DDI 2.1 to 3.0 Translation Notes

Translation Instructions

Read this document before using the 2_1TO3_0 spreadsheet contents. Certain element sets in DDI 2.1 are translated in different ways in different situations. This is particularly true of elements within <var>. For other element sets you may wish to preprocess your metadata to ensure consistent handling during translation or to identify specific set values.

General Structural Items

Because earlier versions of DDI were all backward compatible, version 2.1 contains the full list of available objects. Any comments regarding DDI 2.1 objects will apply to the same objects found in earlier versions.

DDI 2.1 contains:

793	attributes in addition to the %a.global attributes
6	EMPTY elements
50	elements that WRAP other elements but allow no PCDATA content
428	elements containing PCDATA %a.phrase or %e.form with the exception of element <mi> which allows only %a.phrase

All PCDATA elements allow PCDATA, %a.phrase or %e.form with the exception of element <mi> which allows only %a.phrase

The following PCDATA elements also allow table, an XML version of the Exchange Table Model version of CALS Tables (from OASIS) incorporated in DTD (Tables.dtd)

```
txt
notes
catStat
key
```

Translation of %a.phrase content to DDI 3.0:

DDI 2.1 contains the following definition for %a.phrase. "Link" is used to point to internal content and "ExtLink" to external.

ENTITY % a.phrase 'Link|ExtLink'

A.11 Link: This element permits encoders to provide links from any element containing Link as a subelement to other elements in the document. Obviously, it's based on XLink, but loosely.

Link EMPTY @%a.global @refs @role @title

A.12 ExtLink: This element permits encoders to provide links from any element containing ExtLink as a subelement to electronic resources outside the document.

ExtLink EMPTY @%a.global @URI @role @title

Links content needs to be handled based on the role of the link. If the link is for the purpose of replicating content, then the content should be transferred or reused when reuse is provided by the DDI 3.0 structure (e.g., Concept).

If the role is to provide an intellectual link between two pieces of content, create a Note in DDI 3.0 listing:

- "DDI 2.1 Link relationship" and the @title content to DDI 3.0 element r:Note/r:Content
- Both the containing DDI 3.0 element @id (or its parent identifiable) and the DDI 3.0 element @id for the linked element (or its parent identifiable) in r:Note/r:RelatedToReference/r:ID
- @role to r:Note/r:Relationship/r:RelationshipDescription of each Relationship

ExtLink may contain actual content (e.g., Standard Category content) or other materials specifically related to this element. Content should be evaluated as to its current structure (DDI or other) and either translated to DDI 3.0 as appropriate or listed as OtherMaterial with a link to the DDI 3.0 element or its parent identifiable.

- In the case of transferring this information to OtherMaterial:
- @URI to r:OtherMaterial/r:ExternalURLReference or r:ExternalURNReference as appropriate
- @title to r:OtherMaterial/r:Citation/r:Title (if no title one must be provided, e.g., "Untitled")
- @role to r:OtherMaterial/r:Relationship/r:RelationshipDescription
- ID of DDI 3.0 element or parent identifiable to r:OtherMaterial/r:Relationship/r:RelatedToReference/r:ID

DATES should be using the attribute @date with the format YYYY-MM-DD. Verify and transfer directly to 3.0. If CDATA data is also available, place in HistoricalDate.

All elements contain %a.global, a set of attributes including ID, xml-lang, xml:lang, and source.

xml-lang and xml:lang should both transfer directly to the xml:lang attribute of the corresponding element in 3.0. If all xml-lang/xml:lang attributes are identical, the attribute may be placed at the StudyUnit level and Instance level so that it is inherited throughout the instance. If both xml-lang and xml:lang are present, give precedence to xml:lang.

source is limited to archive and producer. First verify whether all attributes contain the same content (producer or archive). If so, producer is the source organization for the instance (and possible owner if XML is being created by the original producer).

All elements in 2.x and earlier versions of DDI may contain an **ID** expressed as an attribute. These can transfer directly or be translated to an alternative ID system.

ALL elements may be referenced by @refs from ALL other elements.

CHECK FOR USE OF <Link> with @refs

Parse to verify which IDs are referenced within the document in order to identify any that may be lost when translated to 3.0.

