RESPONSES FROM DARREN BELL AT UKDA ARE **RED INLINE COMMENTS** AND INCORPORATED   
INTO v0.2 of the CDC PROFILE MARCH 6th 2020

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**From:** Katja Moilanen (TAU) [mailto:katja.moilanen@tuni.fi]   
**Sent:** 22 November 2019 08:48  
**To:** Zenk-Möltgen, Wolfgang <Wolfgang.Zenk-Moeltgen@gesis.org>; Förster, André <Andre.Foerster@gesis.org>; Borschewski, Kerrin <Kerrin.Borschewski@gesis.org>  
**Cc:** Taina Jääskeläinen (TAU) <taina.jaaskelainen@tuni.fi>  
**Subject:** MDO: DDI Profile pre-check

Dear Wolfgang, André and Kerrin,

I did a pre-check to CDC DDI Profile. Now, I first have to confess that I have not got any DDI Profile training, I have just studied DDI Profile specification on my own. So it might be possible that I have misunderstood some of the features of the DDI Profiles.

I used the google docs <https://docs.google.com/spreadsheets/d/1u9NSMvCwH1EMcpgKRDomZV9UW0tbDTrDZnr74vIgKY8/edit#gid=0> for checking but not the CV usage document <https://docs.google.com/spreadsheets/d/10yeLWJnXU6-mIQTs-dXMhYb4LvVbvtnb3L9bhuza-g0/edit#gid=0> because I'm quite sure that CV usage document is not implemented to CDC. If the specifications of CDC metadata is changed, SPs will need to have enough time to change their metadata accordingly.

I really like the way the CDC DDI Profile is organized; it is easy to read. It is also great that instructions of the usage etc. are added to this document.

My comments are in the attachment "MDO\_DDI\_Profile\_pre-check\_20191122.txt". I am glad that you coordinate the pre-check and collect all the comments for Darren.

Have a nice weekend!

Best regards,

Katja

Katja Moilanen

Tietoarkkitehti / Information Architect

Tietoarkisto / Finnish Social Science Data Archive

FI-33014 Tampereen yliopisto / FI-33014 Tampere University

[katja.moilanen@tuni.fi](mailto:katja.moilanen@tuni.fi)

+358 50 462 7370

[https://www.fsd.uta.fi](https://www.fsd.uta.fi/)

[https://services.fsd.uta.fi](https://services.fsd.uta.fi/)

First the "general" comments:

DDI Profile document

----------------------

The document is organized very nicely. It looks beautiful. It is easy to read. There are headings to distinct the different sections. It is great that there are instructions for usage and other useful information, too. Great work!

Structural elements

----------------------

If I have understood the DDI Profile specification right, there is no need to specify structural elements at all but only "content" elements. If DDI Profile specifies that codeBook element is used, it means that all its subelements are supported unless excluded using "NotUsed" element. Of course it is possible to do an implementation that interprets the specification differently. The DDI Profile would also be shorter, if the structural elements were dropped.

**Agreed. I have removed all “structural” elements for v0.2 of the profile. This means that any DDI document provided by the SP should validate against the DDI2.5 XSD schema (as well as the DDI 2.5 CDC Profile) to ensure that all structural elements are valid.**

isRequired and fixedValue attributes

-----------------------------------------

I suggest checking the isRequired-attributes thoroughly. I assume that if DDI Profile specifies an attribute as required (isRequired="true"), it means that the attribute must exist in the document even if its parent element would not exist in the document e.g. if there would not be any topcClas elements there still has to be at least one topcClas element with xml:lang attribute filled in (<topcClas xml:lang="en"></topcClas>). Of course it is, again, possible to decide to interpret this differently in an implementation.

