

Curation report - Dutch Bilingual Database

CLARIN-NL Data Curation Service

November 2013 / February 2014

Maaske Treurniet, Henk van den Heuvel, Vanja de Lint, Eric Sanders
CLST, Radboud University Nijmegen

0. The Dutch Bilingual Database

The DBD is a rather substantial collection of data (over 1,500 sessions) from a number of projects and research programmes that were directed at investigating multilingualism and comprises data originating from Dutch, Sranan, Sarnami, Papiamentu, Arabic, Berber and

Turkish speakers . At the basis of the collection is the research project TCULT[1] (1998-2002) in which intercultural language contacts in the Dutch city of Utrecht were studied. DBD established a first curation of the TCULT data and added many more bilingual data sets collected in the period 1985 – 2005. The current version of the DBD has IMDI[2] metadata files and is made accessible by the MPI at http://corpus1.mpi.nl/ds/imdi_browser/. The audio and text data are stored at the MPI and (expectedly) at the Meertens Institute. The DBD data consists of audio recordings, most of which are in WAV format while some are in MP3. Most transcripts that are available are in CHAT, some in TXT, PDF or EAF. Metadata are available in IMDI. In a number of cases additional metadata are available in TXT or PDF format. Occasionally, additional materials are available. These include descriptions (PDF) or images (JPG; PDF) of the pictures books or cartoons used to elicit the data and protocols/descriptions of the tasks involved (PDF).

Before starting any curation work we had interviews with relevant researchers involved in the collection of the DBD/TCULT data with two objectives in mind:

1. Past: How and in which phases were the data collected. Is there any relevant data missing and easy to retrieve?
2. Future: What is the best presentation of the data for the future in terms of key information, structure and formats?

To this end we had interviews with: Pieter Muysken, Leoni Cornips, Willy Jongenburger, Theo Meder, and Linda van Meel about the data and their representations. With Pieter Muysken a metadata profile was discussed and agreed upon.

[1] <http://ebookbrowse.com/tcult-pdf-d68190469>

[2] <http://www.mpi.nl/isle/>

1. README

For DBD, the directory hierarchy is as follows. Number of sessions is included in brackets.

Berber

Subdirectories: E-rramdani (97), El Aissati (20)

Dutch Ethnolects

Subdirectories: Moroccan (23), Moroccan and Turkish (13), Turkish (16), other (23)

Moroccan Arabic

Subdirectories: Bos 1997 (639), Boumans 2003 (2), El Aissati 1997 (49), Nortier 1986 (10), Nortier 1991 (12), Rooselaar 1990 (96), TCULT (17).

Papiamentu

Subdirectories: Code-Switching (108), Parent-Child Reading (54)

Sranan:Sarnami

Subdirectories:Class Recordings (10), Different language background (13), Dutch (55), MISC (5), Sarnami-Dutch (154), Sranan-Dutch (49), Turkish-Dutch (8), Unknown language background (7).

Turkish

Subdirectories: Backus (110), Eversteijn (21)

Transcripts are available for part of the sessions and are in .cha format. A few sessions have additional .eaf transcripts, that are derived from original .cha files. Some sessions contain a general description in .pdf instead of a .cha transcript. One session has only an .eaf transcript. The conditions for use of DBD are attached as both .txt and .pdf in a file named Conditions_for_use_DBD.

Part of the data of Berber E-rramdani contains an explanation about the methods used for obtaining the data. The file is linked to from the sessions that it is relevant for.

The data of Moroccan Arabic Bos (1997) contains a .pdf document called aarsen_bos_cartoons.one.pdf. This file is linked to from all sessions that it is relevant for.

2. Curation report

In 2012/2013 CLARIN-NL's Data Curation Service undertook the curation of the Dutch Bilingual Database as stored at the Max Planck Institute for Psycholinguistics in Nijmegen. The curation consisted of:

- 1) the addition of several missing transcription files
- 2) the transformation of the metadata files from .imdi into .cmdi
- 3) the definition of standardized ISOcat data categories and transformation of the metadata categories into these standardized categories.