The following list provides the name and location of attributes of type IDREF or IDREFS. The spreadsheet lists all elements intended for reference by the attribute that references them. All elements containing any attribute with a value of IDREF(S) is also identified in the spreadsheet. The list below provides the type of IDREF by its name and location in 2.1.

Note that sdatrefs may have been used locally to provide an IDREF option for locally modified fields like labl, txt, and notes due to their broad availability within the dtd.

Items indicated with "A" may be referenced by @access from elements

2.0
3.0
4.1
4.2
4.3
4.4

Items indicated with "B" may be referenced by @sdatrefs from elements

3.0
4.1
4.2
4.3
4.3.8
4.3.8.2
4.3.18.4
4.4
4.4.11
A.2 labl
A.3 txt
A.4 notes

Items indicated with "C" may be referenced by @methrefs from elements

3.0
4.1
4.2
4.3
4.3.18.4
4.4
4.4.11

Items indicated with "D" may be referenced by @pubrefs from elements

3.0
4.1
4.2
4.3
4.4
4.4.11

Items indicated with "E" may be referenced by @recGrp from elements

3.1.3
3.1.3.1

Items indicated with "F" may be referenced by @keyvar from elements

3.1.3
3.1.3.1

Items indicated with "G" may be referenced by @varRef from elements

3.2.1
4.3.18.5.1
4.4.12
4.4.13

Items indicated with "H" may be referenced by @nCubeRef from elements

3.2.1

Items indicated with "I" may be referenced by @coordValRef from elements

3.2.1.1

Items indicated with "J" may be referenced by @recRef from elements

3.2.1.2

Items indicated with "K" may be referenced by @var from elements

4.1
4.3.8
4.3.22

Items indicated with "L" may be referenced by @varGrp from elements

4.1

Items indicated with "M" may be referenced by @nCube from elements

4.2

Items indicated with "N" may be referenced by @nCubeGrp from elements

4.2

Items indicated with "O" may be referenced by @wgt-var from elements

4.3
4.3.14
4.3.18.4

Items indicated with "P" may be referenced by @weight from elements

4.3
4.3.14
4.3.18.4

Items indicated with "Q" may be referenced by @qstn from elements

4.3
4.3.8.4

4.3.8.5

Items indicated with "R" may be referenced by @files from elements

4.3

Items indicated with "S" may be referenced by @fileid from elements

4.3.1

Items indicated with "T" may be referenced by @locMap from elements

4.3.1
4.4.1

Items indicated with "U" may be referenced by @catgry from elements

4.3.17
4.3.18

Items indicated with "V" may be referenced by @catGrp from elements

4.3.17

Items indicated with "W" may be referenced by @level from elements

4.3.18

Items indicated with "X" may be referenced by @catRef from elements

4.4.12.1

@name is found in var, varGrp, nCube, and nCubeGrp. It is only required in var. The spreadsheet recommends two options for translation, to the Name element of Identification content or as another Label within the parent Variable, VariableGroup, NCube, or NCube group. In examining current DDI 2.1 instances I have found inconsistent use of the field across studies. However, there is general consistency in how this is used within an organization. When translating documents, determine the intent of the content of the DDI 2.1 attribute and translate to Label or Name accordingly.

The element **concept @vocab @URI** is found as an option for PCDATA content in the following DDI 2.1 elements: nation, geogCover, geogUnit, anlyUnit, universe, dataKind, timeMeth, sampProc, collMode, resInstru, and srcOrig. There is no documentation on the intent of this option in these situations, just the standard definition of concept. This may have been used as a means of providing an external controlled vocabulary or of assigning a concept to a specific element. To date we have not identified anyone who has used this structural option. If the DDI 2.1 instance used concept in any of these locations, determine its use in each situation and translate as appropriate. This could be translated to Concept or a subelement or attribute of the parent that is of DDI 3.0 CodeValueType.

The element **txt @level @sdatRef** is found as an option for PCDATA content in the following DDI 2.1 elements: nation, geogCover, geogUnit, anlyUnit, universe, dataKind, timeMeth, sampProc, collMode, resInstru, and srcOrig. Text PCDATA content is handled as normal PCDATA content of the parent element, level is not applicable in DDI 3.0 and the handling of sdatRef has been described.

