari.eu/id/sourcetypes/findingaid
certainty (0/1)	A human-readable indication of the certainty associated with the date.	certain
certaintyURI (0/1)	A URI for a controlled, regularised form of the certainty, as defined in a controlled vocabulary	http://cendari.eu/id/dates/certainty/certain
calendar (0/1) otherCalendar (0/1)	Calendar type: either gregorian, julian, or otherCalendar If @calendar is set to otherCalendar then any alternative may be used.	gregorian zoroastrian

Dates associated the coverage of materials within the archive, as opposed to dates associated with the archive itself, should be included in <content><collection><contentDesc> (see 3.11.5).

Guidelines

Recommended dates included *CCS Full* include types: archive creation, additions to archives, major changes (eg. discard of materials), relocation of the archive contents

3.5 Lacunae (0/1)

The <lacunae> (0/1) element is a container for multiple <lacuna> (1/*) elements which record details of any items missing from the archive.

```
<lacunae>
      <lacuna lang="en"
      type="missing component"
      typeURI="http://cendari.edu/id/lacunatypes/missingcomponent"
      cause="mice"
      causeURI="http://cendari/edu/id/lacunacauses/mice"
      coverageID="cendari-sample-1-component1"
      startDate="1923-02-02"
      endDate="1924-12-12"
      calendar="gregorian">
            <p>Years 1923-25 are missing as a result of being eaten by
      mi9ce</p>
      </lacuna>
</lacunae>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the nature of the lacuna.

The *lang* (0/1) attribute records the language in which the nature of the lacuna is recorded.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Four additional attributes may also be used with the element:

Attribute name	Definition	Example
cause (0/1)	A human-readable indication of the cause associated with the lacuna.	mice
causeURL (0/1)	A URI for a controlled, regularised form of the cause, as defined in a controlled vocabulary.	http://cendari/edu/id/lacunacauses/mice
startDate (0/1)	The start date of the period covered by the lacuna.	1923-02-02
endDate (0/1)	The end date of the period covered by the lacuna.	1924-12-12
calendar (0/1)	Calendar type as defined in a controlled vocabulary.	gregorian

Guidelines

A controlled vocabulary for lacunae causes is currently under-development.

3.6 Subject coverage (0/1)

The <subjectCoverage> (0/1) element is a container for multiple <subjectTerm> (0/*) and <subjectDesc> (0/*) elements which record details of the subject of the collection.

```
<subjectCoverage>
      <subjectTerm
            lang="en"
            regURI="http://id.loc.gov/authorities/subjects/sh85085001"
            reg="Middle Ages"
            coverageID="cendari-sample-1-collection"
            schemeURI="http://id.loc.gov/authorities/subjects"
            scheme="lcsh">
            Middle Ages
      </subjectTerm>
      <subjectDesc
            typeURI="http://cendari.eu/id/subjectdescription/collectionsubj
      ectcoverage"
            type="collection subject coverage"
            coverageID="cendari-sample-1-collection">
            <p>This collection is mainly centred on materials
      covering...</p>
      </subjectDesc>
</subjectCoverage>
```

3.6.1 Subject Terms (0/*)

For each <subjectTerm> (0/*) element:

The *reg* (0/1) and reg*URI* (0/1) attributes may be used to specify a normalized form of the subject term, and a URI for this normalized form.

The *lang* (0/1) attribute records the language in which the subject term is recorded.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Two additional attributes may also be used with the element:

Attribute name	Definition	Example
scheme (0/1)	A human-readable name of the scheme the subject term is taken from.	lcsh
schemeURL (0/1)	A URI for the subject scheme.	http://id.loc.gov/authorities/subjects

Guidelines

<subjectTerm>s are recommended for both *CCS Basic* and *CCS Full.* Terms may be included from any appropriate scheme, although preference should be given to the CENDARI ontology.

3.6.2 Subject Description (0/*)

For each <subjectDesc> (0/*) element:

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of collection creator recorded.

The *lang* (0/1) attribute records the language in which the subject term is recorded.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Guidelines

<subjectDesc> is primarily for encoding information that already exists in a descriptive form, rather than a reiteration of the more semantically rich <subjectTerm>s. However <subjectDesc> may also be used to elaborate on the use of particular subject terms.

3.7 Languages (0/1)

The <languages> (0/1) element is a container for multiple <language> (1/*) elements which record the language and scripts of the collection and its parts.

```
<languages>
      <language
            coverageID="cendari-sample-1-collection"
            regURI="http://id.loc.gov/vocabulary/iso639-1/de"
```

```
            reg="de"
            coverageID="cendari-sample-1-collection"
            schemeURI="http://id.loc.gov/vocabulary/iso639-1"
            scheme="ISO-639"
            scriptSchemeURI="http://standards.iso.org/iso/15294-2004"
            scriptScheme="ISO 15924:2004"
            scriptURI="http://lexvo.org/id/script/Latn"
            script="Latn"
            sourcePerCent="100"
            findingAidPerCet="100">
                  German
      </language>
      <language
            coverageID="cendari-sample-1-component1"
            regURI="http://id.loc.gov/vocabulary/iso639-1/en"
            reg="de"
            coverageID="cendari-sample-1-component1"
            schemeURI="http://id.loc.gov/vocabulary/iso639-1"
            scheme="ISO-639"
            scriptSchemeURI="http://standards.iso.org/iso/15294-2004"
            scriptScheme="ISO 15924:2004"
            scriptURI="http://lexvo.org/id/script/Latn"
            script="Latn"
            sourcePerCent="25">
                  English
      </language>
</languages>
```

The *reg* (0/1) and reg*URI* (0/1) attributes may be used to specify a normalized form of the language of the collection and any associated finding aids, and a URI for this normalized form.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Eight additional attributes may also be used with the element:

Attribute name	**Definition**	**Example**
scheme (0/1)	Specifies the scheme that is being used to state the language. By default ISO-639-1.	ISO-639
schemeURI (0/1)	A URI for the language scheme that is being used.	http://id.loc.gov/vocabulary/iso639-1
scriptScheme (0/1)	Human-readable name of the	ISO 15924:2004

Attribute name	Definition	Example
	scheme the script is taken from.	
scriptSchemeURI (0/1)	A URI for the script scheme.	http://standards.iso.org/iso/15294-2004
script (0/1)	Specifies the script that is being used.	Latn
scriptURI (0/1)	A URI for the script	http://lexvo.org/id/script/Latn
findingAidPercent (0/1)	Estimated percentage of finding aid that is in the stated language and script.	100
sourcePercent (0/1)	Estimated percentage of the source that is in the stated language and script.	75

Guidelines

<language> elements are recommend for both *CCS Basic* and *CCS Full*. ISO 639-1 is recommended for languages (http://id.loc.gov/vocabulary/iso639-1) and ISO 15294 for scripts (http://www.unicode.org/iso15924/codelists.html).

3.8 Rights information (0/1)

The <rightsInformation> (0/1) element is a container for multiple <rightsStatement> (1/*) elements which record details of the rights information associated with the collection.

