

Deliverable 3.1

Metadata Integration Report

DAM-LR

011841

Distributed Access Management for Language Resources

**implemented as
Specific Support Action**

Contract Number: *011841*

Project Coordinator: Peter Wittenburg

Project Web-Site: www.mpi.nl/dam-lr/

Deliverable: D3.1

Authors: Lund, MPI

Responsible: Lund

Date: 1.1.2006

Content

1	INTRODUCTION	3
1.1	CORPUS STRUCTURE	3
1.2	RESOURCE BUNDLES	3
1.3	DATA ENCODINGS.....	3
2	LUND METADATA.....	4
2.1	LUND CORPUS	4
2.2	METADATA UPGRADE.....	4
2.3	FILE TYPES AND FORMATS	5
2.4	LANGUAGE IDENTIFICATION UPGRADE	5
3	INL METADATA	6
3.1	CGN CORPUS	6
3.2	METADATA UPGRADE.....	6
3.3	FILE TYPES AND FORMATS	6
3.4	FUTURE PLANS	7
4	SOAS METADATA.....	8
4.1	INGESTION METADATA	8
4.2	ARCHIVE METADATA.....	8
4.3	DISSEMINATION METADATA	8
5	MPI METADATA.....	10
5.1	MPI CORPUS	10
5.2	DBD CORPUS.....	10
6	REFERENCES	12
	APPENDIX A : IMDI DBD SESSION PROFILE.....	13

1 Introduction

One of the major points of the integration of the different archives is metadata interoperability. Within the DAM-LR project this will be achieved by adapting the existing metadata infrastructures to the [IMDI] metadata concepts and framework. The metadata framework plays a crucial role in the development of the distributed solution and therefore it's important that all repositories at least provide an interface to supply correct IMDI metadata descriptions. The IMDI descriptions are formalised by the IMDI XML Schema [XMLSCHEMA] version 3.0.

Since each partner can have their own local metadata infrastructure it's clear that every archive needs a semantic mapping from existing metadata descriptions and structures to the overall IMDI framework. These processes are described in the following chapters.

1.1 *Corpus Structure*

Most corpora are created in a tree-structure reflecting an initial categorisation of the language resources. These kinds of structures are supported by IMDI in the form of corpus nodes. One corpus node can contain a set of links to other corpus nodes or to a set of resource bundles. Each node has a name and other metadata elements to describe that point in the corpus structure.

1.2 *Resource Bundles*

The concept of resource bundles (or Sessions) is used as a container for all information about the circumstances and conditions of the linguistic event, groups the resources belonging to this linguistic event, records the administrative information of the event and describes the content of the event. An extensive description can be found in the IMDI documentation Metadata Elements for Session Descriptions [SESSIONS].

1.3 *Data encodings*

Although [XML] enables the use of different character encodings it's more practical for tools to use a single encoding within the IMDI framework. It was agreed that all data encodings will be [UNICODE]. This is supported by IMDI by using UTF-8 [RFC2279] to implement the encoding of UNICODE characters.

2 Lund Metadata

At Lund already some part of the archive is described by using early versions of IMDI metadata. The data and metadata of these four sub-corpora are analysed and made available for integration into the distributed IMDI domain.

2.1 Lund Corpus

The following archives are involved:

- Phonogrammarchiv Vienna, Austria (pha)
- Student English Corpus of Uppsala University (use)
- Swedia 2000, subpart of a research corpus of Swedish dialects (swedia)
- The Swedish part of the Spencer project (spencer)

A special case is the Helsinki corpus containing IMDI corpus files and a 'creative' mix of Corpus/Session files. These files should have been IMDI Catalogue files. Communication with the authors is needed to do custom modifications for a complete upgrade to IMDI 3.0.

The table below gives an overview of the amount of IMDI and resource files in the four sub-corpora. The first number is the actual files copied. The second number (between braces) is the number of linked files. A difference between these numbers indicates that not all the linked files could be copied. This happens when a link is broken or when it's not allowed to access the file (read permission).