Common Elements in DDI 2.1 Requiring Consistent Processing

All PCDATA elements allow %e.form. Some of these will transfer into 3.0 fields that do not allow XHTML format. Check the following 2.1 elements for formatting and remove the format, editing where needed for contextual sense.

CREATE standard mapping of html tags allowed in Version 2.1 to Version 3.0 XHTML tags for structured text

> table
> div
> emph
> head
> hi
> itm
> list
> p

Defaults to Declare:

- Agency identification and content for organization definition
- xml:lang
- producer definition (from @source)
- archive definition (from @source)
- Create list of notes @type contents and map to NoteTypeCodeType values OR declare default

NOTES

Version 2.1

notes
@ID
@xml:lang
@source
@type
@subject
@level
@resp
@sdatrefs
     PCDATA

Version 3.0

Note
     Subject 0..1 (xs:string)
     Relationship 1..n
      RelatedToReference 1..1 (ReferenceType)
      RelationshipDescription 0..n (InternationalStringType)
     Responsibility 0..1 (xs:string)
     Header 0..1 (xs:string)
     Content 1..1
@id
@xml:lang
@NoteTypeCodeType [Processing | Footnote | Addendum | System | Problem | Comment | Other] use required

TRANSLATION Instructions

The default location for all notes is s:StudyUnit/r:Note. The spreadsheet provides module specific locations based on the location of the original notes field in 2.1.

@ID	@id [generate if not provided]
@xml-lang/@xml:lang	@xml:lang [use default if not provided]
Concatenate @source (using default definition) and @resp	Responsibility (e.g., <notes source="archive" resp="Wendy Thomas"> and <r:Responsibility>archive - Wendy Thomas</r:Responsibility>)
@type map	Predetermined value to @NoteTypeCodeType or default to "Other"
@subject	Subject
@level [use	Verify parent level for disposition to correct module if available]
@sdatrefs	RelatedToReference AND add "Study description

	item reference" to RelationshipDescription
Transfer ID of parent identifiable	RelatedToReference add "parent element" to RelationshipDescription
PCDATA content	Content

A listing of identifiable parents for DDI 2.1 note fields is provided in DDI2_1_to_3_0.xls/Note Parents sheet. This includes the location of all notes in 2.1, the location of the parent element, and a default identifiable parent path in DDI 3.0.

OTHER MATERIAL

Other material in 2.1 allows for an option of PCDATA or %a.phrase (a Link or ExtLink), %e.form, and/or citation. If citation does not exist, the information in the PCDATA or %e.form content must be parsed to create a citation with minimum content. Alternatively, all information can be placed in default s:StudyUnit/r:OtherMaterial/r:Relationship/r:RelationshipDescription. s:StudyUnit/r:OtherMaterial/r:Citation/r:Title is required and can contain a "See RelationshipDescription for content" as an informative place holder for later cleanup.

Description of 2.1 content indicates "Can take the form of bibliographic citations. This element can contain either PCDATA or a citation or both, and there can be multiple occurrences of both the citation and PCDATA within a single element. May consist of a single URI or a series of URIs comprising a series of citations/references to external materials which can be objects as a whole (journal articles) or parts of objects (chapters or appendices in articles or documents)." Verify the content of any othrStdyMat and otherMat content in your original document and pre-parse into separate entries.

Content fields transferred to r:OtherMaterial/r:Citation fields may also be copied to their related DCElements fields (dc:title, dc:creator, dc:subject, dc:decription (PCDATA content), dc:publisher, dc:contributor, dc:date, dc:format, dc:identifer, dc:source, dc:language, dc:relation, dc:coverage, dc:rights) if available.

Note that relMat contains four additional attributes: @callno, @label, @media, and @type. There is no description of what these should contain and they are not relevent to the DDI 3.0 structure of OtherMaterial except as follows:

- If callno is a recognizable standard type (e.g., Dewey Decimal or Library of Congress in the U.S.) this information could be captured in r:OtherMaterial/r:Citation/r:InternationalIdentifier using the predetermined format for the required @type.
- callno could place in r:OtherMaterial/r:Citation/r:DCElements/dc:identifier.
- If @type contains a descriptor for the type of other material this may be used in the r:OtherMaterial@type.
- If type indicators are not available in the DDI 2.1 instance, use the element name of the parent item (e.g., relMat, relPubl, etc.) as the required r:OtherMaterial@type to provide general information on the Other Material type as originally designated.