An explanation on these curation steps will be given below.

1) Addition of missing files

Our restoring activities concentrated on retrieving 1. missing data in sessions present at MPI and 2. additional sessions from archives. Thus, retrieving existing data was our goal, not creating data ourselves, like missing transcriptions. The latter is not part of our task as DCS. Potential additional sessions that we found on a list by Leonie Cornips were, as it turned out, not archived at the Meertens institute. Our initial hope for this additional material turned out to be in vain.

During the inventory of the different sessions and the files they contain, we found out that some sessions contain only a media file, other sessions contain both a media file and a transcript, and again other sessions contain only a transcript. For a part of DBD, originating from the project of Bos (1997), the part of Moroccan Arabic data contained several .wav files with corresponding .cha transcripts. However some of the .wav files lacked corresponding .cha transcripts. The CHILDES database was found to contain 32 .cha transcripts that were not present in DBD. Although these files were already curated, because they were stored in the CHILDES database, it was decided to add the transcripts to DBD as well, to provide the data as complete as possible to the users. The metadata files were adapted: a component describing the .cha files was added as well as a resource proxy. For 6 of the .cha files the .wav file they belonged to could not be found in the database and these files were not added to the database. A list of the .cha transcripts that were obtained from CHILDES, is provided in Appendix A.

2) Conversion of metadata files

The basic principle for creating a metadata profile for DBD, has been that as much information as possible should be put in dedicated fields in a metadata profile, to optimize search options in the database. To this end a new CMDI metadata profile for DBD was set up, based on the current IMDI structure.

The metadata fields in the DBD profile are copied from the existing database at the MPI and converted to the appropriate CLARIN structure, that is to say corresponding to the CLARIN Component MetaData initiative (CMDI). The transformation was done with a .xslt sheet. This was originally provided by the CLARIN data centre at the MPI, but had to be adapted severely. In the IMDI structure were a lot of “key” metadata fields that function as a catch to store metadata that are otherwise not defined. All these “key” fields were either mapped to an existing or newly created CMDI element or removed when empty or redundant. The xslt sheet was adapted in order to solve these issues. A list of all “key” fields and the action taken with respect to them can be found in appendix D.

In CMDI a strict order of the components and element is demanded. The order has to be exactly the same as in the profile. The .xslt script was adapted to this need.

A postprocessing script was created that solved a couple of issues. The size of all media and annotation files was added in a consistent way. A component describing the added .cha files was added (in WrittenResource). Components describing media files or written resources that are not present in the corpus were removed. Resource proxies were added or changed where necessary. All CMDI files were validated against the DBD.xsd of the profile.

The profile can be found in the component registry: http://catalog.clarin.eu/ds/ComponentRegistry?item=clarin.eu:cr1:p_1391763610422

In Appendix B, an example of a CMDI file is included.

3) Standardisation of metadata categories

CLARIN developed the ISOcat data category registry to create standardized metadata categories. For the curation of DBD, ISOcat references were created for all the existing and added metadata categories. A profile called DBD was created, consisting of all the ISOcat metadata categories needed to describe the DBD data and new bilingual research data that is likely to be added to the database in the near future.

A list of newly introduced ISOcat metadata elements can be found in Appendix C.

Appendix A

Files obtained from the CHILDES database

dbd_dut_03_51_01_253.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_05/Annotation
dbd_dut_03_51_01_254.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_05/Annotation
dbd_dut_03_51_01_953.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_05/Annotation
dbd_dut_03_51_01_255.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_05/Annotation
dbd_dut_03_51_01_256.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_05/Annotation
dbd_dut_03_51_01_952.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_05/Annotation
dbd_dut_03_51_01_272.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_07/Annotation
dbd_dut_03_51_01_273.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_07/Annotation
dbd_dut_03_51_01_274.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_07/Annotation
dbd_dut_03_51_01_275.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_07/Annotation
dbd_dut_03_51_01_276.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_07/Annotation
dbd_dut_03_51_01_243.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_04/Annotation
dbd_dut_03_51_01_943.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_04/Annotation
dbd_dut_03_51_01_245.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_04/Annotation
dbd_dut_03_51_01_942.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_04/Annotation
dbd_dut_03_51_01_941.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_04/Annotation
dbd_dut_03_51_01_241.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_04/Annotation
dbd_dut_03_51_01_843.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_04/Annotation
dbd_dut_03_51_01_262.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_06/Annotation

