What’s new with DDI 3.2 and beyond

Jeremy Iverson
Wendy Thomas
NADDI 2014
Agenda

- Changes and New Content in 3.2
  - Changes in content that effect decisions to stay with an earlier DDI version or use 3.2
- Improvements for Developers
  - Structural and other improvements that support the work of developers
- Moving Forward
  - The projects, sprints, activities to date, and where to find the details
Changes and New Content in 3.2
Changes in the structure of existing content

- Bundling language equivalents in a single element
- Representations and missing values
- Managing identification content
- Name clashes
- New Code Value Types
Changes in the structure of existing content

- Bundling language equivalents in a single element
  - Make sure multiple language content is complete
  - Note where cardinality has been constrained in 3.2
  - Loss of Identified Structured String – when multiple languages are involved
- Representations and missing values
- Managing identification content
- Name clashes
- New Code Value Types
Changes in the structure of existing content

- Bundling language equivalents in a single element
- Representations and missing values
  - Valid and Invalid (Missing) Values are handled separately in 3.2 – and probably into the future
  - Create CodeSchemes that combine a CodeScheme of valid values and a CodeScheme of invalid (missing) values
  - Managed Representations (Numeric, Text, DateTime, ...)
- Managing identification content
- Name clashes
- New Code Value Types
Changes in the structure of existing content

- Bundling language equivalents in a single element
- Representations and missing values
- Managing identification content
  - Use URN’s! This is the only place where all the content of the full object identification can be expressed for Identifiable and Versionable types
- Name clashes
- New Code Value Types
Changes in the structure of existing content

• Bundling language equivalents in a single element
• Representations and missing values
• Managing identification content
• Name clashes
  ▫ Note where object names are changing (these may be namespace specific)
• New Code Value Types
Changes in the structure of existing content

- Bundling language equivalents in a single element
- Representations and missing values
- Managing identification content
- Name clashes
- New Code Value Types
  - Many attributes were changed to elements to support external controlled vocabularies (CodeValue\_Type)
  - Constrain current content of these fields to values you plan to provide in an external controlled vocabulary
New Content

• Complex Questionnaires
  ▫ Question Grids and Question Blocks
  ▫ Data Processing
  ▫ New representations and response domains
• Concept Management
  ▫ Conceptual modeling
  ▫ Managed representations and missing (invalid) values
• Content Management
  ▫ Quality and authority statements
  ▫ Reorganizing Organizations and Individuals
  ▫ Archiving submitted content
• Consistency
  ▫ Multilanguage content
  ▫ Geographic content consistency
  ▫ Scheme structure consistency
Complex Questionnaires

• The primary reason for using 3.2
• Complex questionnaires are defined as those using:
  ▫ Complex Question Grids
  ▫ Question Blocks - where a set of questions are related to a specific stimulus item (common in educational testing)
  ▫ Need for a broader set of response domains
    • Scales, capturing a mark on an object, geographic codes, ranking, distribution, etc.
  ▫ Complex routing and reuse of data for calculations
    • Incorporating pre-loaded content
    • Extensive use of response information to inform the content of dynamic text
    • Data validation done during the collection process
Question Grids

- Simple grids where a single response type such as a Value or “Yes/No” response is required for each item in a list
- Complex grids where the list may be provided or entered by the respondent and multiple responses may be required for each item in a list (“Yes/No” plus a Value plus a Percent, etc.)
Did you visit any one of the following performances or facilities over the past 12 months? (Circle the correct response for each)

Yes  No  a theatre performance  
Yes  No  a cabaret performance  
Yes  No  a concert of classical music  
Yes  No  an opera or operetta  
Yes  No  a concert of popular music, pop, jazz, musical or pop opera  
Yes  No  a 'dance' event, houseparty  
Yes  No  a ballet performance  
Yes  No  the cinema  
Yes  No  an art gallery  
Yes  No  a museum
14. Please write down the names of the places you stayed overnight and the duration of the stays in chronological order separately.