9

DDI 2.1 Other Material found in 5.0 and nested 5.0 element as well as the content of Citation

titlStmt
    titl
    subTitl
    altTitl
    parTitl
    IDNo
        @agency
        @level
rspStmt
    AuthEnty
        @affiliation
    othId
        @type
        @role
        @affiliation
prodStmt
    producer
        @abbr
        @affiliation
        @role
    copyright
    prodDate
        @date
    prodPlac
distStmt
    distrbtr
        @URI
    contact
        @affiliation
        @URI
        @email
    depositr
        @abbr
        @affiliation
    depDate
        @date
    distDate
        @date
serStmt
    @URI
    serName
        @abbr
    serInfo
verStmt
    version
        @date

```
                @type
        verResp
                @affil


Version 3.0

OtherMaterial

    Citation
            Title 1..n InternationalString
                    @xml:lang 1..1
                    @translated default="false"
                    @translatable default="true"
            SubTitle 0..n InternationalString
                    @xml:lang 1..1
                    @translated default="false"
                    @translatable default="true"
            AlternateTitle 0..n InternationalString
                    @xml:lang 1..1
                    @translated default="false"
                    @translatable default="true"
            Creator 0..n
                    InternationalString
                    @xml:lang 1..1
                    @translated default="false"
                    @translatable default="true"
                    @affiliation 0..1 xs:string
            Publisher 0..n InternationalString
            Contributor 0..n
                    InternationalString
                    @xml:lang 1..1
                    @translated default="false"
                    @translatable default="true"
                    @role 0..1 xs:string
                    @affiliation 0..1 xs:string
            PublicationDate 0..1
            Language 0..1 xs:string
            InternationalIdentifier 0..n TypedStringType
            Copyright 0..1 InternationalString
                    @xml:lang 1..1
                    @translated default="false"
                    @translatable default="true"
            dc:DCElements 0..1
    ExternalURLReference 0..1 xs:anyURI
    ExternalURNReference 0..1 xs:anyURI
    Relationship 0..n
```

```
        RelatedToReference 1..1 (ReferenceType)
        RelationshipDescription 0..n (InternationalStringType)
                @xml:lang 1..1
                @translated default="false"
                @translatable default="true"
        MIMEType 0..1 xs:string
@id
@xml:lang
@type 1..1 xs:string
```

TRANSLATION Instructions

@ID of top available level in the DDI 2.1 Other Material generic to @id of OtherMaterial in DDI 3.0

titlStmt, @level, rspStmt, prodStmt are wrappers and do not translate to 3.0

Default location of OtherMaterial should be StudyUnit. If your original documents are well structured in terms of automatically identifying material that should be located in other modules, you may wish to designate specific modules for those OtherMaterial contents (e.g., relPubl contents might go in a:Archive/r:OtherMaterial to provide ease of updating without having to version the s:StudyUnit maintainable object).

Titl	Title
subtitle	SubTitle
alttitl	AlternateTitle
parTitl	Title making use of @xml:lang to designate alternate language
IDNo IF @agency is an international identifier (ISBN, ISSN, etc.)	IDNo to InternationalIdentifier, @agency to InternationalIdentify/@type
AuthEnty	Creator
AuthEnty/@affiliation	Creator/@affiliation
othID	Contributor
othID@type	NO HOME - no explanation of the content of @type in DDI 2.1
othID@role	Contributor@role
othID@affiliation	Contributor@affiliation
producer	concatenated with prodPlac to Publisher
producer@abbr	NO HOME (an Organization entry for the producer can be created)
producer@affiliation	NO HOME (an Organization entry for the producer can be created)
producer@role	NO HOME (an Organization entry for the producer can be created)
copyright	Copyright
prodDate	The non-ISO date will need to be translated if not available from the data attribute
prodDate@date	PublicationDate/SimpleDate
distrbtr@URI	ExternalURLReference or ExternalURNReference depending on type

Organization Scheme (Organization and Individual Entries)

Preview the content of the DDI 2.1 instance and identify all individuals or organizations and the amount of information available regarding them (name, affiliation, abbreviation, email, URI, etc.). Create a DDI 3.0 OrganizationScheme (a:Archive/a:OrganizationScheme), nesting individual in organizations and organizations in parent organizations (or referencing a relation to an organization) in advance. This will facilitate the assignment of r:ReferenceType content when creating the DDI 3.0 version of the documentation.