```
<rightsInformation>
      <rightsStatement
            typeURI="http://cendari.edu/id/rightsdecltypes/copyrightdeclati
      on"
            type="copryright declaration"
            coverageID="cendari-sample-1-collection">
            <RightsDeclarationMD
                        xmlns="http://cosimo.stanford.edu/sdr/metsrights/">
                  <RightsDeclaration>
                        Collection is open to registered users of the
                  archive
                  </RightsDeclaration>
                  <RightsHolder>
                        <RightsHolderName>
                              European Imaginary Archive
                        </RightsHolderName>
                  </RightsHolder>
```

```
            </RightsDeclarationMD>
      </rightsStatement>
</rightsInformation>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of holding institution recorded.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

The <rightsDeclarationMD> element was taken up from the METS Rights schema (http://www.loc.gov/standards/rights/METSRights.xsd).

Guidelines

It is recommended that the <RightsDeclaration> and <RightsHolder> elements of METSRights are included as a minimum.

3.9 Geographic information (0/1)

The <geogInformation> (0/1) element is a container for multiple <geogTerm> (1/*) elements which record details of geographic locations associated with the collection.

```
<geogInformation>
      <geogTerm
            lang="en"
            regURI="http://id.loc.gov/vocabulary/countries/gw"
            reg="Germany"
            coverageID="cendari-sample-1-collection"
            schemeURI="http://id.loc.gov/vocabulary/countries"
            scheme="MARC countries"
            termValidEnd="9999-12-31"
            termValidStart="1871-01-18"
            otherCalendar=""
            modernEquivalentURI="http://id.loc.gov/vocabulary/countries/gw"
            modernEquivalent="Germany"
            dateCalendar="gregorian">
                  Germany
      </geogTerm>
</geogInformation>
```

The *reg* (0/1) and reg*URI* (0/1) attributes may be used to specify a normalized form of the name of the institution, and a URI for this normalized form.

The *lang* (0/1) attribute records the language in which the name of the institution is recorded.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Eight additional attributes may also be used with the element:

Attribute name	**Definition**	**Example**
scheme (0/1)	Specifies the scheme that is being used to state the associated country.	MARC countries
schemeURI (0/1)	A URI for the country scheme that is being used.	http://id.loc.gov/vocabulary/countries
termValidStart (0/1)	The historic date from which the term is valid.	1871-01-18
termValidEnd (0/1)	The date at which the term is no longer valid.	9999-12-31
modernEquivalent (0/1)	Human readable form of the modern equivalent of the associated country.	Germany
modernEquivalent URI (0/1)	A URI for the modern equivalent of the associated country.	http://id.loc.gov/vocabulary/countries/gw
calendar (0/1) otherCalendar (0/1)	Calendar in which the dates are given. If @calendar is set to otherCalendar then any alternative may be used.	gregorian zoroastrian

Guidelines

<geogInformation> and <geogTerm>s are recommended for *CCS Basic* as well as *CCS Full*. Preference should be given to those schemes identified as appropriate within the CENDARI ontologies, e.g., http://www.getty.edu/research/tools/vocabularies/tgn/.

3.10 Source Information (0/1)

The <sourceInformation> (0/1) element is a container for multiple <provenance> (1/*) elements which record details of the historical provenance of the collection.

```
<sourceInformation>
      <provenance
            lang="en"
            coverageID="cendari-sample-1-collection"
            dateCalendar="gregorian"
            endDate="9999-12-31"
            startDate="1922-01-01"
            agentURI="http://lccn.loc.gov/n97062016"
            agent="Smith, John 1883-1964"
            role="collection donor"
            roleURI="http://cendari.eu/id/collectioncreatortype/collectiond
      onor"
            digitalOriginal="original">
                  John Smith donated the collection to the archive in 1922
      </provenance>
</sourceInformation>
```

The *lang* (0/1) attribute records the language in which the historical provenance information is recorded.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Eight additional attributes may also be used with the element:

Attribute name	Definition	Example
dateCalendar (0/1)	Calendar in which the startDate and endDate are given.	gregorian
startDate (0/1)	The start date of the period covered by the collection.	1922-01-01

Attribute name	Definition	Example
endDate (0/1)	The end date of the period covered by the collection.	9999-12-31
Agent (0/1)	Human readable form of an agent associated with the historical provenance of the collection.	Smith, John 1883-1964
agentURI (0/1)	A URI for the agent.	http://lccn.loc.gov/n97062016
role (1/1)	Human readable form of the role of an agent in the historical provenance of a collection.	collection donor
roleURI (0/1)	A URI for the role.	http://cendari.eu/id/collectioncreatortypes/collectiondonor

startDate and endDate should be in a normalized form, following the profile of ISO 8601 defined by W3C (http://www.w3.org/TR/NOTE-datetime).

Guidelines

<sourceInformation> and the associated <provenance> elements are recommended for *CCS Basic* as well as *CCS Full,* due to the strong association between the provenance of an archive and the topic of the archive.

Agents and agentURI should be taken from the CENDARI ontology or VIAF (http://viaf.org/).

A controlled vocabulary for roles is currently under development, which will be based upon the MARC relators: http://www.loc.gov/marc/relators/.

3.11 Contents (0/1)

The <contents> (0/1) element is a container element for multiple <collection> (1/1) elements which contain information on a collection as a whole and any components thereof, which may share the same structure as a component element.

```
<contents>
      <collection ID="cendari-sample-1-collection">
            ...
            <component1 ID="cendari-sample-1-component1">
                  <!-- same structure as collection element, with recursive
                  component elements down to component6 -->
            </component1>
      </collection>
</contents>
```

The collection elements and any component parts can have three attributes:

Attribute name	Definition	Example
ID (1/1)	A unique ID referring to the collection as a whole or part of the collection.	cendari-sample-1-collection
level (0/1)	Human readable form of archival level associated with metadata record.	fonds
levelURI (0/1)	A URI for the level.	http://cendari.eu/vocabularies/ISADG/fonds

In addition to allowing recursive components, each <collection> (1/1) element is a container for:

Multiple <title> (0/*) elements which allow for the recording of the title in multiple languages.

An <identifier> (0/*) element which allows for the association with a unique identifier with the collection.

The <physicalLocation> (0/*) is a container element for stating the location of the collection.

A <materialDesc> (0/*) element which is a container element used for describing the physical nature of the collection.

A <contentDesc> (0/1) element which is used for describing the content of the collection.

Guidelines

Distinct *ID* attributes are required for each <collection> and <component> so that parts of the collection can be referred to.

It is recommended that the ISAD(G) terms (http://www.icacds.org.uk/eng/ISAD(G).pdf#page=36) are used to describe the *level* of collection that a particular <collection> element or <component> element refers to: fonds, sub-fonds, series, sub-series, file, item.

Detailed item level records should be represented in MODS (with additional extensions) (see http://www.cendari.eu/metadata#item for details) and be linked to within the relations element (see 3.12).