IMDI overview

type	pha	use	swedia	spencer
imdi	112 (117)	1957 (1957)	313 (313)	50 (50)
imdi session	96	1489	224	40
imdi corpus	6	468	89	10
description files	2	1	1	4

Resource overview

format	pha	use	swedia	spencer
wav	0 (93)	-	223 (224)	-
gif	-	-	438 (448)	-
jpeg	-	-	116 (224)	-
txt	63 (66)	1489 (1489)	-	-
html	-	-	1	-
doc	-	-	-	0 (44)
cha	-	-	-	0 (40)
pdf	-	0 (1)	-	-
unknown	1 (4)	-	-	-

2.2 Metadata upgrade

To upgrade the Lund IMDI files to IMDI 3.0 several generic repairs and a few custom modifications were needed. A serious problem with the older IMDI files was that they contain characters encoded in UTF-8 mixed with characters in [ISO8859-1] and sometimes even Windows specific codes [WINLATIN1]. A tool was made to correct these non-UTF8 codes.

Generic modifications include upgrading IMDI 1.x to 3.0 and correcting all 3.0 files with available tools to make the IMDI files valid. Some custom corrections were done to synchronise the closed vocabulary values and correct syntax errors.

2.3 File types and formats

The resources in the Lund archive are correctly specified using IMDI Type/Format values. One exception is the Spencer corpus which uses the format 'text/doc' which is not a known MIME type and should be corrected.

2.4 Language identification upgrade

Language names and identifiers are based on the 14th edition of the Ethnologue list from SIL. An upgrade to the new list (15th edition, ISO 639-3) can be expected in the near future. The table below gives an indication of the changes to be made.

Edition 14		Edition 15	
Name	Code	Name	Code
Serbo-Croatian	SRC	Bosnian	bos
Serbo-Croatian	SRC	Croatian	hrv
Serbo-Croatian	SRC	Serbian	srp
Armenian	ARM	Armenian	hye
Cantonese	YUH	Chinese, Yue	yue
Kurdi	KDB	Kurdish, Central	ckb
Kurdi	KDB	Kurdish, Southern	sdh
Norwegian, bokmal	NRR	Norwegian, Bokmål	nob

3 INL Metadata

3.1 CGN Corpus

Recently, a 2nd version of the CGN corpus has been released. Information for this version has been added to the tables below.

IMDI overview

file type	CGN	CGN 2.0
imdi	12,893	12,907
imdi session	12,767	12,780
imdi corpus	126	126
description files	1	1

Resource overview

format	CGN	CGN 2.0
wav ¹	12,767	12,780
bpt ²	14,199	14,212
lxk	12,767	12,780
pri	12,767	12,780
prx ³	796	796
skp ⁴	26,966	26,992
tag	12,767	12,780
tig ⁵	1,303	1,303
eaf ⁶	12,767	12,780

3.2 Metadata upgrade

The 1st version of the Dutch Spoken Corpus (or CGN) at the INL uses an older version of IMDI and is therefore the first choice to be integrated. The upgrade tool from IMDI 1.x to 3.0 is used to make the metadata valid according to the new schema and also corrections of cv values had to be done.

3.3 File types and formats

The format specifier of several text resources are upgraded to conform to the most recent IMDI format for written resources which is based on MIME types. The table below gives an overview of this conversion.

old value	new value
text/x-bpt	text/x-cgn-bpt+xml
text/x-lxk	text/x-cgn-lxk+xml
text/x-pri	text/x-cgn-pri+xml
text/x-prx	text/x-cgn-prx+xml
text/x-skp	text/x-cgn-skp+xml
text/x-tag	text/x-cgn-tag+xml
text/x-tig	text/x-cgn-tig+xml

¹ Although there are 12,780 wav files in the CGN, only 12,767 had annotations and metadata in the 1st version. CGN 2.0 contains all available annotations and metadata.

² There are 12,780 auto-generated bpts in CGN 2.0. 1,432 were manually verified and stored separately.

³ The prx files contain manually verified prosodic annotations. Two (groups of) annotators annotated (the same) 398 resources.

⁴ The skp files link annotations to the correct time frames in the sound files. Skps were created for three sets of annotations: 12,780 + 12,780 + 1,432.

⁵ Syntactic annotations were created for 1,303 fragments.