dbd_dut_03_51_01_263.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_06/Annotation
dbd_dut_03_51_01_264.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_06/Annotation
dbd_dut_03_51_01_963.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_06/Annotation
dbd_dut_03_51_01_265.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_06/Annotation
dbd_dut_03_51_01_266.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_06/Annotation
dbd_dut_03_51_01_962.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_06/Annotation
dbd_dut_03_51_01_961.cha => boumans/Bos_1997/Monolingual_Dutch/
Monolingual_Dutch_06/Annotation

Not added because of missing .wav

dbd_dut_03_05_01_481.cha
dbd_dut_03_51_01_341.cha
dbd_dut_03_51_01_351.cha
dbd_dut_03_51_01_361.cha
dbd_dut_03_51_01_371.cha
dbd_dut_03_51_01_744.cha

Appendix B

Example of CMDI file

```
<CMD xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.clarin.eu/cmd/" CMDVersion="1.1"
xsi:schemaLocation="http://www.clarin.eu/cmd/ http://catalog.clarin.eu/ds/ComponentRegistry/
rest/registry/profiles/clarin.eu:cr1:p_1391763610422/xsd">
  <Header>
    <MdCreator>imdi2clarin.xsl</MdCreator>
    <MdCreationDate>2014-02-25</MdCreationDate>
    <MdSelfLink>hdl:1839/00-0000-0000-0005-B3A8-2@format=cmdi</MdSelfLink>
    <MdProfile>clarin.eu:cr1:p_1390399932755</MdProfile>
  </Header>
  <Resources>
    <ResourceProxyList>
      <ResourceProxy id="d530e331">
        <ResourceType mimetype="audio/x-wav">Resource</ResourceType>
        <ResourceRef>hdl:1839/00-0000-0000-0005-B3A9-F</ResourceRef>
      </ResourceProxy>
      <ResourceProxy id="d530e393">
        <ResourceType mimetype="application/pdf">Resource</ResourceType>
        <ResourceRef>hdl:1839/00-0000-0000-0009-7D96-5</ResourceRef>
      </ResourceProxy>
    </ResourceProxyList>
    <JournalFileProxyList/>
    <ResourceRelationList/>
  </Resources>
  <Components>
    <DBD>
      <Session>
        <Name>DBD_DUT_14_11_05_135</Name>
        <Title>Dutch Bilingualism Database, Ethnic Dutch, Session 135</Title>
        <Date>2000-10-20</Date>
        <Descriptions>
          <Description xml:lang="nl">Deze sessie is opgenomen op twee audio files van in totaal
01:02:23. Het gaat om een interview met een Turkse jeugdwerker. De interviewer is een
vrouwelijke onderzoeker. </Description>
          <Description xml:lang="en">This session is recorded on 2 audio files of a total of
01:02:23. It is an interview with a Turkish youth social worker. The interviewer is a Dutch female
researcher.</Description>
        </Descriptions>
        <Location>
```



```

<Continent>Europe</Continent>
<Country>Netherlands</Country>
<Address>Utrecht</Address>
</Location>
<Project>
  <Name>T-Cult</Name>
  <Title>Research on Language and Culture in the Multicultural Neighbourhoods Lombok
and Transvaal in Utrecht, The Netherlands</Title>
  <Id>14</Id>
  <Contact>
    <Person>Ms L. Cornips, PhD. and/or Ms H.Dibbits, PhD.</Person>
    <Address>Joan Muyskenweg 25, 1096 CJ Amsterdam, tel. +31 (0)20 462 85 00</
Address>
    <Email>info@meertens.knaw.nl</Email>
    <Organisation>Meertens Institute</Organisation>
  </Contact>
  <Descriptions>
    <Description xml:lang="nl">TCULT is een interdisciplinair onderzoeksproject waarbinnen
onderzoek wordt gedaan naar taal en cultuur in de multiculturele wijk Lombok en Transvaal
in Utrecht. Hiertoe wordt het culturele en linguïstische repertoire van individuen
en van groepen autochtone Nederlanders, Marokkanen, Turken, Surinamers, en andere
etnische groepen bestudeerd. Het gaat daarbij niet alleen om de interne structuur van taal en
cultuuruitingen, maar ook om de functie en betekenis ervan voor de verschillende groepen. De
structuur van taalvarianten, verhalen, liedjes, feesten, rituelen en gebruiksvoorwerpen
worden door middel van veldwerk in kaart gebracht, waarna de resultaten getoetst worden
aan theoretische invalshoeken. Het project duurt van 1 mei 1998 - 1 mei 2002 en wordt
gerealiseerd in samenwerking met de Katholieke Universiteit Brabant, de Rijksuniversiteit
Leiden, de Universiteit van Amsterdam en de Universiteit Utrecht. Daarnaast wordt het
financieel gesteund door NWO en het Fonds Sociale Vernieuwing en Integratie van de
Gemeente Utrecht.</Description>
  </Descriptions>
</Project>
<Content>
  <Genre>Discourse</Genre>
  <SubGenre>Interview</SubGenre>
  <Task>No task</Task>
  <Modalities>Speech</Modalities>
  <Subject>Traditions, cultural differences</Subject>
  <CommunicationContext>
    <Interactivity>interactive</Interactivity>
    <PlanningType>semi-spontaneous</PlanningType>
    <Involvement>non-elicited</Involvement>
    <SocialContext>Controlled environment</SocialContext>
    <EventStructure>Conversation</EventStructure>