<table>
<thead>
<tr>
<th>Name of the place (City, town, etc.)</th>
<th>Total number of nights</th>
<th>Hotel</th>
<th>Motel</th>
<th>Village</th>
<th>Pension</th>
<th>Camp/ Caravan</th>
<th>Youth And Camp</th>
<th>Own House</th>
<th>Rented House</th>
<th>House of a relative or friend</th>
<th>In the vehicle (Cruise, Yacht, Train, TIR)</th>
<th>Other (Please specify)</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Question Blocks

• Contains “Stimulus Material” (image, sound, text, etc.) to which a set of questions pertain.
• The set of specific questions generally make sense only in context of the stimulus material.
• Commonly used in educational testing
• Difference between a Question Block and sequence of questions is the need to tightly tie the questions to a specific stimulus material in order to assess the comparability of the question content
QUESTION 9.4
The director positions the actors on the stage. On a diagram, the director represents Amanda with the letter A and the Duchess with the letter D.
Put an A and a D on the following diagram of the set to show approximately where Amanda and the Duchess are when the Prince arrives.

QUESTION 9.5
Towards the end of the extract from the play, Amanda says, "He didn’t recognise me...". What does she mean by that?
A. That the Prince didn’t look at Amanda.
B. That the Prince didn’t realise that Amanda was a shop assistant.
C. That the Prince didn’t realise that he’d already met Amanda.
D. That the Prince didn’t notice that Amanda looked like Leocadia.
New representations/response domains

- Response domains needed to capture information in specialized situations:
  - Marks on an image or audio/visual
  - Ranking behaviors
  - Controlled distributions

- Ability to use Geographic Structure and Geographic Location content directly as a specialized code representation

- Scales as response domains and value representations
  - The actual layout of scales may effect the accuracy of the response
I am aware of the presence of God or the Divine

<table>
<thead>
<tr>
<th>Never</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contradiction</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Attraction</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Repulsion</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Indifference</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Strongly Disagree</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
Concept Management

• DDI is not fully there but 3.2 dramatically improves support for concept management and expression
• Alignment with GSIM content regarding Concepts, Conceptual Variables, and Represented Variables
• Managed representations allow numeric, textual, datetime and other representations to be captured once and reused
  ▫ BONUS: managed representations can be mapped for comparison, i.e. a numeric representation of age to a coded set of age cohorts
Concept

Universe

Conceptual Variable

Represented Variable

Representation

Variable

May constrain the Universe and Representation for a specific applied usage
Content Management: Quality

- Capture information on Quality Frameworks, Quality Standards, and other means of assuring the quality of the data, metadata, and process.
- Quality statements can be referenced from many levels within DDI
- Additional content to reflect authorization, legal status, budget, and evaluation procedures have been added
Content Management: Actors/Agents

- Organization Scheme content has been cleaned up and stripped down
- Organization and Individual information can be better managed over time (specific contact information is time stamped, etc.)
- Relationships between any two Organizations and/or Individuals is tracked separately under “Relation”. Roles are assigned as part of the Relation
- Locations in DDI where names are associated with specific roles within the context of the Study can now reference an Individual or Organization
Content Management: Archiving

• Supports inclusion of all major packaging types as submitted (deposited) content (StudyUnit, Group, ResourcePackage)
• Supports Local Added Content of all packaging types
• Links Local Added Content to Submitted (Deposited) content by mapping
  ▫ Local Object Reference
  ▫ Depository Object Reference
  ▫ Relationship Action (CodeValueType) i.e. Overrides, Deletes, Extends, Restricts, etc.
  ▫ Description
Language Consistency

• Bundling equivalent content expressed in multiple languages

<r:Label xml:lang="en">What is it?</r:Label>
<r:Label xml:lang="de">Was ist das?</r:Label>

<r:Label>
  <r:String xml:lang="en">What is it?</r:String>
  <r:String xml:lang="de">Was ist das?</r:String>
</r:Label>
Geographic content consistency

- Added composite geographies (Metropolitan Areas – composite of counties)
- Improved management of geographic change over time
- Added ability to use geographic structures (names and codes) and geographic locations (names and codes) directly as representations or response domains
- Added ability to reference specific locations for the purpose of defining Spatial Coverage
Scheme structure consistency

- All schemes have the following structural features
  - `XxxxSchemeName`, Label, Description
  - Inclusion of a scheme by reference
  - Object type included in-line or by reference
  - Object Group included in-line or by reference
- All schemes have an `XxxxGroup` to support grouping for administrative or conceptual purposes
  - `XxxxGroupName`, Label, Description, and `TypeOfXxxxGroup`
  - Universe and Concept references
  - Subject and Keyword
  - Group and object included by reference
What if I’m just starting out?