3.4 *Future plans*

The CGN is the INL's test case. When the CGN has been integrated into the distributed DAM-LR solution, more language materials will follow.

4 SOAS Metadata

The Endangered Languages Archive [ELAR], SOAS, has possibly the largest variety of potential contributors and users of the archive materials, and so the archive design and metadata policies are designed to provide the greatest possible flexibility in materials and associated metadata, while conforming to a large number of metadata standards.

After an extensive review of digital archive design principles, an archive architecture was designed that allows the maintenance a local metadata set that is continuously extensible, while serving and IMDI compliant version of the metadata in order to meet the requirements of the DAM-LR project. The archive design is heavily influenced by the Open Archive Information System [OAIS] and in particular the division between ingestion (accession), archive and dissemination formats. The OAIS model defines sets of data in terms of 'information packages', which is defined as a bundle of resources plus the associated metadata. For ingestion, archiving and dissemination packages are respectively known as a Submission Information Package (SIP), Archival Information Package (AIP) and a Dissemination Information Package (DIP).

4.1 *Ingestion metadata*

The following metadata elements are the minimum required for accessioning materials into ELAR:

- Identifier: A means to uniquely identify each item in the SIP. This might be either:
 - a unique name for each item listed together with the full filename (and media carrier label if relevant) or
 - a unique filename for each item
- Format: Describe formats
 - file format
 - mark-up format
 - character encoding format
- Creator: Entity primarily responsible for making the content
- Subject.language: The language(s) which is described or documented
- Language: The language in which the content is expressed or introduced.
- Rights: Information about rights held in and over the resource (ELAR will apply default values if required)

Additionally, ELAR is developing dedicated applications for the automated ingestion of data in IMDI and [OLAC] compliant formats.

4.2 *Archive metadata*

ELAR encourages users to add metadata and new metadata categories. Therefore, a primary role in managing the metadata is moderating the ongoing development of metadata supplied by users. The AIP metadata is initiated with fields primarily taken from the IMDI metadata schema, with some initial extensions.

ELAR will be the first language archive to actively encourage users to translate metadata into different languages, to support the browsing of metadata via these different languages and to allow multiple concurrent values for metadata fields. This requires individual metadata values to be uniquely and unambiguously identifiable, which is a more fine-grained model of metadata than that which is supported by most current metadata schemas. While this doesn't prevent mapping archive metadata to well-known standards, it does prevent ELAR being 'IMDI-native', meaning that our metadata integration strategy for the DAM-LR project is to dynamically map our archive metadata to an IMDI compliant format.

4.3 *Dissemination metadata*

ELAR's catalogue serving system is part of the dissemination strategy, serving DIPs with metadata in IMDI, OLAC (Open Language Archives Community), OAI (Open Archives Initiative) and Text Encoding Initiative [TEI] formats. The AIP metadata will be dynamically mapped to each of these formats in order to serve a variety of user communities in parallel. An advantage of this mapping system is that the archive will be able to map to future versions of metadata formats without needing to change the archive metadata, thereby saving considerable future resources and the need to re-verify the changes for preservation purposes. Further details of the mappings to OLAC, OAI, and TEI formats are outside the scope of the DAM-LR project.

The mapping to IMDI is to meet the requirements of the DAM-LR project and to support users of the archive who wish to take advantage of IMDI compliant software tools. As the AIP metadata system has been initiated with IMDI fields, the mapping to IMDI metadata in the DIP metadata has been a simple task, and has not presented any significant design problems.

5 MPI Metadata

The last two years the major part of the MPI archive was upgraded from IMDI 1.x to 3.0. Also the diverse domain of resource formats was synchronised with standard formats and a clean-up of cv values was done. However, there are still a lot of resources to be processed and integrated which is an ongoing task of corpus management. Tools are continuously developed to assist corpus managers with checking and correction of new resources to be ingested into the archive.

5.1 MPI Corpus

The table below gives a global overview of the IMDI archive from last year.