**I have changed this so that if @isRequired=false for an element, then its children will also be false as well. Hopefully this will be less confusing.  
The** <r:Content>Required: **element will indicate if an attribute is required when its parent element is present:  
For example:**<pr:Used xpath="/codeBook/stdyDscr/citation/titlStmt/parTitl/@xml:lang" isRequired="false">  
<r:Description>  
<r:Content>Required: Mandatory if 'parTitl' element is present</r:Content>

If I have understood the meaning of fixedValue attributes right, I would suggest dropping the fixedValue attributes from the profile. I understood that they will overdrive the information from DDI document with the defaultValue, but I might have got this wrong.

**I have left the fixedValue attributes in. Basically, if you specify a defaultValue in the profile,  
then fixedValue should also be present.  
If fixedValue=true, then the SP CANNOT override the defaultValue.  
If fixedValue-false, then the SP CAN override the defaultValue**

Instructions element

---------------------

Each "Used" element has subelement:

<pr:Instructions>

<r:Content><![CDATA[<instruction type="example">instruction\_parameter</instruction>]]></r:Content>

</pr:Instructions>

could these be dropped?

**Agreed –these were “placeholders” in case we wanted to insert machine-readable parameters for the DDI Validator tool. I have removed them for now. They can be re-instated later if required.**

Language codes

----------------

For the attribute xml:lang, it is mandatory to use ISO 639-1 codes. ISO 3166 2-letter code is used for country only.

**All instances replaced in v0.2 of the profile apart from the “nation” element**

XPaths for attributes

-------------------------

Some XPaths (especially @vocab and @vocabURI) were missing slash character ("/") before the attribute e.g. /codeBook/stdyDscr/method/dataColl/timeMeth/concept@vocabURI would be better as /codeBook/stdyDscr/method/dataColl/timeMeth/concept/@vocabURI

**All corrected with forward slasges. To align with XPath 1.0 specification, all attributes in XPaths are in format of:  
elementname/@attributename**

CV fields topcClas, keyword, anlyUnit, timeMeth, sampProc and collMode

-------------------------------------------------------------------------

Topic classification and keyword have different structure in DDI-C than many other vocabulary field (e.g. anlyUnit, collMode, timeMeth, sampProc). topcClas and keyword do not have concept subelement at all. Therefore, the code/id of term is put in the ID attribute (which is not a perfect solution). anlyUnit, collMode, timeMeth and sampProc do have concept subelement and the code/id of the term is put in there. Therefore, anlyUnit, collMode, timeMeth and sampProc do not need concept/@ID at all.

"concept" element could have same usage text everywhere e.g. "Usage: Contains the value of the code (term ID) that is the same across languages".

The optional (not mandatory) vocabURI attribute in CDC is used mostly for the vocabulary URN. Its usage could be the same for all vocabulary fields e.g. "Usage: Use the URN of the CV version used". Because the value of vocabURI is version depended, it cannot have any default value.

Please note that only sampProc may have free text and topcClas, keyword, anlyUnit, timeMeth and collMode are used only with vocabulary.

**OK this a little bit messy. Because the CESSDA CV Manager does not currently provide working item-specific URNs or URIs (this is a separate issue to be discussed with MDO), I have removed @ID attributes altogether for sampProc, anlyUnit, timeMeth and collMode.**

**For sampProc, anlyUnit, timeMeth and collMode, I have taken the following approach**

|  |  |  |
| --- | --- | --- |
| Base element (sampProc or anlyUnit or timeMeth or collMode and topcClas) | Contains the local language CV item (except for sampProc, which can be freetext) i.e. the “Code descriptive term” in [CV Manager](https://vocabularies.cessda.eu/#!detail/AnalysisUnit?url=http%3A%2F%2Fwww.ddialliance.org%2FSpecification%2FDDI-CV%2FAnalysisUnit%2Fen%2F1.0) |  |
| Base/@xml:lang | The language of the item in the base element |  |
| Base/concept | The “code value”in [CV Manager](https://vocabularies.cessda.eu/#!detail/AnalysisUnit?url=http%3A%2F%2Fwww.ddialliance.org%2FSpecification%2FDDI-CV%2FAnalysisUnit%2Fen%2F1.0) – this is the same for all languages |  |
| Base/concept/@vocab | The Name of the Controlled Vocbaulary as in in [CV Manager](https://vocabularies.cessda.eu/#!detail/AnalysisUnit?url=http%3A%2F%2Fwww.ddialliance.org%2FSpecification%2FDDI-CV%2FAnalysisUnit%2Fen%2F1.0) e.g. “DDI Analysis Unit” |  |
| Base/concept/@vocabURI | The URN of the CV version e.g. “urn:ddi:int.ddi.cv:AnalysisUnit:2.0” |  |