- A number of new tools are already available for 3.2
  - Colectica
  - Rogatus Platform
- If DDI-C or DDI-L 3.1 does it for you and there are tools to support your work, USE THEM
  - But....pay attention to the consistency changes made in 3.2 to make your work easy to migrate forward. These changes were made in anticipation of future version needs.
Improvements for Developers
What’s New in DDI 3.2 and Beyond
Controversial Thesis

DDI 3.2 is Simple

DDI 3.2 is Powerful
## Document a Single Variable with DDI 3.1

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>casid</td>
<td>age</td>
<td>marital</td>
<td>company</td>
<td>country</td>
<td>rank</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>85</td>
<td>Married</td>
<td>Vitae Posuere Corp.</td>
<td>Tunisia</td>
<td>4.5882396116</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>29</td>
<td>Common-Law</td>
<td>Luctus Consulting</td>
<td>Isle of Man</td>
<td>5.251625872</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>70</td>
<td>Common-Law</td>
<td>Purus Associates</td>
<td>Uruguay</td>
<td>5.327125737</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>76</td>
<td>Common-Law</td>
<td>Pellentesque Tellus Sem Institute</td>
<td>Tunisia</td>
<td>4.872900924</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>32</td>
<td>Single</td>
<td>Tempus Non Lacinia Incorporated</td>
<td>Romania</td>
<td>5.007552577</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>33</td>
<td>Divorced</td>
<td>Aenean Inc.</td>
<td>Tonga</td>
<td>4.716042835</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>Single</td>
<td>Sum Socites Natoque Limited</td>
<td>Montserrat</td>
<td>5.541866538</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>49</td>
<td>Common-Law</td>
<td>UT LLP</td>
<td>Thailand</td>
<td>4.850531953</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>42</td>
<td>Common-Law</td>
<td>Cum Sociis Limited</td>
<td>Chad</td>
<td>5.283005818</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>26</td>
<td>Single</td>
<td>Eget Company</td>
<td>Croatia</td>
<td>5.019311745</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>42</td>
<td>Married</td>
<td>Arcu LLP</td>
<td>Saint Vincent and the Grenadines</td>
<td>5.154959312</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>89</td>
<td>Common-Law</td>
<td>Auctor Vitae Incorporated</td>
<td>Saint Vincent and the Grenadines</td>
<td>5.212204562</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>61</td>
<td>Single</td>
<td>Lorem Auctor Quis Inc.</td>
<td>Philippines</td>
<td>4.739571551</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>81</td>
<td>Common-Law</td>
<td>Luctus Ipsum Leo LLP</td>
<td>Kazakhstan</td>
<td>5.066390012</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>82</td>
<td>Common-Law</td>
<td>Neque Pellentesque LLP</td>
<td>Sri Lanka</td>
<td>4.897902</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>29</td>
<td>Common-Law</td>
<td>Rhoncus Corporation</td>
<td>Seychelles</td>
<td>4.9453120184</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>28</td>
<td>Divorced</td>
<td>Luctus Felis Ltd</td>
<td>South Africa</td>
<td>4.9412923129</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>24</td>
<td>Single</td>
<td>Dolor Foundation</td>
<td>Tokelau</td>
<td>5.145752551</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>58</td>
<td>Single</td>
<td>Odio Etiam Incorporated</td>
<td>Guam</td>
<td>4.39812833</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>33</td>
<td>Single</td>
<td>Mauris Ut LLP</td>
<td>Bosnia and Herzegovina</td>
<td>5.265284389</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>31</td>
<td>Common-Law</td>
<td>Tempus Inc.</td>
<td>Japan</td>
<td>5.265583113</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>66</td>
<td>Single</td>
<td>Quam Elementum At LLP</td>
<td>Laos</td>
<td>4.894495887</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>Divorced</td>
<td>Sem Consequit Foundation</td>
<td>Sierra Leone</td>
<td>5.168156291</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>85</td>
<td>Divorced</td>
<td>Elefend Nunc Limited</td>
<td>Iran</td>
<td>4.915958645</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>43</td>
<td>Single</td>
<td>Aenean Institute</td>
<td>Luxembourg</td>
<td>4.589390092</td>
<td></td>
</tr>
</tbody>
</table>
<ddi:DDIInstance id="ddi1" version="1.0.0" agency="int.example">
  <r:Citation>
    <r:Title xml:lang="en">MyDdiInstance</r:Title>
  </r:Citation>
  <g:ResourcePackage id="rp1">
    <r:Citation>
      <r:Title xml:lang="en">MyResourcePackage</r:Title>
    </r:Citation>
    <g:Purpose id="purpose1">
      <r:Content xml:lang="en">Not specified</r:Content>
    </g:Purpose>
    <l:VariableScheme id="varscheme1">
      <r:Label xml:lang="en">MyVariableScheme</r:Label>
      <l:Variable id="var1">
        <l:VariableName xml:lang="en">age</l:VariableName>
        <r:Label xml:lang="en">Age of the respondent</r:Label>
        <l:Representation>
          <l:NumericRepresentation type="Integer" />
        </l:Representation>
      </l:Variable>
    </l:VariableScheme>
  </g:ResourcePackage>
</ddi:DDIInstance>
<ddi:DDIInstance id="ddi1" version="1.0.0" agency="int.example">
  <r:Citation>
    <r:Title xml:lang="en">MyDdiInstance</r:Title>
  </r:Citation>
  <g:ResourcePackage id="rp1">
    <r:Citation>
      <r:Title xml:lang="en">MyResourcePackage</r:Title>
    </r:Citation>
    <g:Purpose id="purpose1">
      <r:Content xml:lang="en">Not specified</r:Content>
    </g:Purpose>
    <l:VariableScheme id="varscheme1">
      <r:Label xml:lang="en">MyVariableScheme</r:Label>
      <l:Variable id="var1">
        <l:VariableName xml:lang="en">age</l:VariableName>
        <r:Label xml:lang="en">Age of the respondent</r:Label>
        <l:Representation>
          <l:NumericRepresentation type="Integer"/>
        </l:Representation>
      </l:Variable>
    </l:VariableScheme>
  </g:ResourcePackage>
</ddi:DDIInstance>
Document a Single Variable with DDI 3.2