IMDI overview

type	count
imdi	12893
imdi session	22004
imdi corpus	3576
description files	8024

Media Resources

format	count
jpg	8987
mov	3357
mpeg	1226
mpg	11543
wav	7909

Text Resources

format	count
chat	3323
eaf	1877
pdf	260
sht	387
tr	723
txt	5833
html	4
typ	8

5.2 DBD Corpus

A new IMDI session profile (see Appendix A) was made to enable IMDI metadata creation for the Dutch Bilingual Database (DBD). The corpus was added to the IMDI domain.

The following additional keys were defined:

Name	CV
DBD.LanguageMode	http://www.mpi.nl/IMDI/Schema/DBD.LanguageMode.xml
DBD.CountryOfBirth	http://www.mpi.nl/IMDI/Schema/Countries.xml
DBD.AgeAtImmigration	http://www.mpi.nl/IMDI/Schema/DBD.AgeAtImmigration.xml
DBD.LevelOfBilingualism	http://www.mpi.nl/IMDI/Schema/DBD.LevelOfBilingualism.xml

IMDI overview

type	count
imdi	1366
imdi session	1190
imdi corpus	176
description files	633

Media Resources

format	count
jpg	4
pdf	5
wav	194

Text Resources

format	count
chat	668
eaf	3
pdf	21
txt	1

6 References

[ELAR] Endangered Languages Archive
<http://www.hrelp.org/archive/>

[IMDI] ISLE Metadata Initiative
<http://www.mpi.nl/IMDI/>

[ISO8859-1] ISO-8859. International Standard, Information Processing, 8-bit Single-Byte Coded Graphic Character Sets, Part 1: Latin alphabet No. 1, 1987

[OAI] Open Archives Initiative
<http://www.openarchives.org/>

[OAIPMH] The Open Archives Initiative Protocol for Metadata Harvesting, Version 2.0, 2002-06-14
<http://www.openarchives.org/OAI/openarchivesprotocol.html>

[OAIS] Reference model for an Open Archival Information System, January 2002
<http://www.ccsds.org/documents/650x0b1.pdf>

[OLAC] Open Language Archive Community
<http://www.language-archives.org/>

[RFC1738] Uniform Resource Locators (URL), December 1994
<http://www.ietf.org/rfc/rfc1738.txt>

[RFC2279] UTF-8, a transformation format of ISO 10646, January 1998
<http://www.ietf.org/rfc/rfc2279.txt>

[SESSIONS] Metadata Elements for Session Descriptions, Version 3.0.4, MPI Nijmegen, 2003
<http://www.mpi.nl/IMDI/>

[TEI] Text Encoding Initiative
<http://www.tei-c.org/>

[UNICODE] Unicode Standard
<http://www.unicode.org/>

[WINLATIN1] Windows-1252 codepage (WinLatin1)
<http://en.wikipedia.org/wiki/Windows-1252>

[XML] Extensible Markup Language (XML) 1.0 (Third Edition), W3C, 2004
<http://www.w3.org/TR/REC-xml/>

[XMLSCHEMA] XML Schema Part 1: Structures Second Edition, W3C, October 2004
<http://www.w3.org/TR/xmlschema-1/>