3.11.1 Title elements (0/1)

Multiple <title> (0/*) elements which allow for the recording of the title in multiple languages.

```
<title
      lang="en"
      typeURI="http://purl.org/dc/elements/1.1/title"
      type="title"
      regURI="http://cendari.eu/id/collection/title/sample-1"
      reg="Cendari Sample Collection 1">
            Cendari Sample Collection 1
</title>
<title
      lang="fr"
      typeURI="http://purl.org/dc/elements/1.1/title"
      type="title"
      regURI="http://cendari.eu/id/collection/title/sample-1"
      reg="Cendari Sample Collection 1">
            Collection d'Echantillons Cendari 1
</title>
<title
      lang="de"
      typeURI="http://purl.org/dc/elements/1.1/title"
      type="title"
      regURI="http://cendari.eu/id/collection/title/sample-1"
      reg="Cendari Sample Collection 1">
            Cendari Probenentnahme 1
</title>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of title recorded.

The *reg* (0/1) and reg*URI* (0/1) attributes may be used to specify a normalized form of the title, and a URI for this normalized form.

The *lang* (0/1) attribute records the language in which the title is recorded.

Guidelines

In both *CCS Basic and CCS Full* it is recommended that <collectionDescTitle> should be of type title (http://purl.org/dc/elements/1.1/title).

3.11.2 Identifier element (0/*)

An <identifier> element which allows for the association with a unique identifier with the collection or component in addition to the *ID* attribute used to reference the collection within CENDARI>

```
<identifier
      typeURI="http://cendari.eu/id/identifier-type/collection-number"
      type="cendari-collection-number">
            cendari-sample-1
</identifier>
```

The *type* (1/1) and *typeURI* (0/1) attributes may be used to specify the type of title recorded.

CENDARI Guidelines

The <identifier> is recommended for encoding existing external identifiers for parts of a collection.

3.11.3 Physical location element (0/*)

The <physicalLocation> (0/*) is a container element for stating the location of the collection which is structured according to five sub-elements: <name> (0/*), <location> (1/1), <telephone> (0/1), <fax> (0/1), and <email> (0/1).

```
<physicalLocation
      typeURI="http://cendari.eu/id/physicallocation/administration"
      type="administration"
      descURI="http://cendar.eu/descriptions/eag/eia.xml>
      <name>
            …
      </name>
      <location>
            …
      </location>
      <telephone>
            …
      </telephone>
      <fax>
            …
      </fax>
      <email>
            …
      </email>
</physicalLocation>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of physical location recorded.

One additional attribute may also be used with the element:

Attribute name	Definition	Example
desc (0/1)	A URI to the associated eag file.	http://cendar.eu/descriptions/eag/eia.xml

Guidelines

The <physicalLocation> is recommended only when the information differs from the EAG(CENDARI) record for the associated institution.

3.11.3.1 Name (0/*)

Multiple <name> (0/*) elements enable the encoding of the name of the physical location, e.g., the site name.

```
<name
      type="site name"
      regURI="http://cendari.edu/id/sites/main"
      typeURI="http://cendari/edu/id/nametypes/sites"
      reg="European Imaginary Archive"
      lang="en">European Imaginary Archive Main Site
</name>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of name recorded.

The *reg* (0/1) and reg*URI* (0/1) attributes may be used to specify a normalized form of the name, and a URI for this normalized form.

The *lang* (0/1) attribute records the language in which the name is recorded.

3.11.3.2 Location (1/1)

The <location> (1/1) is a container element for the physical address of the collection or component and has not attributes. It has four sub-elements: <country>(0/1), <postalcode>(0/1), <municipality>(0/1), <street>(0/*).

```
<location>
      <country lang="en"
            regURI="http://dbpedia.org/resource/United_Kingdom"
            reg="United_Kingdom">United Kingdom
      </country>
      <postalcode>WC2B 5RL</postalcode>
      <municipality
            lang="en"
            regURI="http://dbpedia.org/resource/London,_Greater_London"
            reg="London,_Greater_London">London
      </municipality>
      <street>26-29 Drury Lane</street>
</location>
```

<location> (1/1) has no associated attributes.

<country> (0/1) is used for encode the country of the location. It has three associated attributes. The *reg* (0/1) and reg*URI* (0/1) attributes may be used to specify a normalized form of the country, and a URI for this normalized form. The *lang* (0/1) attribute records the language in which the name is recorded.

<municipality> (0/1) is used for encode the country of the location. It has three associated attributes. The *reg* (0/1) and reg*URI* (0/1) attributes may be used to specify a normalized form of the country, and a URI for this normalized form. The *lang* (0/1) attribute records the language in which the name is recorded.

<postalcode>(0/1) is used for encoding the postal code or zip code of a location. It has no associated attributes.

<street> (0/*) is used for encoding any additionally address information. It has no associated attributes.

3.11.3.3 Telephone (0/*)

The <telephone> (0/*) element may be used multiple times to encode telephone numbers.

```
<telephone
      type="primary"
      typeURI="http://cendari.edu/id/telephone/primary">+44 7848 9999
</telephone>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of telephone.

3.11.3.4 Fax (0/1)

The <fax> (0/*) element may be used multiple times to encode fax numbers.

```
      <fax
            type="fax"
            typeURI="http://cendari.edu/id/telephone/fax">+44 7848 0000
      </fax>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of fax.

3.11.3.5 Email (0/1)

The <fax> (0/*) element may be used multiple times to encode fax numbers.

```
<email
      typeURI="http://cendari.edu/id/email/primarycontact"
      type="primary contact">acontact@iea.org
</email>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of email recorded.

3. 11.4 Material description element (0/*)

The <materialDesc> (0/*) element is used for describing the physical nature of the collection. It may have multiple <materialType> (0/*) elements for providing greater detail about specific types of material, each of which may contain a <conditionReport> (0/1) element describing the condition of a particular material type and a <desc> (0/1) element for describing a particular set of materials in a collection.

```
<materialDesc
      unitSizeUnits="mm"
      unitSizeUnitURI="http://dbpedia.org/resource/Milllimetre"
      unitSize="250x170"
      unitTypeURI="http://cendari.eu/id/unittype/box"
      unitType="box"
      unitCountUnitsURI="http://cendari.eu/id/unitcountunit/box"
      unitCountUnits ="box"
      unitCount="295"
      perCentDigitised="95"
      extentUnitURI="http://cendari.eu/id/extentunit/metre"
      extentUnits="m"
      extent="250">
            <materialType>...<materialType>
            <materialType>...<materialType>
</materialDesc>
```

Eleven attributes may be used with the <materialDesc> (0/*) element:

Attribute name	Definition	Example
unitSizeUnits	Unit of measurement used to describe the units.	mm
unitSize	Size of units in the collection.	250x170
unitTypeURI	A URI for the type of units in the collection.	http://cendari.eu/id/unittype/box
unitType	Human readable type of units in the collection.	box
unitCountUnitsURI	Number of units in the collection.	http://cendari.eu/id/unitcountunit/box
unitCountUnits		box

Attribute name	Definition	Example
unitCount		295
percentDigitised	Percentage of the collection digitised.	95
extentUnitURI		http://cendari.eu/id/extentunit/metre
extentUnit		m
extent		250

Guidelines

<materialDesc>, <materialType>, <conditionReport>, and <desc> are not a recommendation for *CCS Basic.*