```

```

    <Channel>Face to Face</Channel>
  </CommunicationContext>
  <Languages>
    <Language>
      <LanguageID>ISO639-3:nld</LanguageID>
      <LanguageName>Dutch</LanguageName>
      <Dominant>true</Dominant>
      <SourceLanguage>Unspecified</SourceLanguage>
      <TargetLanguage>true</TargetLanguage>
    </Language>
  </Languages>
  <Descriptions>
    <Description xml:lang="nl">In dit interview worden vragen gesteld over tradities, cultuur
en culturele verschillen.</Description>
    <Description xml:lang="en">In this interview questions are being asked about traditions,
culture and cultural differences.</Description>
  </Descriptions>
</Content>
<Actors>
  <Actor>
    <Role>Interviewer</Role>
    <Name>Margreet</Name>
    <FullName>Margreet Dorleijn</FullName>
    <Code>MAR</Code>
    <FamilySocialRole>Unspecified</FamilySocialRole>
    <EthnicGroup>Dutch</EthnicGroup>
    <Age>Unspecified</Age>
    <BirthDate>Unspecified</BirthDate>
    <Sex>Female</Sex>
    <Education>PhD</Education>
    <Anonymized>Unspecified</Anonymized>
    <Contact>
      <Person/>
      <Address/>
      <Email/>
      <Organisation/>
    </Contact>
  </Actor>
</Actors>
<Languages>
  <Language>
    <LanguageID>ISO639-3:nld</LanguageID>
    <LanguageName>Dutch</LanguageName>
    <MotherTongue>true</MotherTongue>
    <PrimaryLanguage>true</PrimaryLanguage>
    <LanguageBackground>