This is an example:

<anlyUnit xml:lang="fi">Henkilö<concept vocab="DDI Analysis Unit"  
vocabURI="urn:ddi:int.ddi.cv:AnalysisUnit:2.0">Individual</concept>  
</anlyUnit>

And now to more detailed comments:

Root element

-------------

Attribute xml:lang is recommended to be used in DDI rather than xml-lang

Suggested amendment: <pr:Used xpath="/codeBook/@xml:lang" isRequired="true">

**Agreed – corrected from xml-lang to xml@lang**

Usually the URL is given.

Suggested amendment: <pr:Used xpath="/codeBook/@xsi:schemaLocation" defaultValue="ddi:codebook:2\_5 http://www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/codebook.xsd" fixedValue="true" isRequired="true">

**Agreed – updated in v0.2 of the profile**

Title

-------------

Only the title of the stdyDscr part (/codeBook/stdyDscr/citation/titlStmt/titl) is used for CDC. Definitions related to /codeBook/docDscr/citation/titlStmt/titl can be deleted.

**Agreed – these have been removed in v0.2 of the profile**

IDNo

------------

The same DDI2.5 field IDNo is used for two different purposes in CDC: first it is used for the study/archival number given by the SP and secondly it is used for PID. These two cases have different specifications. IDNo in the sense of study number is mandatory but the PID is not, so there could be guidance for this too e.g. in the usage. For PID, @xml:lang is mandatory but for Study number it is not. In the DDIProfile, @xml:lang is marked as mandatory. It could be presented so that "xml:lang" is not mandatory and note that it is mandatory for PID. For Study number the agency attribute is mandatory, but not for PID and this could be said in the usage too.

**I have done my best to put in a helpful usage note. My understanding is that there should always be one “IDNo’ element for the Study Number (xml:lang is not required). There \*may\* be a second ‘IDNo’ element for the PID. If so, then xml:lang is mandatory. I think @agency should always be mandatory to keep this simple.**

Principal Investigator (AuthEnty)

----------------------------

Usage texts have information about references. Nevertheless, DDI2.5 does not use references so the texts could be changed accordingly.

**Agreed – have removed “references” in the usage text**

Publisher (would it be possible to change the heading "Producer" to "Publisher"?)  
**Agreed – updated to “PUBLISHER”**

---------------------------------------------------------------------------------

This CDC field is for telling which SP takes care of this dataset. Therefore the DDI element used in CDC is /codeBook/docDscr/citation/prodStmt/producer that usually is the producer of the metadata. Producer in the studyDscr part is the producer of the original study.

Please note that the there is also another option for giving information about publisher for CDC: "distrbt" element. The publisher field is mandatory for CDC but because there are two options, /codeBook/docDscr/citation/prodStmt/producer and /codeBook/stdyDscr/citation/distStmt/distrbtr, both of them cannot be mandatory. It is possible to write XPath with two options e.g. xpath="/codeBook/docDscr/citation/prodStmt/producer | /codeBook/stdyDscr/citation/distStmt/distrbtr" with isRequired="true" or put two distinct Used elements with isRequired="false".