3.11.4.1 Material type element (0/*)

The <materialType> (0/*) element provides greater detail about a specific type of material in a collection, as well a container for a <conditionReport> (0/1) element describing the condition of a particular material type and a <desc> (0/1) element for describing a particular set of materials in a collection.

```
<materialType
      lang="en"
      ID="cendari-sample-1-material1"
      unitSizeUnitURI="http://cendari.eu/id/extentunit/millimetre"
      unitSizeUnits="mm"
      unitSize="250x170"
      unitCountUnitsURI="http://cendari.eu/id/unitcountunit/leaf"
      unitCountUnits="leaf"
      unitCount="3795"
      extentUnitURI="http://cendari.eu/id/extentunit/metre"
      extentUnits="m"
      extent="245"
      materialTypeURI="http://cendari.eu/id/unittype/paper"
      materialType="paper" >
      <conditionReport
            lang="en"
            conditionURI="http://cendari.eu/id/condition/poor"
            condition="poor">
            <p>The paper contents of the archive are generally in poor
      condition, having been eaten by mice</p>
      </conditionReport>
      <desc>
```

```
            <p>The bulk of the collection comprises 245 metres of boxed
      individual paper leaves</p>
      </desc>
</materialType>
<materialType
      lang="en"
      ID="cendari-sample-1-material2"
      unitSizeUnits="mm"
      unitSize="300"
      unitCountUnitsURI="http://cendari.eu/id/unitcountunit/unitaryobject"
      unitCountUnits="unitary object"
      unitCount="12"
      extentUnitURI="http://cendari.eu/id/extentunit/metre"
      extentUnits="m"
      extent="0.3"
      materialTypeURI="http://cendari.eu/id/unittype/gramophonerecording"
      materialType="gramophone recording"
      unitSizeUnitURI="http://cendari.eu/id/extentunit/millimetre">
      <conditionReport
            lang="en"
            conditionURI="http://cendari.eu/id/condition/good"
            condition="good">
                  <p>The gramophone records are in good condition, and are
            playable on suitable equipment.</p>
      </conditionReport>
      <desc>
            <p>12 78rpm gramophone records</p>
      </desc>
</materialType>
```

The *lang* attribute in the <materialType> element records the language in which the description is written.

Twelve additional attributes may be added to the <materialType> element.

Attribute name	Definition	Example
ID		cendari-sample-1-material1
unitSizeUnits	Unit of measurement used to describe the units.	mm
unitSize	Size of material type units.	250x170
materialTypeURI	A URI for the type of material.	http://cendari.eu/id/unittype/paper
materialType	Human readable type of material.	paper
unitCountUnitsURI	A URI for the material type type of units.	http://cendari.eu/id/unitcountunit/leaf

Attribute name	Definition	Example
unitCountUnits	Human readable material type type of units.	leaf
unitCount	Number of units	3795
percentDigitised	Percentage digitised.	95
extentUnitURI		http://cendari.eu/id/extentunit/metre
extentUnit		m
extent		250

The <conditionReport> (0/1) element describes the condition of a material type.

The *lang* attribute in the <conditionReport> element records the language of the free text description.

Two additional attributes may be added to the <conditionReport> element.

Attribute name	Definition	Example
condition	A human-readable term for the condition of the material.	Good.
schemeURL	A URI for the condition for the condition of the material.	http://cendari.eu/id/condition/good

The <desc> (0/1) element allows for the unstructured

Guidelines

It is recommend that the *materialType* and the *materialTypeURI* is selected from the RDA vocabulary for base materials: http://metadataregistry.org/concept/list/vocabulary_id/35.html

3.11.5 Content Description (0/1)

The <contentDesc> (0/1) element is a container element for one or more <contentInfo> (1/*) used for describing the content of the collection.

```
<contentDesc>
      <contentInfo
            lang="en"
            typeURI="http://cendari.eu/id/unittype/mixed"
            type="mixed"
            sourceURI="http:iea.org/findingaid/399302"
            source="IEA:399302"
            dateCalendar="gregorian"
            endDate="1920-01-01"
```

```
            startDate="1918-01-01"
            sourceTypeURI="http://cendari.eu/id/sourcetype/findingaid"
            sourceType="finding aid"
            authorURI="http://cendari.eu/id/findingaidauthor/0001"
            author="Archivist, A.N.">
                  <p>The collection comprises 3795 leafs of private
            correspondance and 12 gramophone records of spoken word
            covering the period 1918-19..</p>
      </contentInfo>
</contentDesc>
```

The *type* (0/1) and *typeURI* (0/1) attributes in the <contentInfo> element may be used to specify the type of content that is included.

The *lang* (0/1) attribute in the <contentInfo> element records the language of the content information.

Nine additional attributes may be added to the <contentInfo> element.

Attribute name	Definition	Example
sourceURI		http:iea.org/findingaid/399302
source		IEA:399302
dateCalendar		gregorian
endDate		1920-01-01
startDate		1918-01-01
sourceTypeURI		http://cendari.eu/id/sourcetype/findingaid
sourceType		finding aid
authorURI		http://cendari.eu/id/findingaidauthor/0001
author		Archivist, A.N.

Guidelines

<contentDesc> and <contentInfo> are recommended for *CCS Basic* and *CCS Full.*

A controlled vocabulary of source types is currently under-development.

3.12 Relations (0/1)

The <relations> (0/1) element is a container element for one or more <relation> (1/*) elements used for describing any relationship between a collection or parts of a collection and an external entity.

```
<relations>
```

```
      <relation
            typeURI="http:/cendari.eu/id/partnerinstitution"
            type="partner institution"
            targetURI="http:/cendari.eu/id/institution/281830"
            target="United States Imaginary Archive"/
            coverageID="cendari-sample-1" >
</relations>
```

The *type* (1/1) and *typeURI* (0/1) attributes may be used to specify the type of relationship recorded.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Two additional attributes may be added to the <relation> element.

Attribute name	Definition	Example
targetURI	A URI representing the related item.	http://cendari.eu/id/institution/281830
target	A human-readable name of the related object	United States Imaginary Archive

Guidelines

A <relations> element, and the subsequent mandatory <relation> elements, is recommended for *CSS Basic* as well as *CSS Full.*

Any type of relationship could theoretically be expressed within the <relation> element. It is particularly useful for referencing individual items within the collection (for instance, those which have CENDARI item-level records). It is recommended that the relationship between the collection and items within the collection are expressed with the Dublin Core term *hasPart* (http://purl.org/dc/terms/hasPart).

A controlled vocabulary of source types is currently under-development.

3.13 Usage impediment (0/1)

The <usageImpediments> (0/1) element is a container element for one or more <usageImpediment> (1/*) elements used for describing any impediments to the usage of the collection.

```
<usageImpediments>
      <usageImpediment
            lang="en"
            typeURI="http://cendari.eu/id/impediment/illegibility"
```

```
            type="illegibility"
            coverageID="cendari-sample-1-collection"
            severity="75">
                  Approx. 75% of texts illegibility owing to mice damage.
            </usageImpediment>
</usageImpediments>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of impediment to usage.

The *lang* (0/1) attribute records the language in which the impediment to usage is recorded.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

One additional attribute may be added to the <usageImpediment> element.

Attribute name	Definition	Example
severity (0/1)	An approximate proportion of the collection or component suffering.	75

Guidelines

A controlled vocabulary of types of impediment is currently under-development.

3.14 Collection Future (0/1)

The <collectionFuture> (0/1) element is used for describing the likely future availability of the collection, or the future plans for the collection.

```
<collectionFuture lang="en"
         startDate="1920-01-01"
         endDate="9999-12-31"
         coverageID="cendari-sample-1-collection">
                  <p>The collection will be maintained indefinitely at the
European Imaginary Archive.</p>
</collectionFuture>
```

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Two additional attributes may also be used with the element:

Attribute name	Definition	Example
startDate	The start date of the period covered by the <collectionFuture> element.	1920-01-01
endDate	The end date of the period covered by the <collectionFuture> element.	9999-12-31

3.15. Bibliography (0/1)

The <bibliography> (0/1) element is a container element for one or more <biblItem> (1/1) elements used for describing any bibliographic items associated with a collection or component.