```

```

    <LanguageKnowledge/>
  </LanguageBackground>
</Language>
</Languages>
</Actor>
<Actor>
  <Role>Consultant</Role>
  <Name>Emin</Name>
  <FullName>Emin Yilmaz</FullName>
  <Code>emi</Code>
  <FamilySocialRole>Unspecified</FamilySocialRole>
  <EthnicGroup>Turkish</EthnicGroup>
  <Age>Unspecified</Age>
  <BirthDate>Unspecified</BirthDate>
  <Sex>Male</Sex>
  <Education>Unspecified</Education>
  <Anonymized>>false</Anonymized>
  <Contact>
    <Person/>
    <Address/>
    <Email/>
    <Organisation/>
  </Contact>
  <Languages>
    <Language>
      <LanguageID>ISO639-3:nld</LanguageID>
      <LanguageName>Dutch</LanguageName>
      <MotherTongue>Unknown</MotherTongue>
      <PrimaryLanguage>Unknown</PrimaryLanguage>
      <LanguageBackground>
        <LanguageKnowledge/>
      </LanguageBackground>
    </Language>
    <Language>
      <LanguageID>ISO639-3:tur</LanguageID>
      <LanguageName>Turkish</LanguageName>
      <MotherTongue>Unknown</MotherTongue>
      <PrimaryLanguage>Unknown</PrimaryLanguage>
      <LanguageBackground>
        <LanguageKnowledge/>
      </LanguageBackground>
    </Language>
  </Languages>
</Actor>

```

```

</Actors>
<Resources>
  <MediaFile ref="d530e331">
    <ResourceLink>../Media/DBD_DUT_14_11_05_135_1.wav</ResourceLink>
    <Type>audio</Type>
    <Format>audio/x-wav</Format>
    <Quality>Unspecified</Quality>
    <RecordingConditions/>
    <TotalSize>
      <Number>330031676</Number>
      <SizeUnit>bytes</SizeUnit>
    </TotalSize>
    <TimePosition>
      <Start>Unspecified</Start>
      <End>Unspecified</End>
    </TimePosition>
    <Access>
      <Availability/>
      <Date>Unspecified</Date>
      <Owner/>
      <Publisher/>
      <Contact>
        <Person/>
        <Address/>
        <Email/>
        <Organisation/>
      </Contact>
    </Access>
  </MediaFile>
  <WrittenResource ref="d530e393">
    <ResourceLink>../Annotations/DBD_DUT_14_11_05_135_tr.pdf</ResourceLink>
    <MediaResourceLink/>
    <Date>Unspecified</Date>
    <Type>Annotation</Type>
    <SubType>Unspecified</SubType>
    <Format>application/pdf</Format>
    <Derivation>Analysis</Derivation>
    <ContentEncoding/>
    <CharacterEncoding/>
    <LanguageId/>
    <Anonymized>Unspecified</Anonymized>
    <TotalSize>
      <Number>111196</Number>
      <SizeUnit>bytes</SizeUnit>

```

```
</TotalSize>
<Validation>
  <Type>Unspecified</Type>
  <Methodology>Unspecified</Methodology>
  <Level>Unspecified</Level>
</Validation>
<Access>
  <Availability/>
  <Date>Unspecified</Date>
  <Owner/>
  <Publisher/>
  <Contact>
    <Person/>
    <Address/>
    <Email/>
    <Organisation/>
  </Contact>
</Access>
</WrittenResource>
</Resources>
<References/>
</Session>
</DBD>
</Components>
</CMD>
```

Appendix C

Newly introduced ISOcat metadata elements

Name	Type
ageAtImmigration	complex/open
contentEncoding	complex/open
discursive	complex/open
familyAge	complex/open
familyStructure	complex/open
friendsStructure	complex/open
grade	complex/open
homeLanguage1	complex/closed
homeLanguage2	complex/closed
interactional	complex/open
languageBackground	complex/open
languageKnowledge	complex/open
languageMode	complex/open
languageProfile	complex/open
levelOfBilingualism	complex/open
originParents	complex/open
originParticipant	complex/open
owner	complex/open
primaryLanguage	complex/open
publisher	complex/open
residenceHistory	complex/open
resourceRef	complex/open