Determining Variable Representation Type

DDI 2.1 does not cleanly differentiate between the representation types available in DDI 3.0. Variables with category codes may also provide information on the number range of the valid codes. References to external coding structures through standard category descriptions may not be explicit or may point to information in a wide variety of formats. The following process will provide some consistency in translating 2.1 variables to the proper 3.0 variable representation type.

1. Determine if the 2.1 variable is used as a dimension of an nCube. If so these should be held in an l:NCubeLogicalProduct/l:VariableScheme. All of these variables will be l:CodeRepresentation. Note that not all variables in an DDI 2.1 file using nCubes are used as dimensions in nCubes. If the nCubes are replicated for a large number of areas, such as in a Census file, there may also be variables that provide a geographic or other record identifying code string. These may be of any representation type.
2. Does the variable use a standard category (stdCatgry)? If yes, either create/locate a DDI 3.0 version of the standard category content and reference it from an l:CodeRepresentation OR use l:TextRepresentation providing the following attributes:
    2.1. maxLength using the 2.1 content of codebook/dataDscr/var/location@width
    2.2. minLength using the 2.1 content of codebook/dataDscr/var/location@width if the field has a standard length, otherwise do not include
    2.3. regExp containing the allowed syntax
    2.4. EXAMPLE: United States 5-Digit ZIP Code @maxLength="5" @minLength="5" @regExp="(0-9)*"
3. If NOT a standard category, does the variable use category (catgy)? If yes, use l:CodeRepresentation creating both a set of l:LogicalProduct/l:CategoryScheme/l:Category objects and the l:LogicalProduct/l:CodeScheme.
    3.1. If catgryGrp elements are used, these should be entered as l:Category rather than as l:CategoryGroup as these do not allow for labels. The resulting l:CodeScheme will be hierarchical. Use l:CodeScheme/l:Level and the associated level attributes on l:Code to allow inclusion of only those categories that originated from DDI 2.1 catgry fields to be included. This may also be done using the attribute isDiscrete which identifies the lowest level. Restrict l:CodeRepresentation to include only those levels with data.
        3.1.1. EXAMPLE: A 2.1 Variable "labor force" with the following irregular hierarchy structure:

LABEL	DDI 2.1	DDI 3.0	Code notations
In labor force:	catgryGrp	Category	[no data] Level 1 isDiscrete="false"
Employed	catgry	Category	[data] Level 2 isDiscrete="true"
Unemployed	catgry	Category	[data] Level 2 isDiscrete="true"
Not in labor force	catgry	Category	[data] Level 1 isDiscrete="true"

```
<l:Variable>…
      <l:Representation>…
            <l:CodeRepresentation>
                  <r:CodeSchemeReference> reference to CodeScheme
                  <l:CodeSubsetInfo>
                        <l:DataExistence>
                              <l:DiscreteCategory>true</l:DiscreteCategory>
                        </l:DataExistence>
                  </l:CodeSubsetInfo>
            </l:CodeRepresentation>
      </l:Representation>
</l:Variable>
```

3.1.2.EXAMPLE: A 2.1 Variable "Household Type" with the following regular hierarchy structure:

LABEL	DDI 2.1	DDI 3.0	Code notations
Household:	catgryGrp	Category	[no data] Level 1 isDiscrete="false"
Family	catgry	Category	[data] Level 2 isDiscrete="true"
Non-Family	catgry	Category	[data] Level 2 isDiscrete="true"
Group Quarters:	catgryGrp	Category	[data] Level 1 isDiscrete="false"
Institution	catgry	Category	[data] Level 2 isDiscrete="true"
Non-Institution	catgry	Category	[data] Level 2 isDiscrete="true"

```
<l:Variable>…
      <l:Representation>…
            <l:CodeRepresentation>
                  <r:CodeSchemeReference> reference to CodeScheme
                  <l:CodeSubsetInfo>
                        <l:DataExistence>
                              <l:Level>2</l:Level>
                        </l:DataExistence>
                  </l:CodeSubsetInfo>
            </l:CodeRepresentation>
      </l:Representation>
</l:Variable>
```

3.2. If catgry elements are nested, they should be entered as l:Category and the l:CodeScheme used to express the nesting levels and patterns. Additional level name information may be available in catLevel@levelnm. The resulting DDI 3.0 Variable would use the full CodeScheme without use of the l:CodeSubsetInfo.

4. If NOT a CodeRepresentation, is the attribute var@isTemporal="Y" OR does 4.3.23 format@category equal "date" or "time"?

4.1. Evaluate the content and use l:DateTimeRepresentation if the content is consistently structured.

4.2. If not consistently structured used l:NumericRepresentation or l:TextRepresentation as appropriate.

5. If NOT a DateTimeRepresentation, use l:NumericRepresentation if any of the following are true?

5.1. Does the attribute var@decml, var@scale, or var@origin contain content?

5.2. Is attribute var@intvl="contin"? [note that this has a default value of "discrete" in DDI 2.1]

5.3. Is attribute var@wgt = "wgt"?

5.4. Is 4.3.23 format@type="numeric"? [note that this has a default value of "numeric" in DDI 2.1]

6. If NOT a NumericRepresentation, use l:TextRepresentation.

Name attribute, label and text in DDI 2.1

The translation of the name attribute in DDI 2.1 as found in variable (var), nCube, variable group (varGrp), and nCube group (nCubeGrp) has been the source of considerable discussion. DDI 2.1 documentation defines name as a "unique ID" and is required for variable although not in other locations. In addition, the documentation for name in variable states "Following the rules of many statistical analysis systems such as SAS and SPSS, names are usually up to eight characters long."

In DDI 3.0 Name is a repeatable element in all identified elements with the documentation "Human-readable name given the entity being identified. May be repeated to provide language and/or geographic alternatives." As such it has a different usage than the attribute name in DDI 2.1. DDI 3.0 Label as defined by LabelType in reusable has the following definition, "An unstructured label for the element. DDI does not impose any length limitations on Label. If length of Label is constrained due to use of the element in a specific application, the maximum length supported should be noted in the attribute maxLength. Label may be repeated to provide content for systems that have length constraints (e.g., some versions of the following statistical packages have character length limits: SAS 40-character, SPSS 120 characters, and Stata 80 characters). Mnemonics associated with a variable should include both of the following attributes: type="Nickname" and maxLength. Software packages will use this information to select the appropriate mnemonic for use." Note that this definition includes the common usage of the attribute name in DDI 2.1.

Similarly, text (txt) in DDI 2.1 is defined in variable as "An extended description, beyond that provided in Variable Name and Label, of the variable." Its usage has also been to contain an extended label, for example in one of the common SPSS to DDI converters labels over a certain length were routinely set to txt rather than labl so that the variable would have either a label OR a text but not both.

The confusion over these similarly named but differently defined elements means that one must determine the specific usage of these elements in the DDI 2.1 instance prior to deciding where the information should be moved. In general, the spreadsheet has provided options. Future versions of DDI will seek to clarify these differentiations. With this in mind the following approach is recommended:

DDI 2.1	DDI 3.0
Attribute name	If the attribute contains a locally unique identifier related to a specific system, place the content in Name. If the content is a commonly used mnemonic within a study or series, place the content in Label with the type indicator "Nickname" and maxLength if appropriate. This will provide an easy means of identifying a standard mnemonic for any future version changes in DDI.
labl	Label including maxLength and type indicator when appropriate and available.
txt	If this is an extended label, place content in a second Label element providing a type differentiation indicating either its source or usage for example "extended label". If this is a description, place the content in description, creating a second Description element if needed or making it an additional paragraph in the structured content of Description.