<table>
<thead>
<tr>
<th>caseid</th>
<th>age</th>
<th>sex</th>
<th>marital_status</th>
<th>company</th>
<th>country</th>
<th>rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>65</td>
<td>M</td>
<td>Widowed</td>
<td>Drupal</td>
<td>United States</td>
<td>4.586238116</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>F</td>
<td>Widowed</td>
<td>Luctus</td>
<td>United Kingdom</td>
<td>4.231580072</td>
</tr>
<tr>
<td>1000</td>
<td>70</td>
<td>M</td>
<td>Single</td>
<td>Drupal</td>
<td>United States</td>
<td>4.0721580072</td>
</tr>
<tr>
<td>1000</td>
<td>75</td>
<td>M</td>
<td>Married</td>
<td>Luctus</td>
<td>United States</td>
<td>4.586238116</td>
</tr>
<tr>
<td>1000</td>
<td>80</td>
<td>M</td>
<td>Single</td>
<td>Luctus</td>
<td>United States</td>
<td>4.0721580072</td>
</tr>
<tr>
<td>1000</td>
<td>85</td>
<td>F</td>
<td>Married</td>
<td>Luctus</td>
<td>United States</td>
<td>4.586238116</td>
</tr>
<tr>
<td>1000</td>
<td>90</td>
<td>M</td>
<td>Single</td>
<td>Luctus</td>
<td>United States</td>
<td>4.0721580072</td>
</tr>
<tr>
<td>1000</td>
<td>95</td>
<td>F</td>
<td>Married</td>
<td>Luctus</td>
<td>United States</td>
<td>4.586238116</td>
</tr>
</tbody>
</table>
<Fragment>
  <Variable>
    <URN>urn:ddi:int.example:var1:1</URN>
    <Agency>int.example</Agency>
    <ID>var1</ID>
    <Version>1</Version>
    <VariableName>
      <String xml:lang="en">age</String>
    </VariableName>
    <Label>
      <Content xml:lang="en">Age of the respondent</Content>
    </Label>
    <VariableRepresentation>
      <NumericRepresentation>
        <NumericTypeCode>Integer</NumericTypeCode>
      </NumericRepresentation>
    </VariableRepresentation>
  </Variable>
</Fragment>
The DDI 3.1 Way

```xml
<ddi:DDIInstance id="ddii1" version="1.0.0" agency="example.org">
  <r:Citation>
    <r:Title xml:lang="en">MyDdiInstance</r:Title>
  </r:Citation>
  <g:ResourcePackage id="rp1">
    <r:Citation>
      <r:Title xml:lang="en">MyResourcePackage</r:Title>
    </r:Citation>
    <g:Purpose id="purpose1">
      <r:Content xml:lang="en">Not specified</r:Content>
    </g:Purpose>
    <l:VariableScheme id="varscheme1">
      <r:Label xml:lang="en">MyVariableScheme</r:Label>
      <l:Variable id="var1">
        <l:VariableName xml:lang="en">age</l:VariableName>
        <r:Label xml:lang="en">Age of the respondent</r:Label>
        <l:Representation>
          <l:NumericRepresentation type="Integer" />
        </l:Representation>
      </l:Variable>
    </l:VariableScheme>
  </g:ResourcePackage>
</ddi:DDIInstance>
```
The DDI 3.2 Way

```xml
<Fragment>
  <Variable>
    <URN>urn:ddi:int.example:var1:1</URN>
    <Agency>int.example</Agency>
    <ID>var1</ID>
    <Version>1</Version>
    <VariableName>
      <String xml:lang="en">age</String>
    </VariableName>
    <Label>
      <Content xml:lang="en">Age of the respondent</Content>
    </Label>
    <VariableRepresentation>
      <NumericRepresentation>
        <NumericTypeCode>Integer</NumericTypeCode>
      </NumericRepresentation>
    </VariableRepresentation>
  </Variable>
</Fragment>
```
DDI 3.2: Focus on the Content