Appendix A : IMDI DBD Session Profile

```
<?xml version="1.0" encoding="UTF-8"?><!DOCTYPE METATRANSCRIPT [<!ENTITY
annotationunitPrefix "">
<!ENTITY infolinkPrefix "">
<!ENTITY globalPrefix "">
<!ENTITY anonymousPrefix "">
<!ENTITY mediafilePrefix "">
]>
<METATRANSCRIPT Date="2005-06-17" FormatId="IMDI 3.0" Originator="Editor -
Profile:local/DBD_Profile.xml" Type="SESSION.Profile" Version="1"
xmlns="http://www.mpi.nl/IMDI/Schema/IMDI"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.mpi.nl/IMDI/Schema/IMDI ./IMDI_3.0.xsd">
  <Session XXX-HelpText="A session bundles information about the circumstances and
conditions of the linguistic event, groups the resources belonging to this
linguistic event, records the administrative information of the event, and
describes the content of the event." XXX-Multiple="false">
    <Name XXX-HelpText="A short name to identify the session. The name of the
session can be considered as shorthand of the session title." XXX-Multiple="false"
XXX-Type="String">DBD Profile</Name>
    <Title XXX-Type="String"/>
    <Date XXX-Type="Date">Unspecified</Date>
    <Description LanguageId="" Link="" XXX-Multiple="true" XXX-Type="String"/>
    <MDGroup XXX-Visible="false">
      <Location>
        <Continent Link="http://www.mpi.nl/IMDI/Schema/Continents.xml"
Type="ClosedVocabulary" XXX-Type="FixedCV">Unspecified</Continent>
        <Country Link="http://www.mpi.nl/IMDI/Schema/Countries.xml"
Type="ClosedVocabulary" XXX-Type="FixedCV">Unspecified</Country>
        <Region XXX-Multiple="true" XXX-Type="String">Unspecified</Region>
        <Address XXX-Type="String">Unspecified</Address>
      </Location>
      <Project>
        <Name XXX-Type="String"/>
        <Title XXX-Type="String"/>
        <Id XXX-Type="String"/>
        <Contact>
          <Name XXX-Type="String"/>
          <Address XXX-Type="String"/>
          <Email XXX-Type="String"/>
          <Organisation XXX-Type="String"/>
        </Contact>
        <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
      </Project>
      <Keys>
        <Key Name="" Type="OpenVocabulary" XXX-Multiple="true"/>
      </Keys>
      <Content XXX-Multiple="false">
        <Genre Link="http://www.mpi.nl/IMDI/Schema/Content-Genre.xml"
Type="OpenVocabulary" XXX-Type="FixedCV">Unspecified</Genre>
        <SubGenre DefaultLink=""
Link="http://www.mpi.nl/IMDI/Schema/Content-SubGenre.xml" Type="OpenVocabularyList"
XXX-FollowUpDepend="Genre" XXX-Type="FixedCV">Unspecified</SubGenre>
        <Task Link="http://www.mpi.nl/IMDI/Schema/Content-Task.xml"
Type="OpenVocabulary" XXX-Type="FixedCV"/>
        <Modalities Link="http://www.mpi.nl/IMDI/Schema/Content-
Modalities.xml" Type="OpenVocabularyList" XXX-Type="FixedCV"/>
        <Subject DefaultLink=""
Link="http://www.mpi.nl/IMDI/Schema/Content-Subject.xml" Type="OpenVocabularyList"
XXX-Type="FreeCV"/>
      <CommunicationContext>
        <Interactivity Link="http://www.mpi.nl/IMDI/Schema/Content-
Interactivity.xml" Type="ClosedVocabulary" XXX-
Type="FixedCV">Unspecified</Interactivity>
    </Session>
</METATRANSCRIPT>
```

```

        <PlanningType Link="http://www.mpi.nl/IMDI/Schema/Content-
PlanningType.xml" Type="ClosedVocabulary" XXX-
Type="FixedCV">Unspecified</PlanningType>
        <Involvement Link="http://www.mpi.nl/IMDI/Schema/Content-
Involvement.xml" Type="ClosedVocabulary" XXX-
Type="FixedCV">Unspecified</Involvement>
        <SocialContext Link="http://www.mpi.nl/IMDI/Schema/Content-
SocialContext.xml" Type="ClosedVocabulary" XXX-
Type="FixedCV">Unspecified</SocialContext>
        <EventStructure Link="http://www.mpi.nl/IMDI/Schema/Content-
EventStructure.xml" Type="ClosedVocabulary" XXX-
Type="FixedCV">Unspecified</EventStructure>
        <Channel Link="http://www.mpi.nl/IMDI/Schema/Content-
Channel.xml" Type="ClosedVocabulary" XXX-Type="FixedCV">Unspecified</Channel>
        </CommunicationContext>
        <Languages>
        <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
        <Language ResourceRef="" XXX-Multiple="true">
        <Id XXX-Type="String">Unspecified</Id>
        <Name Link="http://www.mpi.nl/IMDI/Schema/MPI-
Languages.xml" Type="OpenVocabulary" XXX-Type="FixedCV">Unspecified</Name>
        <Dominant XXX-Type="Boolean">Unspecified</Dominant>
        <SourceLanguage XXX-
Type="Boolean">Unspecified</SourceLanguage>
        <TargetLanguage XXX-
Type="Boolean">Unspecified</TargetLanguage>
        <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
        </Language>
        </Languages>
        <Keys XXX-Multiple="false">
        <Key DefaultLink="" Name="DBD.LanguageMode" XXX-
Multiple="false" XXX-
Type="CV:http://www.mpi.nl/IMDI/Schema/DBD.LanguageMode.xml">Unspecified</Key>
        </Keys>
        <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
        </Content>
        <Actors XXX-Multiple="false" XXX-Visible="false">
        <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
        <Actor ResourceRef="" XXX-Multiple="true" XXX-Tag="Consultant">
        <Role Link="http://www.mpi.nl/IMDI/Schema/Actor-Role.xml"
Type="OpenVocabularyList" XXX-Multiple="false" XXX-Type="FixedCV">Consultant</Role>
        <Name XXX-Type="String"/>
        <FullName XXX-Type="String"/>
        <Code XXX-Type="String"/>
        <FamilySocialRole Link="http://www.mpi.nl/IMDI/Schema/Actor-
FamilySocialRole.xml" Type="OpenVocabularyList" XXX-
Type="FixedCV">Unspecified</FamilySocialRole>
        <Languages>
        <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
        <Language XXX-Multiple="true" XXX-Type="Language">
        <Id XXX-Type="String">Unspecified</Id>
        <Name Link="http://www.mpi.nl/IMDI/Schema/MPI-
Languages.xml" Type="OpenVocabulary" XXX-Type="FixedCV">Unspecified</Name>
        <MotherTongue XXX-
Type="Boolean">Unspecified</MotherTongue>
        <PrimaryLanguage XXX-
Type="Boolean">Unspecified</PrimaryLanguage>
        <Description LanguageId="" Link="" XXX-Multiple="true"
XXX-Type="String"/>
        </Language>
        </Languages>
        <EthnicGroup Type="OpenVocabulary" XXX-Type="String"/>
        <Age XXX-Type="Age">Unspecified</Age>