```
<bibliography>
      <biblItem
            typeURI="http://cendari.edu/id/bibltype/secondaryliterature"
            type="secondary literature"
            coverageID="cendari-sample-1-collection">
            <mods:modsCollection>
                  <mods:mods>
                        <mods:titleInfo>
                              <mods:title>
                                    A guide to Cendari Sample Collection
                              </mods:title>
                        </mods:titleInfo>
                        <mods:originInfo>
                              <mods:publisher>
                                    Imaginary Publishers Inc.
                              </mods:publisher>
                        </mods:originInfo>
                  </mods:mods>
            </mods:modsCollection>
      </biblItem>
</bibliography>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of bibliographic item.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Each <biblItem> is a container for the MODs element set (http://www.loc.gov/standards/mods/v3/mods-3-0-outline.html) to describe the documents.

Guidelines

It is recommended that the MODS <identifier> with attribute type="Open Library" is used to reference the Open Library ID (https://openlibrary.org) if one is available.

3.16 Record information (1/1)

The <recordInformation> (1/1) element is a container element for one <recordCreation> (1/1) element and one or more <recordChanges> (0/*) elements, these elements document the creation of the records and any changes that have been made.

```
<recordInformation>
      <recordCreation>…</recordCreation>
      <recordChanges>…</recordChanges>
</recordInformation>
```

The <recordInformation> element is only one of two mandatory top-level elements.

3.16.1 Record creation (1/1)

The <recordCreation> (1/1) element is for documenting the creation of the record. It is a container for three elements:

<responsibilities> (0/1) contains multiple <responsibility> (1/*) elements providing details about the person(s) with the responsibility for creating the record.

<creationDates> (0/1) contains multiple <creationDate> (1/*) elements providing the date of the creation of the record.

<notes> (0/1) is a container element for multiple <note> (1/*) elements providing additional details about the creation of the record.

```
<recordCreation
      typeURI="http://cendari.eu/id/recordcreation/recordgenerationfromsour
ce"
      type="record generation from source"
      coverageID="cendari-sample-1-collection">
      <responsibilities>
            <responsibility
                  lang="en"
                  typeURI="http://cendari.eu/id/responsibility/recordcreato
            r"
                  type="record creator"
                  regURI="http://cendari.eu/id/findingaidauthor/0001"
```

```
                reg="Archivist, A.N.">
                Record created by A.N. Archivist
            </responsibility>
        </responsibilities>
        <creationDates>
            <creationDate
                typeURI="http://cendari.eu/id/datetype/recordcreation"
                type="record creation"
                reg="2012-04-08"/>
            </creationDates>
        <notes>
            <note>
                <p>This record derived from archival finding aid ref
            399302 </p>
            </note>
        </notes>
</recordCreation>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of record creation.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6>element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Guidelines

<recordCreation> is used for encoding information about the creation of the CCS record for CENDARI, not the creation of any source such a record may be based upon.

3.16.1.1 Responsibilities (0/1)

The <responsibilities> (0/1) element contains multiple <responsibility> elements providing details about the person(s) with the responsibility for creating the record.

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of responsibility.

The *reg* (0/1) and reg*URI* (0/1) attributes may be used to specify a normalized form of the name of the person with the responsibility, and a URI for this normalized form.

The *lang* (0/1) attribute records the language in which the name of the responsibility is recorded.

Guidelines

Anyone creating CCS records for CENDARI should have a CENDARI profile, and the appropriate CENDARI URI should be stored in the regURI.

3.16.1.2 Creation Dates (0/1)

The <creationDates> (0/1) element contains multiple <creationDate> (1/*) elements providing the date of the creation of the record.

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of date.

The *reg* (0/1) attribute is used to specify a normalized form of the creation date.

Guidelines

The date should follow the profile of ISO 8601 defined by W3C (http://www.w3.org/TR/NOTE-datetime).

3.16.1.3 Notes (0/1)

The <notes> (0/1) element is a container element for multiple <note> (1/*) elements providing additional details about the creation of the record. The <notes> and <note> elements have no associated

Guidelines

The note section may be used for storing any additional information about the reason for record creation, or the source for record.

3.16.2 Record changes (0/1)

The <recordChanges> (0/1) element is a container element for multiple <recordChange> (1/*) elements for documenting any changes to a record. It is a container for three elements:

<responsibilities> (0/1) contains multiple <responsibility> (1/*) elements providing details about the person(s) with the responsibility for creating the record.

<changeDates> (0/1) contains multiple <changeDate> (1/*) elements providing the date of the creation of the record.

<notes> (0/1) is a container element for multiple <note> (1/*) elements providing additional details about the creation of the record.

```
<recordChanges>
      <recordChange
            typeURI="http://cendari.eu/id/recordcreation/recordedit"
            type="record edit"
            coverageID="cendari-sample-1-collection">
            <responsibilities>
                  <responsibility
```

```
                        lang="en"
                        typeURI="http://cendari.eu/id/responsibility/record
                  editor"
                        type="record editor"
                        regURI="http://cendari.eu/id/findingaidauthor/0001"
                        reg="Archivist, A.N.">
                              Record edited by A.N. Archivist
                  </responsibility>
            </responsibilities>
            <changeDates>
                  <changeDate