**Study**
- Group
- StudyUnit
- Quality

**Survey**
- DataCollection
- Instrument
- ControlConstructs
- Question

**Data**
- PhysicalInstance
- DataRelationship
- Variable

**Foundational**
- Concept
- Universe
- Organization
- CodeList
- CategoryList
- Category
### Overview

#### Schemas

- Simplified packaging
- Clearer identification
- More consistency
- Bug fixes

#### Content

- Data Collection Instruments
- Harmonization
- Groups
- Data
Simplified Packaging
Simplified Packaging

- New DDI Fragment allows simple listing of DDI items
- Web services can be item-specific
- Useful for import and export
Ambiguous Location of Information

Where are Variables defined?

<DDIInstance>
  <ResourcePackage>
    <VariableScheme>
      <Variable />
    </VariableScheme>
  </ResourcePackage>
</DDIInstance>

<DDIInstance>
  <Group>
    <StudyUnit>
      <LogicalProduct>
        <VariableScheme>
          <Variable />
        </VariableScheme>
      </LogicalProduct>
    </StudyUnit>
    <StudyUnit>
      <LogicalProduct>
        <VariableScheme>
          <Variable />
        </VariableScheme>
      </LogicalProduct>
    </StudyUnit>
  </Group>
</DDIInstance>

<DDIInstance>
  <Group>
    <StudyUnit>
      <LogicalProduct>
        <VariableScheme>
          <Variable />
        </VariableScheme>
      </LogicalProduct>
    </StudyUnit>
  </Group>
</DDIInstance>
Consistent Location of Information

<FragmentInstance>
    <Fragment>
    <Variable />
    </Fragment>
</FragmentInstance>
### Explicit Top Level Element