Appendix D

List of all “key” fields and the actions taken with respect to them

Key	Action
<i>IMDI_file_creator</i>	<i>added to CMDI profile renamed to “IMDICreator”</i>
<i>IMDI_file_corrector</i>	<i>added to CMDI profile renamed to “IMDICorrector”</i>
<i>TranscriptionFile</i>	<i>left out of CMDI profile n=109. all keys referred to .txt files that do not exist in our version of the database.</i>
<i>MediataggerFile</i>	<i>left out of CMDI profile n=109. 26 of which were empty fields; 83 of which referred to .mt files that do not exist in our version of the database.</i>
<i>MediaFile1 MediaFile2</i>	<i>left out of CMDI profile either empty, redundant or non-existent reference</i>
<i>Experimenter</i>	<i>added to CMDI profile</i>
<i>Condition</i>	<i>added to CMDI profile</i>
<i>comment Comment</i>	<i>left out of CMDI profile fields were always empty</i>
<i>DBD.CountryOfBirth</i>	<i>added to CMDI profile renamed to “BirthCountry”</i>
<i>Countries</i>	<i>added to CMDI profile renamed to “BirthCountry”</i>

	<p>correspondence of values to birth country of actor was verified with text in other fields in the same file; also, key "Countries" was mutually exclusive with keys "DBD.CountryOfBirth" and "BirthCountry"</p>
<i>MonoLingualNonDutch</i>	<p>added to CMDI profile</p> <p>renamed to "BirthCountry"</p> <p>text string "MonoLingualNonDutch" occurs very often (n>900) in the corpus, as value of field "DBD.LanguageMode". very probable that it has been used erroneously as name of key. correspondence of values to birth country of actor was verified with text in other fields in the same file (a.o. "AgeAtImmigration")</p>
<i>DBD.AgeAtImmigration</i>	<p>added to CMDI profile</p> <p>renamed to "AgeAtImmigration"</p>
<i>DBD.LevelOfBilingualism</i>	<p>added to CMDI profile</p> <p>renamed to "LevelOfBilingualism"</p>
<i>DBD.LanguageMode</i>	<p>added to CMDI profile</p> <p>renamed to "LanguageMode"</p> <p>occurs in 2 different places in the profile</p>
<i>IMDI_1_9.Interactional</i>	<p>added to CMDI profile</p> <p>renamed to "Interactional"</p>
<i>IMDI_1_9.Discursive</i>	<p>added to CMDI profile</p> <p>renamed to "Discursive"</p>
<i>Date of Immigration in The Netherlands</i> <i>Date of Immigration in The Netherlands 2</i>	<p>added to CMDI profile</p> <p>renamed to "AgeAtImmigration"</p> <p>the field "AgeAtImmigration" was made unbounded to allow for multiple values (migrations).</p>

<i>Birth date</i>	<p><i>left out of CMDI profile</i></p> <p><i>n=4. key, including value, was redundant: each occurrence of "Birth date" co-occurred with an instance of "BirthDate" which corresponded in value.</i></p>
<i>School grade</i>	<p><i>added to CMDI profile</i></p> <p><i>renamed to "Grade"</i></p>
<i>topic</i>	<p><i>left out of CMDI profile</i></p> <p><i>added value to field "Description" with preceding text "The topic is".</i></p> <p><i>n=1. in this file the field "Description" was empty. the value of "topic" in this file corresponds to the value of the field "Description" in other files.</i></p>
<p><i>Turkish proficiency</i></p> <p><i>Turkish</i></p> <p><i>Kurdish profinciency (sic!)</i></p> <p><i>Dutch</i></p>	<p><i>left out of CMDI profile</i></p> <p><i>n=1: the 4 keys co-occurred in 1 file: Backus/bilingual/Metadata/DBD-TXX_16_28_03_001.imdi. information stated in these fields is represented elsewhere in the file, on a par with other files.</i></p>
<i>Berber</i>	<p><i>left out of CMDI profile</i></p> <p><i>n=1 and empty</i></p>
<i>Literacy</i>	<i>added to CMDI profile</i>