      typeURI="http://cendari.eu/id/datetype/recordupdate"
                        type="record update"
                        reg="2012-04-09"/>
            </changeDates>
            <notes>
                  <note>
                        <p>This record was updated with new data</p>
                  </note>
            </notes>
      </recordChange>
</recordChanges>
```

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of record change recorded.

The *coverageID* (0/1) attribute records the parts of the collection to which this element refers by containing the XML ID of the <component1..6> element which contains information on this part. It may refer to the collection as a whole by recording the ID of the <collection> element in <contents>.

Guidelines

<recordChanges> and <recordChange> are recommended for *CCS Basic* as well as *CCS Full* for any subsequent changes following a record's creation.

3.16.2.1 Responsibilities (0/1)

The <responsibilities> (0/1) element contains multiple <responsibility> (1/*) elements providing details about the person(s) with the responsibility for change.

The *type* (0/1) and *typeURI* (0/1) attributes may be used to specify the type of responsibility.

The *reg* (0/1) and reg*URI* (0/1) attributes may be used to specify a normalized form of the name of the person with the responsibility, and a URI for this normalized form.

The *lang* (0/1) attribute records the language in which the name of the responsibility is recorded.

Guidelines

As with record creator, anyone creating CCS records for CENDARI should have a CENDARI profile, and the appropriate CENDARI URI should be stored in the regURI.

3.16.2.2 Change Dates (0/1)

The <changeDates> (0/1) element contains multiple <changeDate> (1/*) elements providing the date of the changes to the record.

The *type* (0/1) and *typeURI* (1/*) attributes may be used to specify the type of change.

The *reg* (0/1) attribute is used to specify a normalized form of the change date.

Guidelines

The date should follow the profile of ISO 8601 defined by W3C (http://www.w3.org/TR/NOTE-datetime).

16.2.3 Notes (0/1)

The <notes> (0/1) element is a container element for multiple <note> (1/1) elements providing additional details about changes to the record.

Guidelines

The notes field should provide sufficient information for a user to ascertain who is responsible for any particular content in the record.

4. From CSS to EAG

CCS	EAD
<collectionDescIdentifiers>	<eadheader> /<eadid>
<collectionDescTitles>	<eadheader> /<filedesc> <titlestmt> /<titleproper>
<holdingInstitutions>	<archdesc>/<did>/<repository> /<corpname>
<dates>	<archdesc>/<did>/<unitdate>
<lacunae>	<note type="lacunae">/<p>
<subjectCoverage>	<archdesc>/<controlaccess> /<subject> **or** <archdesc>/<scopecontent> /<p>
<languages>	<archdesc>/<did>/<langmaterial>/<language>
<rightsInformation>	<archdesc>/<did>/<accessrestrict>/<p>
<geogInformation>	<archdesc>/<controlaccess> /<geogterm>
<sourceInformation>	<archdesc>/<custodhist>/ <chronlist>/ <chronitem>
<contents>	<archdesc>/ (for <collection> **or** <archdesc>/<dsc> /<c01> etc (for <component1> etc)
<relations>	<archdesc>/<note type="relations">
<usageImpediments>	<archdesc>/<note type="usage impediment">
<collectionFuture>	<archdesc>/<note type="collection future">
<bibliography>	<archdesc>/<bibliography>
<recordInformation>	<eadheader> /<profiledesc> /<creation> **and** <eadheader> /<revisondesc> /<change>

Document 4

CENDARI Item-level Schema Guidelines

CENDARI Item-level Schema Guidelines

[version 1.0 – 07/02/14]

1. Introduction

For item-level descriptions, CENDARI uses the pre-existing metadata schema MODS (Metadata Object Description Schema), supplemented by elements from the TEI P5 Manuscript Description Schema and a small number of additional elements from the CENDARI Collection Schema (CCS). MODS is used as the foundation of CENDARI item-level requirements as it is both generic enough to be suitable for the wide range of resource that may be incorporated within CENDARI, and can also be extended where a higher level of detail is necessary.

Within CENDARI there are distinct differences in the requirements of the two research communities at the item level. Whereas the general descriptive functionality of MODS is adequate for the wide range of resources that might be included by World War 1 scholars, as the item-level is the focus of medieval research it is important that the resources can be described to the fullest extent possible. The standard MODS element set is too generic for the more specific, particularly codicological, features necessary for many items in the medieval domain. It is not capable, for instance, of recording such important features as provenance information, detailed physical descriptions, information on scripts or information on incipits, explicts and colophons. To alleviate problems of this type, MODS allows 'extension' elements to be embedded within its architecture: any XML marked-up data may be incorporated into a MODS record by inclusion in its <extension> element.

Two extensions are recommended for the CENDARI item-level descriptions: an extension for the inclusion of TEI elements for the encoding of detailed descriptions of manuscripts; and an extension for the inclusion of CCS elements, for the potential encoding of lacunae and an associated bibliography.

The rest of this document provides an overview of the three parts of the item-level descriptions in more detail. Firstly, the MODS elements, then the TEI elements, and finally the CCS elements. For a full example see Appendix 1.

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3">
      <!-- MODS elements -->
      <extension xmlns:cen="http://www.tei-c.org/ns/1.0”>
            <!-- TEI elements -->
      </extension>
```

```
      <extension xmlns:cen="http://www.cendari.eu/metadatacendari-
      collection-desc.xsd”>
            <!-- CCS elements -->
      </extension>
</mods>
```

The guidelines are particularly important at the item level due to the potential for redundancy between the element sets of MODS and TEI. For instance, the choice of whether to use the relatively simple physical description components of MODS, or its potentially more complex but more powerful counterpart in TEI, requires the consistent use of cataloguing rules to ensure interoperability between these descriptions.

2. MODS Elements

MODS (Metadata Object Description Schema) is a generic schema devised by the Library of Congress for bibliographic object descriptions. MODS is derived from the MARC (Machine Readable Cataloging) standard which has been used as the basis of interoperable online catalogues for almost 50 years: it is designed specifically, however, to address the needs of both analogue and digital objects, including multiple levels of granularity which are often a feature of the latter.

There are 20 top-level elements in MODS:

<titleInfo> - The name of the resource.
<name> - The name of a person, organization or event associated with the resource.
<typeOfResource> - A term reflecting the type of resource, e.g., still image, text.
<genre> - A term to denote a particular style.
<originInfo> - Information about the origin of a resource, e.g., publisher, place of publication.
<language> - The language of the resource.
<physicalDescription> - Physical attributes of the information resource.
<abstract> - Summary of the content of the resource.
<tableOfContents> - Details of the contents of the resource.
<targetAudience> - Intellectual level of the resource.
<note> - Any information for which there is not another more specific element.
<subject> - The main topic(s) of a work.
<classification> - The main topic of a work according to a formal classification system.
<relatedItem> - Information about related resources.
<identifier> - Unique identifier for the resource.
<location> - The physical or electronic location of the resource.
<accessCondition> - Information about access rights to a resource.
<part> - Details of the physical parts of a resource, e.g., page numbers.

<extension> - Allows for the inclusion of information not covered by MODS.

<recordInfo> - Information about the creation, origins, and changes in the metadata record.

Any of the MODS elements may be part of an item level record, and the full documentation on applying these element is available at http://www.loc.gov/standards/mods/userguide/generalapp.html#top_level. The eight MODS elements that are likely to form the basis of a typical CENDARI item-level document are discussed in briefly below.

2.1 <titleInfo>

<titleInfo> provides information about the item. When the title information has not been transcribed from the item itself, the attribute supplied should be set to "yes". The title itself is given with the <title> element, and this may have a lang attribute indicating the language of the title. The <title> element is repeatable, and a 'type' attribute may be used to identify alternative forms of the title, e.g., type="translated" or type="alternative".

Nb. The primary title of a resource should not use a type attribute.

```
<titleInfo supplied="yes">
      <title lang="en"></title>
</titleInfo>
```

2.2 <name>

The repeatable <name> element records any names associated with the item. The <role>/ <roleTerm> element within <name> can be used to specify the form of the association with the item. It is recommended to use MARC relator codes where possible (see http://www.loc.gov/marc/relators/relaterm.html for full list), in which case the type attribute of <roleTerm> should be set to **code** and the authority attribute set to marcrelation as in the example (if not, set type to **text).**

```
<name>
      <role>
            <roleTerm type="code" authority="marcrelator"></roleTerm>
      </role>
      <namePart></namePart>
</name>
```

Each component of the name is given in a separate <namePart> element, the type attribute of which should be set to one of **family, given, termsOfAddress** or **date.**

2.3 <subject>

The repeatable <subject> element contains subject terms associated with the item. Within the <subject> element use one of <topic> <geographic> <temporal> <titleInfo> <name> <genre> <hierarchicalGeographic> <cartographics> <geographicCode> <occupation> to indicate the type of subject being recorded. The attributes authority, authorityURI, and valueURI may be used to reference the authority scheme, the scheme's URI, and the URI associated with a particular subject term.

```
<subject authority="lcsh"
authorityURI="http://id.loc.gov/authorities/#conceptscheme"
valueURI=" http://id.loc.gov/authorities/sh85148236#concept">
      <topic>World War, 1914-1918</topic>
</subject>
```

Subject terms may be taken from any ontology, although preference should be given to the CENDARI ontology.

2.4 <language>

The repeatable <language> element allows the recording of any languages used in the item. The language should be given as an ISO 639-2 code (http://www.loc.gov/standards/iso639-2/langhome.html) within the <languageTerm> element, with the type and authority attributes set as in the example.

```
<language>
      <languageTerm type= "code" authority="iso639-2”>en</languageTerm>
</language>
```

2.5 <physicalDescription>

The MODS <physicalDescription> element should be used to record the physical medium of the item and its extent, using the <form> and <extent> sub-elements respectively.

```
<physicalDescription>
      <form>parchment</form>
      <extent>220 folios</form>
</physicalDescription>
```

All other features of the physical description should be encoded using the <tei:physDesc> element (see 3.3 below).