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>10001</td>
<td>85</td>
<td>Married</td>
<td>Vitae Posuere Corp.</td>
<td>Tunisla</td>
<td>4.586298116</td>
</tr>
<tr>
<td>3</td>
<td>10002</td>
<td>29</td>
<td>1</td>
<td>Common-Law</td>
<td>Luctus Consulting</td>
<td>Isle of Man</td>
</tr>
<tr>
<td>4</td>
<td>10003</td>
<td>70</td>
<td>1</td>
<td>Common-Law</td>
<td>Purus Associates</td>
<td>Uruguay</td>
</tr>
<tr>
<td>5</td>
<td>10004</td>
<td>76</td>
<td>1</td>
<td>Common-Law</td>
<td>Pellentesque Tellus Sem Institute</td>
<td>Tunisla</td>
</tr>
<tr>
<td>6</td>
<td>10005</td>
<td>32</td>
<td>2</td>
<td>Single</td>
<td>Tempus Non Lacinia Incorporated</td>
<td>Romania</td>
</tr>
<tr>
<td>7</td>
<td>10006</td>
<td>33</td>
<td>2</td>
<td>Divorced</td>
<td>Ac Nulla Inc.</td>
<td>Tonga</td>
</tr>
<tr>
<td>8</td>
<td>10007</td>
<td>50</td>
<td>2</td>
<td>Common-Law</td>
<td>Cum Sociis Natoque Limited</td>
<td>Montserrat</td>
</tr>
<tr>
<td>9</td>
<td>10008</td>
<td>49</td>
<td>2</td>
<td>Common-Law</td>
<td>UT LLP</td>
<td>Thailand</td>
</tr>
<tr>
<td>10</td>
<td>10009</td>
<td>42</td>
<td>1</td>
<td>Common-Law</td>
<td>Cum Sociis Natoque Limited</td>
<td>Chad</td>
</tr>
<tr>
<td>11</td>
<td>10010</td>
<td>26</td>
<td>1</td>
<td>Single</td>
<td>Eget Company</td>
<td>Croatia</td>
</tr>
<tr>
<td>12</td>
<td>10011</td>
<td>42</td>
<td>Married</td>
<td>Arca LLP</td>
<td>Saint Vincent and The Grenadines</td>
<td>5.154959373</td>
</tr>
<tr>
<td>13</td>
<td>10012</td>
<td>89</td>
<td>1</td>
<td>Common-Law</td>
<td>Auctor Vitae Incorporated</td>
<td>Saint Vincent and The Grenadines</td>
</tr>
<tr>
<td>14</td>
<td>10013</td>
<td>61</td>
<td>2</td>
<td>Single</td>
<td>Lorem Auctor Quis Inc.</td>
<td>Philippines</td>
</tr>
<tr>
<td>15</td>
<td>10014</td>
<td>81</td>
<td>2</td>
<td>Common-Law</td>
<td>Luctus Ipsum Leo LLP</td>
<td>Kazakhstan</td>
</tr>
<tr>
<td>16</td>
<td>10015</td>
<td>82</td>
<td>2</td>
<td>Common-Law</td>
<td>Neque Pellentesque LLP</td>
<td>Sri Lanka</td>
</tr>
<tr>
<td>17</td>
<td>10016</td>
<td>29</td>
<td>1</td>
<td>Common-Law</td>
<td>Rohan Corporation</td>
<td>Seychelles</td>
</tr>
<tr>
<td>18</td>
<td>10017</td>
<td>28</td>
<td>2</td>
<td>Divorced</td>
<td>Luctus Felis Ltd</td>
<td>South Africa</td>
</tr>
<tr>
<td>19</td>
<td>10018</td>
<td>24</td>
<td>1</td>
<td>Single</td>
<td>Dolour Foundation</td>
<td>Tokelau</td>
</tr>
<tr>
<td>20</td>
<td>10019</td>
<td>58</td>
<td>2</td>
<td>Single</td>
<td>Odio Etiam Incorporated</td>
<td>Guam</td>
</tr>
<tr>
<td>21</td>
<td>10020</td>
<td>33</td>
<td>2</td>
<td>Single</td>
<td>Mauris Ut LLP</td>
<td>Bosnia and Herzegovina</td>
</tr>
<tr>
<td>22</td>
<td>10021</td>
<td>31</td>
<td>2</td>
<td>Common-Law</td>
<td>Tempus Inc.</td>
<td>Japan</td>
</tr>
<tr>
<td>23</td>
<td>10022</td>
<td>66</td>
<td>1</td>
<td>Common-Law</td>
<td>Quam Elementum At LLP</td>
<td>Laos</td>
</tr>
<tr>
<td>24</td>
<td>10023</td>
<td>60</td>
<td>2</td>
<td>Divorced</td>
<td>Sem Consequat Foundation</td>
<td>Sierra Leone</td>
</tr>
<tr>
<td>25</td>
<td>10024</td>
<td>85</td>
<td>2</td>
<td>Divorced</td>
<td>Ettelfeld Nunc Limited</td>
<td>Iran</td>
</tr>
<tr>
<td>26</td>
<td>10025</td>
<td>43</td>
<td>1</td>
<td>Single</td>
<td>Aenean Institute</td>
<td>Luxembourg</td>
</tr>
</tbody>
</table>
Explicit Top Level Element

```xml
<FragmentInstance>
  <TopLevelReference>
    <!-- Points to the PhysicalInstance--> 
  </TopLevelReference>
  <Fragment>
    <PhysicalInstance />
  </Fragment>
  <Fragment>
    <Variable />
  </Fragment>
  <Fragment>
    <Variable />
  </Fragment>
  <Fragment>
    <Category />
  </Fragment>
  <Fragment>
    <Category />
  </Fragment>
  ...
</FragmentInstance>
```
Clearer Identification
Clearer identification