```

```

        <BirthDate XXX-Type="Date">Unspecified</BirthDate>
        <Sex Link="http://www.mpi.nl/IMDI/Schema/Actor-Sex.xml"
Type="ClosedVocabulary" XXX-Type="FixedCV">Unspecified</Sex>
        <Education XXX-Type="String">Unspecified</Education>
        <Anonymized XXX-Type="Boolean">Unspecified</Anonymized>
        <Contact>
            <Name XXX-Type="String"/>
            <Address XXX-Type="String"/>
            <Email XXX-Type="String"/>
            <Organisation XXX-Type="String"/>
        </Contact>
        <Keys XXX-Multiple="false">
            <Key DefaultLink="" Name="DBD.CountryOfBirth" XXX-
Multiple="false" XXX-
Type="CV:http://www.mpi.nl/IMDI/Schema/Countries.xml">Unspecified</Key>
            <Key DefaultLink="" Name="DBD.AgeAtImmigration" XXX-
Multiple="false" XXX-
Type="CV:http://www.mpi.nl/IMDI/Schema/DBD.AgeAtImmigration.xml">Unspecified</Key>
            <Key DefaultLink="" Name="DBD.LevelOfBilingualism" XXX-
Multiple="false" XXX-
Type="CV:http://www.mpi.nl/IMDI/Schema/DBD.LevelOfBilingualism.xml">Unspecified</Ke
y>
        </Keys>
        <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
    </Actor>
    <Actor ResourceRef="" XXX-Multiple="true" XXX-Tag="other">
        <Role Link="http://www.mpi.nl/IMDI/Schema/Actor-Role.xml"
Type="OpenVocabularyList" XXX-Type="FixedCV">Unspecified</Role>
        <Name XXX-Type="String"/>
        <FullName XXX-Type="String"/>
        <Code XXX-Type="String"/>
        <FamilySocialRole Link="http://www.mpi.nl/IMDI/Schema/Actor-
FamilySocialRole.xml" Type="OpenVocabularyList" XXX-
Type="FixedCV">Unspecified</FamilySocialRole>
        <Languages>
            <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
            <Language XXX-Multiple="true" XXX-Type="Language">
                <Id XXX-Type="String">Unspecified</Id>
                <Name Link="http://www.mpi.nl/IMDI/Schema/MPI-
Languages.xml" Type="OpenVocabulary" XXX-Type="FixedCV">Unspecified</Name>
                <MotherTongue XXX-
Type="Boolean">Unspecified</MotherTongue>
                <PrimaryLanguage XXX-
Type="Boolean">Unspecified</PrimaryLanguage>
                <Description LanguageId="" Link="" XXX-Multiple="true"
XXX-Type="String"/>
            </Language>
        </Languages>
        <EthnicGroup Type="OpenVocabulary" XXX-Type="String"/>
        <Age XXX-Type="Age">Unspecified</Age>
        <BirthDate XXX-Type="Date">Unspecified</BirthDate>
        <Sex Link="http://www.mpi.nl/IMDI/Schema/Actor-Sex.xml"
Type="ClosedVocabulary" XXX-Type="FixedCV">Unspecified</Sex>
        <Education XXX-Type="String">Unspecified</Education>
        <Anonymized XXX-Type="Boolean">Unspecified</Anonymized>
        <Contact>
            <Name XXX-Type="String"/>
            <Address XXX-Type="String"/>
            <Email XXX-Type="String"/>
            <Organisation XXX-Type="String"/>
        </Contact>
        <Keys>
            <Key Name="" Type="OpenVocabulary" XXX-Multiple="true" XXX-
Type="*" />
        </Keys>