2.6 <originInfo>

The <originInfo> element records information on the origin of item, including dates and places associated with its creation.

Five date elements may be used as appropriate: <dateCreated> <dateCaptured> <dateValid> <dateModified> or <dateOther>. It is recommended that dates are encoding according to W3C recommendations (http://www.w3.org/TR/NOTE-datetime).

Place names should be put in the <placeTerm> element within <place>. The attributes authority, authorityURI, and valueURI may be used to reference the authority scheme, the scheme's URI, and the URI associated with a particular subject term.

```
<originInfo>
      <dateCreated endcoding="w3cdtf">1974</dateCreated>
      <place>
            <placeTerm type="code" authority="iso3166">GB</placeTerm>
            <placeTerm type="text">United Kingdom</placeTerm>
      </place>
</originInfo>
```

2.7 <accessCondition>

<accessCondition> is used to record information on restrictions and condition on access to the item. This is a free-text element.

```
<accessCondition></accessCondition>
```

2.8 <relatedItem>

<relatedItem> may be used to provide information on related items, with the type attribute being used to encode the nature of the relationship. Popular uses of the <relatedItem> in CENDARI may express the relationship with digital versions of a resource, type="otherFormat" (as in the example below) or expressing the relationship with the hosting collection, type="host".

The <identifier> element may be used to provide its URL, URI or other identifier; if the identifier is a URI or URL, the type attribute should be set to 'uri'. Within the example below the <physicalDescription> element and <internetMediaType> element are used to record the format of the digital surrogates.

```
<relatedItem type="otherFormat">
      <physicalDescription>
            <internetMediaType>tiff</internetMediaType>
      </physicalDescription>
      <identifier type="uri"></identifier>
</relatedItem>
```

2.9 <note>

Any note may be recorded in the repeatable <note> element. Its type attribute may be set to any meaningful value.

```
<note></note>
```

2.10 <recordInfo>

This element may be used to record information on the creation of the metadata record and changes made to it. It may also be used to record the source of the metadata itself.

```
<recordInfo>
      <recordCreationDate encoding="marc">030211</recordCreationDate>
      <recordContentSource>Finding aid</recordContentSource>
</recordInfo>
```

3. TEI Extension

The TEI extension includes the subset of TEI designed for encoding detailed descriptions of manuscripts, particularly codicological and physical information.

The TEI <msDesc> element, which can contain a description of a single manuscript may be composed of seven further elements.

<msIdentifier> contains any identifiers associated with the manuscript.

<head> contains any associated headings.

<msContents> describes the content of the manuscript.

<physDesc> provides a physical description of the manuscript.

<history> contains the history of the manuscript.

<additional> contains additional bibliographic, surrogate, and administrative information.

<msPart> contains information about a part of the manuscript that was originally distinct.

Within CENDARI <msIdentifier>, <msContents>, <physDesc>, and <history> will be sufficient in most instances. Title information will not be held within the <tei:head> element, but within <mods:titleInfo> (see 2.1), and additional bibliographic information can be encoded in <ccs:bibliography> (see 4.2). Although <tei:msPart> will still be used to describe

parts of the manuscript that were previously separate, this is supplemented by <ccs:lacunae> to describe content is now missing (see 4.1).

```
<extension xmlns:tei="http://www.tei-c.org/ns/1.0”>
      <tei:msDesc>
            <tei:msIdentifier>…</tei:msIdentifier>
            <tei:msContents>…</tei:msContents>
            <tei:physDesc>…</tei:physDesc>
            <tei:history>…</tei:history>
      </tei:msDesc>
```

Each of the four elements are discussed in more briefly below, although further documentation on all the TEI MSDesc elements is available at http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html#msdo.

3.1 <tei:msIdentifier>

The <tei: msIdentifier> element records the city, repository and identification number (such as shelfmark) of the item in its <tei:settlement> <tei:repository> and <tei:idno> sub-elements respectively. A URL for the repository may be given in the ref attribute of <tei:repository>

```
<tei:msIdentifier>
      <tei:settlement></tei:settlement>
      <tei:repository ref="http://repository-url.eu"></tei:repository>
      <tei:idno></tei:idno>
</tei:msIdentifier>
```

3.2 <tei:msContents>

The <tei: msItem> element may be used to record the inciplit, explicit and colophon of the item in its <tei:incipit> <tei:explicit> and <tei:colophon> sub-elements respectively. The <tei:msItem> element is repeatable and so may record multiple components of the same item (in the case of composite items).

```
<tei:msContents>
      <tei:msItem>
            <tei:incipit></tei:incipit>
            <tei:explicit></tei:explicit>
            <tei:colophon></tei:colophon>
      </tei:msItem>
</tei:msContents>
```

3.3 <tei:physDesc>

The physical description of the object can comprise of three types of information: the object description; information on writing styles and notations; and descriptions of bindings and seals. The TEI <physDesc> element should be used for all facets of the physical description, except for its physical medium and its extent (for which the MODS <form> and <extent> sub-elements of <physicalDescription> should be used (see 2.5 above).

3.3.1 <tei:objectDesc>

The <tei:objectDesc> is used to describe the physical nature of the text-bearing part of the object (as opposed to describing the text itself or any additional associated items), and may use the form attribute to record the form of the item (e.g. codex).

There are two parts of the object description: <tei:supportDesc> which groups the elements that physically support writing, and <tei:layout> describes how the layout of the text.

```
<tei:objectDesc form="codex">
      <tei:supportDesc material="paper">
            <tei:extent>55 leaves
                  <tei:dimensions>
                        <tei:height></tei:height>
                        <tei:width></tei:width>
                        <tei:depth></tei:depth>
                  </tei:dimensions>
            </tei:extent>
            <tei:condition>
                  <tei:p></tei:p>
            </tei:condition>
      </tei:supportDesc>
      <tei:layoutDesc>
            <tei:layout columns="2">In double columns</tei:layout>
      </tei:layoutDesc>
</tei:objectDesc>
```

The material attribute of the <tei: supportDesc> element records the material of the item is composed (e.g. paper, vellum). The number of leaves is recorded in the <tei:extent> element as shown. The dimensions of the item are recorded in the <tei:dimensions> element, using its sub-elements <tei:height>, <tei:width> and <tei:depth>. The <tei:condition> element records information on the condition of the item: it contains repeatable <tei:p> elements for paragraphs of the description.

Information on the layout of the item is given in the <tei:layout> element within <tel:layoutDesc>. If arranged in columns, the number is given in the columns attribute.

It is recommended that the form and material attributes make use of terms from the RBMS Controlled Vocabularies (http://www.rbms.info/committees/bibliographic_standards/controlled_vocabularies/).

3.3.2 <tei:musicNotation>, <tei:scriptDesc>, <tei:decoDesc>, <bindingDesc>

The <tei:musicNotation> element records information on musical notation used within the item: it contains repeatable <tei:p> elements for paragraphs of the description.

```
<tei:musicNotation>
      <tei:p></tei:p>
</tei:musicNotation>
```

The <tei:scriptDesc> element records information on the script(s) used within the item: it contains repeatable <tei:p> elements for paragraphs of the description.

```
<tei:scriptDesc>
      <tei:p></tei:p>
</tei:scriptDesc>
```

The <tei:decoDesc> element records information on decorations used within the item: it contains repeatable <tei:p> elements for paragraphs of the description.

```
<tei:decoDesc>
      <tei:p></tei:p>
</tei:decoDesc>
```

The <tei:bindingDesc> element records information on the item's binding(s): it contains repeatable <tei:p> elements for paragraphs of the description.

```
<tei:bindingDesc>
      <tei:p></tei:p>
</tei:bindingDesc>
```

3.4 <tei:history>

The <tei:history> element can contain information about the origin, provenance, and acquisition of a manuscript.