- DDI items are identified using three pieces that create a key
  - Agency Name
  - Item Identifier
  - Item Version

- Item Identifier should be unique
  - Recommend using UUID

- No more containership rules
Consistency

- DDI items can be used by reference everywhere they are used
- Most previously required elements are now optional
- DDI schemes all look similar

Cleanup
- Spelling mistakes
- Repeated use of the same element names for different uses
Bug Fixes
Bug Fixes

- Hundreds of bugs reported in 3.1 are addressed
- Can describe hierarchical datasets!
Surveys and Data Collection

New Question Types
Surveys with Data Flows
DDI Parameters and Bindings
For each of your children, please provide the following information.

<table>
<thead>
<tr>
<th>Name</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M / F</td>
</tr>
<tr>
<td>2</td>
<td>M / F</td>
</tr>
<tr>
<td>3</td>
<td>M / F</td>
</tr>
<tr>
<td>4</td>
<td>M / F</td>
</tr>
<tr>
<td>5</td>
<td>M / F</td>
</tr>
</tbody>
</table>
Please rank your favorite ways to travel.

<table>
<thead>
<tr>
<th>CodeDomain</th>
<th>QuestionText</th>
<th>RankingDomain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plane</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automobile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bike</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foot</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. What city is this?
2. How many buildings can you count?
3. How many boats can you count?
Other: Specify

1. What is your name? ____________________________.

2. How did you get to NADDI?
   - Plane
   - Train
   - Automobile
   - Other (please specify): ________________________

3. Thanks for taking the survey, Jeremy.
Scales

How great is this year’s NADDI conference?

Anchors

Great

Low 1 2 3 4 5 Greatest

High

Value Increments
Surveys with Data Flows

1. What is your name? Jeremy

2. How did you get to NADDI?
   - Plane
   - Train
   - Automobile
   - Other (please specify): ______________________

3. Thanks for taking the survey, Jeremy.
DDI Parameters and Bindings

What is your name?

Output: FirstName

How did you get to NADDI?

Binding
Source: FirstName
Target: Name

Input: Name

Thank you, {Name}.
Groups
DDI 3.1: VariableGroups

- Group variables together
  - Topic
  - Section
  - ...

- Nested
  - Create a hierarchy of items using subgroups
DDI 3.2: Lots of Groups

- Groups can be created of most versionable item types
- Groups can have assigned concepts, subjects
InstrumentGroups and Batteries

InstrumentGroup: “Wechsler Adult Intelligence Scale”
TypeOfGroup: Battery

[Diagram showing the sequence of WAIS subtests: Information, Comprehension, Arithmetic, Similarities, ...]
Harmonization with Groups

QuestionGroup: “Age Questions”
TypeOfGroup: similar
Concept: age

- How old are you?
- What is your age?
- What age are you?
- What’s your current age?
Harmonization
Extended Data
What about when information does not fit the DDI content model?