**Writing the XPath with 2 different options is not really desirable, To me, it seems that codeBook/stdyDscr/citation/distStmt/distrbtr is what we should use instead of the current “producer” element, so I have updated this in v0.2 of the profile.**

Year of publication

-----------------------

Only /codeBook/stdyDscr/citation/distStmt/distDate/@date is used for CDC, so other Used elements may be dropped. Year of publication is not mandatory in CDC, therefore, isRequired for distDate/@date should be false.

**Changed distDate element to “Required: Optional” and isRequired to false for the distDate/@date attribute**

Keyword

---------------------------------

keyword/@vocab is not mandatory in CDC.

**fixedValue="true" removed and isRequired changed to false**

Data collection dates

-----------------------

collDate may be used both for free text (e.g. <collDate>Spring 2007</collDate>) and for having an exact date in its date attribute (e.g. <collDate date="2007-01-01"></collDate>). collDate as a freetext element is optional not recommended. Only when collDate is used for free text the language attribute xml:lang is mandatory. For the same reason collDate/@date is not mandatory because there must be the possibility to use collDate for free text without date attribute collDate/@date : <r:Content>Required: Mandatory if 'collDate' element is present and not used as free text element</r:Content>

**Agreed – changed collDate element to “Required: Optional”. Changed @date description to “Required: Mandatory if 'collDate' element is present and empty”**

It would be better not have to defaultValue for collDate/@event - event may have 3 different values "start", "end" and "single". Please note that if the collDate is used as free text field, then the event attribute is optional: "Required: Mandatory if date attribute is present, optional if collDate is used as free text element"

**Agreed – I have removed the defaultValue=’start’ attribute, and the fixedValue=’false’ attribute”**

Unit of analysis

-----------------------

/codeBook/stdyDscr/stdyInfo/sumDscr/anlyUnit is content element, it is for the term in local language.

Proposed amendment:

<pr:Used xpath="/codeBook/stdyDscr/stdyInfo/sumDscr/anlyUnit" isRequired="false">

<r:Description>

<r:Content>Required: Recommended</r:Content>

<r:Content>ElementType: Content element</r:Content>

<r:Content>ElementRepeatable: Yes</r:Content>

<r:Content>Usage: Contains the term in local language</r:Content>

<r:Content>CDC UI Label: Analysis unit</r:Content>

</r:Description>

</pr:Used>

Proposed amendment to anlyUnit/@xml:lang : <r:Content>Usage: Language of the unit of analysis term</r:Content>

anlyUnit/concept/@vocab has quite long explination for the usage. There is some text that probably would work better in the anlyUnit and anlyUnit/concept usage text.

**Have attempted to simply this a little**

Sampling Procedure

-------------------------

Proposed amendments to sampProc element:

<r:Content>Required: Optional</r:Content>

<r:Content>Usage: Contains either the term in local language (when DDI Sampling Procedure vocabulary is used) or free text description of sampling procedure</r:Content>

**Agreed and implemented for v2.0**

Data Collection Method

------------------------

Proposed amendment to collMode element:

<r:Content>CDC UI Label: Data collection method</r:Content>

Proposed amendment to collMode/@xml:lang attribute:

<r:Content>Usage: Language of the data collection method term</r:Content>

Proposed amendment to collMode/concept element:

Delete: <r:Content>CDC UI Label: Data collection method</r:Content>

**Agreed and corrected for v2.0**

Language of the data file(s) (File Description)

-------------------------------------------------

There is no proper element/attribute for data file language in DDI2.5. This is why there are two potential attributes named to this purpose in CDC specification:

/codeBook/fileDscr/fileTxt/@xml:lang and /codeBook/fileDscr/fileTxt/fileName/@xml:lang . File language is recommended.

**I think to keep this simple, we should have ‘filename’ as the standard element, so standard usage for would be:** <fileDscr>  
 <fileTxt>  
 <fileName xml:lang="en">datafile.csv</fileName>  
 <fileName xml:lang="en">datafile2.csv</fileName>  
 </fileTxt>  
 </fileDscr>