The skeletal example below shows the elements necessary for the encoding of the provenance of a record. The <tei:provenance> elements contains sub-elements detailing previous owners or other persons associated with the item. It contains a <tei:listPerson> element within which are multiple <tei:person> elements, one for each associated with it.

Each <tei:person> element can have a role attribute to indicate whether they were the owner, curator etc of the item. Within the <tei:person> element are multiple <tei:event> elements which contain a when attribute used to record the date itself and multiple <tei:p> elements to record what form of provenance event took place. The name of the person associated with each provenance item is given in the <tei:persName> element.

```
<tei:history>
      <tei:provenance>
            <tei:listPerson>
                  <tei:person role="curator">
                        <tei:event when="1622">
                              <tei:p></tei:p>
                        </tei:event>
                        <tei:persName></tei:persName>
                  </tei:person>
            </tei:listPerson>
      </tei:provenance>
</tei:history>
```

4. CCS Extension

The CCS extension enables the addition of two top-level elements from the CENDARI Collection Schema: <lacunae> and<bibliography>. The <lacunae> element allows the

```
<extension xmlns:css="http://www.cendari.eu/metadatacendari-collection-
desc.xsd">
      <ccs:lacunae>…</ccs:lacunae>
      <cen:bibliography>…</cen:bibliography>
</extension>
```

Brief examples of the use of the <ccs:lacunae> and <ccs:bibliography> are provided below, with further documentation on CCS elements is available at http://www.cendari.eu/metadata#collection.

2.1 <lacunae>

The <ccs: lacunae> element is a container for multiple <ccs:lacuna> elements which record details of any items missing from the archive.

```
<ccs:lacunae>
      <ccs:lacuna lang="en"
            type="missing component"
            typeURI="http://cendari.edu/id/lacunatypes/missingcomponent"
```

```
            causeURI="http://cendari/edu/id/lacunacauses/mice"
            coverageID="cendari-sample-1-component1">
                  <p>Years 1923-25 are missing as eaten by mice</p>
      </ccs:lacuna>
</ccs:lacunae>
```

4.2 <bibliography>

The <ccs:bibliography> element is a container element for one or more <ccs:biblItem> elements used for describing any bibliographic items associated with a collection or component. The type and typeURI attributes may be used to specify the type of bibliographic item. Each <ccs:biblItem> contains a <modCollection> (for multiple entries) or <mods> element, which contains the standard MODS elements for bibliographic entries.

```
<ccs:bibliography>
      <ccs:biblItem
            type="secondary literature"
            typeURI="http://cendari.eu/id/bibltype/secondaryliterature">
            <modsCollection>
                  <mods>
                        <titleInfo>
                              <title>A guide to Cendari</title>
                        </titleInfo>
                        <originInfo>
                              <publisher>Imaginary Publishers</publisher>
                        </originInfo>
                  </mods>
            </modsCollection>
      </cen:biblItem>
</cen:bibliography>
```

Appendix 1.

Skeletal example of a CENDARI item-level description record.

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3"
   xmlns:tei="http://www.tei-c.org/ns/1.0"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/mods.xsd
   http://www.tei-c.org/ns/1.0 http://www.tei-
c.org/release/xml/tei/custom/schema/xsd/tei_ms.xsd">

   <titleInfo supplied="yes">
      <title lang="en"></title>
   </titleInfo>

   <name>
      <role>
         <roleTerm type="code" authority="marcrelator"></roleTerm>
      </role>
      <namePart></namePart>
   </name>

    <subject>
      <topic></topic>
   </subject>

   <language>
      <languageTerm type= "code" authority="iso639-2">en</languageTerm>
   </language>

   <originInfo>
      <dateCreated></dateCreated>
      <place>
         <placeTerm></placeTerm>
      </place>
   </originInfo>

   <accessCondition></accessCondition>
```

```
<relatedItem type="otherFormat">
  <physicalDescription>
    <form>parchment</form>
   <extent>220 folios</extent>
  </physicalDescription>
  <identifier type="uri"></identifier>
</relatedItem>

<note></note>

<extension>
  <tei:msDesc>
    <tei:msIdentifier>
      <tei:settlement></tei:settlement>
      <tei:repository ref="http://repository-url.eu"></tei:repository>
      <tei:idno></tei:idno>
    </tei:msIdentifier>
    <tei:msContents>
      <tei:msItem>
        <tei:incipit></tei:incipit>
        <tei:explicit></tei:explicit>
        <tei:colophon></tei:colophon>
      </tei:msItem>
    </tei:msContents>

    <tei:physDesc>
      <tei:objectDesc form="codex">

        <tei:supportDesc material="paper">

         <tei:extent>55 leaves

            <tei:dimensions>
              <tei:height></tei:height>
              <tei:width></tei:width>
              <tei:depth></tei:depth>
             </tei:dimensions>
          </tei:extent>

          <tei:condition>
            <tei:p></tei:p>
```

```
            </tei:condition>



         </tei:supportDesc>
         <tei:layoutDesc>
            <tei:layout columns="2">In double
               columns</tei:layout>
         </tei:layoutDesc>

      </tei:objectDesc>
      <tei:musicNotation>
         <tei:p></tei:p>
      </tei:musicNotation>
      <tei:scriptDesc>
         <tei:p></tei:p>
      </tei:scriptDesc>
      <tei:decoDesc>
         <tei:p></tei:p>
      </tei:decoDesc>

      <tei:bindingDesc>
         <tei:p></tei:p>
      </tei:bindingDesc>

   </tei:physDesc>
   <tei:history>
      <tei:provenance>
         <tei:listPerson>
            <tei:person role="">
               <tei:person role="curator">
               <tei:event when="1622">
                  <tei:p></tei:p>
               </tei:event>
             <tei:persName></tei:persName>

            </tei:person>
         </tei:listPerson>
      </tei:provenance>
   </tei:history>
```

```
    </tei:msDesc>

 <extension xmlns:cen="file:/home/richard/Dropbox/CENDARI/cendari-collection-desc.xsd">
    <cen:lacunae>
     <cen:lacuna lang="en"
      type="missing component"
      typeURI="http://cendari.edu/id/lacunatypes/missingcomponent"
      cause="mice"
      causeURI="http://cendari/edu/id/lacunacauses/mice"
      coverageID="cendari-sample-1-component1">
     <p>Years 1923-25 are missing as eaten by mice</p>
      </cen:lacuna>
    </cen:lacunae>

   <cen:bibliography>
    <cen:biblItem
     type="secondary literature"
     typeURI="http://cendari.edu/id/bibltype/secondaryliterature">
        <modsCollection>
          <mods>
            <titleInfo>
              <title>A guide to Cendari</title>
            </titleInfo>
            <originInfo>
               <publisher>Imaginary Publishers</publisher>
            </originInfo>
          </mods>
        </modsCollection>
     </cen:biblItem>
    </cen:bibliography>
  </extension>

</mods>
```