- Previous Options
  - Abuse tags
  - Extend the schemas
  - Don’t use DDI
<s:StudyUnit>
  <r:Citation>
    <r:Title xml:lang="en">My Study</r:Title>
  </r:Citation>
  ...
  <r:UserAttributePair>
    <r:AttributeKey>Contact</r:AttributeKey>
    <r:AttributeValue>Alice</r:AttributeValue>
  </r:UserAttributePair>
  <r:UserAttributePair>
    <r:AttributeKey>Data Custodian</r:AttributeKey>
    <r:AttributeValue>Bob</r:AttributeValue>
  </r:UserAttributePair>
</s:StudyUnit>
3.2: A Little Understated
Schemas

• Simplified packaging
• Clearer identification
• More consistency
• Bug fixes

Content

• Data Collection Instruments
• Harmonization
• Groups
• Data
Focus on the Content, not Packaging and Arbitrary Rules

Study
- Group
- StudyUnit
- Quality

Survey
- DataCollection
- Instrument
- ControlConstructs
- Question

Data
- PhysicalInstance
- DataRelationship
- Variable

Foundational
- Concept
- Universe
- Organization
- CodeList
- CategoryList
- Category
DDI 3.2 is Simpler

“It all seems so simple.”

- Attendee, DDI 3.2 Training Workshop at the University of Michigan
DDI 3.2 is More Powerful

- Active DDI 3.2 Projects
  - Longitudinal Data Harmonization and Extraction
  - Cognition Measurement Item Batteries
  - Question Banks
  - Survey Instrument Designers and Code Generators
    (at least 4 of these projects)
  - Data Curation and Archive Workflow Management
  - Classification Management
  - Eurostat Quality Reporting

- What else?
Thank You

Jeremy Iverson  j Jeremy Iverson

Jeremy Iverson
Jeremy Iverson
jeremy@colectica.com

Web
colectica.com

Blog
blogs.colectica.com

Twitter
@Colectica

YouTube
youtube.com/colectica

NADDI 2014 - Vancouver, Canada
Moving Forward
DDI Moving Forward Project

- The DDI project began around 20 years ago with the aim of creating a structured metadata standard for the social sciences.

We are seeking to again improve the value we provide to our members by creating a 'model-based DDI'. This next generation DDI will:
  ▫ be easier to use,
  ▫ be able to be quickly adapted to changing future needs
  ▫ provide an even smarter and more economical way for organisations to manage their metadata, and
  ▫ make it easier for you to interact with data from other disciplines and standards.
An Agile Project

• In order to accelerate the development of DDI4, it has been decided to use an Agile approach

• Agile project management is a 'start and adjust' model, rather than a fixed plan

• This allows for the adjustment of plans to ensure that user needs are being met
Project Deliverables

- Project deliverables will include a UML model for DDI 4 but also XML and RDF bindings.

- The DDI Alliance will release two types of products to meet the needs of a diverse audience:
  
    - **Functional Views** - These views are in essence profiles of the full specification oriented around specific user needs – for example, simple data description, simple codebook, and discovery are functional views. Most users would interact only with the Functional View(s) that interest them. The DDI Alliance would publish a set of official standard views.

    - **The Library of Objects**, which encompasses the entire DDI 4.0 model. Objects are organized into packages in the Library. Packages are mutually exclusive and comprehensive. This second product is intended for use by more sophisticated users.
Sprints

• Sprints are a way to further accelerate progress.

• During a sprint, participants brought together for a set period to progress specific outputs and outcomes.

• Key features of sprint process:
  ▫ Priorities are adjusted on an on-going basis
  ▫ The group is “self-organizing”
  ▫ Frequent plenary sessions and flexible assignments of participants
  ▫ Groups are organized and re-organized based on the perceived rate of progress and priorities of the deliverables
Activities to date

• Three sprints:
  ▫ Dagstuhl (October 2013)
  ▫ EDDI (December 2013)
  ▫ NADDI (March 2014)

• Virtual task teams – Following the sprints, a number of task teams were created to follow up work.
Want to know more?

• Minutes and formal products from the Dagstuhl and EDDI sprints are found at the DDI site

• A new public wiki has been established to maintain and organize all working and final documents
  ▫ [http://www1.unece.org/stat/platform/display/DDI4/DDI+Moving+Forward+Project+Home](http://www1.unece.org/stat/platform/display/DDI4/DDI+Moving+Forward+Project+Home)

• Current development platform for content capture
Want to get involved?

- It is open to participation from any interested country or organisation, with a strong interest in DDI.

- Participation is on a voluntary, unfunded basis, therefore the specific outcomes will automatically focus on areas considered to be priorities for the participating organisations.
Questions?

Technical Committee – ddi-srg@icpsr.umich.edu