```

```

        <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
    </Actor>
</Actors>
</MDGroup>
<Resources>
    <MediaFile XXX-Multiple="true">
        <ResourceLink XXX-Type="URL"/>
        <Type Link="http://www.mpi.nl/IMDI/Schema/MediaFile-Type.xml"
Type="ClosedVocabulary" XXX-Type="FixedCV">Unspecified</Type>
        <Format Link="http://www.mpi.nl/IMDI/Schema/MediaFile-Format.xml"
Type="OpenVocabulary" XXX-Type="FixedCV">Unspecified</Format>
        <Size XXX-Type="String"/>
        <Quality XXX-
Type="Regexp:[12345]|Unknown|Unspecified">Unspecified</Quality>
        <RecordingConditions XXX-Type="String"/>
        <TimePosition>
            <Start XXX-Type="Regexp:Unknown|[0-9][0-9]:[0-9][0-9]:[0-9][0-
9](:[0-9]+)?|Unspecified">Unspecified</Start>
            <End XXX-Type="Regexp:Unknown|[0-9][0-9]:[0-9][0-9]:[0-9][0-
9](:[0-9]+)?|Unspecified">Unspecified</End>
        </TimePosition>
        <Access>
            <Availability Type="OpenVocabulary" XXX-Type="String"/>
            <Date XXX-Type="Date">Unspecified</Date>
            <Owner XXX-Type="String"/>
            <Publisher XXX-Type="String"/>
            <Contact>
                <Name XXX-Type="String"/>
                <Address XXX-Type="String"/>
                <Email XXX-Type="String"/>
                <Organisation XXX-Type="String"/>
            </Contact>
            <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
        </Access>
        <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
        <Keys>
            <Key Name="" Type="OpenVocabulary" XXX-Multiple="true" XXX-
Type="*" />
        </Keys>
    </MediaFile>
    <WrittenResource XXX-Multiple="true">
        <ResourceLink XXX-Type="URL"/>
        <MediaResourceLink XXX-Type="URL"/>
        <Date XXX-Type="Date">Unspecified</Date>
        <Type Link="http://www.mpi.nl/IMDI/Schema/WrittenResource-Type.xml"
Type="OpenVocabulary" XXX-Type="FixedCV">Unspecified</Type>
        <SubType Link="http://www.mpi.nl/IMDI/Schema/WrittenResource-
SubType.xml" Type="OpenVocabularyList" XXX-FollowUpDepend="Type" XXX-
Type="FixedCV">Unspecified</SubType>
        <Format Link="http://www.mpi.nl/IMDI/Schema/WrittenResource-
Format.xml" Type="OpenVocabulary" XXX-Type="FixedCV">Unspecified</Format>
        <Size Type="OpenVocabulary" XXX-Type="String"/>
        <Validation>
            <Type Link="http://www.mpi.nl/IMDI/Schema/Validation-Type.xml"
Type="ClosedVocabulary" XXX-Type="FixedCV">Unspecified</Type>
            <Methodology Link="http://www.mpi.nl/IMDI/Schema/Validation-
Methodology.xml" Type="ClosedVocabulary" XXX-
Type="FixedCV">Unspecified</Methodology>
            <Level XXX-Type="Percentage">Unspecified</Level>
            <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
        </Validation>
        <Derivation Link="http://www.mpi.nl/IMDI/Schema/WrittenResource-
Derivation.xml" Type="ClosedVocabulary" XXX-Type="FixedCV">Unspecified</Derivation>
        <CharacterEncoding XXX-Type="String"/>

```

```

<ContentEncoding XXX-Type="String" />
<LanguageId Type="OpenVocabulary" XXX-Type="String" />
<Anonymized XXX-Type="Boolean">Unspecified</Anonymized>
<Access>
  <Availability Type="OpenVocabulary" XXX-Type="String" />
  <Date XXX-Type="Date">Unspecified</Date>
  <Owner XXX-Type="String" />
  <Publisher XXX-Type="String" />
  <Contact>
    <Name XXX-Type="String" />
    <Address XXX-Type="String" />
    <Email XXX-Type="String" />
    <Organisation XXX-Type="String" />
  </Contact>
  <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String" />
</Access>
<Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String" />
<Keys>
  <Key Name="" Type="OpenVocabulary" XXX-Multiple="true" XXX-
Type="*" />
</Keys>
</WrittenResource>
<Source XXX-Multiple="true">
  <Id XXX-Type="String" />
  <Format Link="http://www.mpi.nl/IMDI/Schema/Source-Format.xml"
Type="OpenVocabulary" XXX-Type="FixedCV">Unspecified</Format>
  <Quality XXX-
Type="Regexp:[12345]|Unknown|Unspecified">Unspecified</Quality>
  <CounterPosition>
    <Start XXX-Type="Regexp:Unknown|[0-
9]+|Unspecified">Unspecified</Start>
    <End XXX-Type="Regexp:Unknown|[0-
9]+|Unspecified">Unspecified</End>
  </CounterPosition>
  <TimePosition>
    <Start XXX-Type="Regexp:Unknown|[0-9][0-9]:[0-9][0-9]:[0-9][0-
9](:[0-9]+)?|Unspecified">Unspecified</Start>
    <End XXX-Type="Regexp:Unknown|[0-9][0-9]:[0-9][0-9]:[0-9][0-
9](:[0-9]+)?|Unspecified">Unspecified</End>
  </TimePosition>
  <Access>
    <Availability Type="OpenVocabulary" XXX-Type="String" />
    <Date XXX-Type="Date">Unspecified</Date>
    <Owner XXX-Type="String" />
    <Publisher XXX-Type="String" />
    <Contact>
      <Name XXX-Type="String" />
      <Address XXX-Type="String" />
      <Email XXX-Type="String" />
      <Organisation XXX-Type="String" />
    </Contact>
    <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String" />
  </Access>
  <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String" />
  <Keys>
    <Key Name="" Type="OpenVocabulary" XXX-Multiple="true" XXX-
Type="*" />
  </Keys>
</Source>
<Anonyms>
  <ResourceLink XXX-Type="URL" />
  <Access>
    <Availability Type="OpenVocabulary" XXX-Type="String" />
    <Date XXX-Type="Date">Unspecified</Date>

```

```
<Owner XXX-Type="String"/>
<Publisher XXX-Type="String"/>
<Contact>
  <Name XXX-Type="String"/>
  <Address XXX-Type="String"/>
  <Email XXX-Type="String"/>
  <Organisation XXX-Type="String"/>
</Contact>
<Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
  </Access>
  </Anonyms>
</Resources>
<References>
  <Description LanguageId="" Link="" XXX-Multiple="true" XXX-
Type="String"/>
  </References>
</Session>
</METATRANSCRIPT>
```