End-to-end process and DDI

Sandra Ionescu
Metadata Specialist
ICPSR
C2Metadata: End-to-end process and DDI

• Project at mid-point: progress overview
  • Tools
  • End-to-end process
  • Updated DDI
C2Metadata: End-to-end process and DDI

Tools developed in the first phase of the project:

- Standard Data Transformation Language – SDTL (Colectica and NSD)
- SPSS parser – SPSS scripts to SDTL (Colectica)
- Stata parser – Stata scripts to SDTL (NSD)
- DDI updater – DDI + SDTL -> revised DDI (MTNA)
- Pseudocode generator – SDTL to human readable text (ICPSR)
- Codebook generator – (revised) DDI to HTML codebook (ICPSR)
C2Metadata: End-to-end process and DDI

Tools developed in the first phase of the project:

- User-oriented service (ICPSR)
  - Accepts input files (data transformation script and DDI metadata)
  - Accesses the individual tools and runs the end-to-end process
  - Translates the SDTL into human-readable “pseudocode”
  - Generates HTML codebook based on the revised DDI
  - Delivers the output (updated DDI and human-readable codebook) to user
C2Metadata: End-to-end process and DDI

End user service at ICPSR (currently in open Beta testing):

Test the Software

The software generated by this project will take as input a Stata/SPSS code file and a DDI 2.5 XML file. The software will then update the DDI file to reflect the changes described into the code file and return both the revised DDI file and an html codebook generated from that file. Because processing can take some time, the output will be emailed to the address you used to log in.

DDI description of the original dataset

```plaintext
*compute using the + sign adds values only when it finds valid values on all of the source variables.
get file='da07213_useforcomputes.sav'.
compute Partydeal = VS20041+VS20042+VS20043.
Variable labels Partydeal 'Care who wins elections - Index 1'.
save ufile='da07215_ComputePlus.sav'.
execute.
```
Thank You

Your request has been submitted for processing. You should receive an email in the next few minutes that will provide you with links to download the revised DDI XML file and the HTML codebook. If it's unable to process your files, you will receive an email explaining what occurred. If you don't receive any email from us, please be sure to check your spam folder.

C2Metadata: Files Processed

You can download your revised DDI XML at:

https://c2metadata.dev.icpsr.umich.edu/c2metadata/download/jobs/336/ddi.xml

You can download the HTML codebook at:

https://c2metadata.dev.icpsr.umich.edu/c2metadata/download/jobs/336/codebook.html
C2Metadata: end user service at ICPSR

Also available at [https://c2metadata.dev.icpsr.umich.edu/c2metadata](https://c2metadata.dev.icpsr.umich.edu/c2metadata)

• DDI documentation for substantive subsets of
  • The American National Election Study (ANES) cumulative file (1948-2012)
  • The General Social Surveys (GSS) cumulative file (1972-2016)

• For internal testing:
  • Links for testing or working with the individual tools
    • [Convert SPSS code to SDTL](https://c2metadata.dev.icpsr.umich.edu/c2metadata) or [Convert Stata code to SDTL](https://c2metadata.dev.icpsr.umich.edu/c2metadata)
    • [Merge SPSS/Stata Code into DDI File](https://c2metadata.dev.icpsr.umich.edu/c2metadata)
    • [Build HTML Codebook from DDI XML](https://c2metadata.dev.icpsr.umich.edu/c2metadata)
    • [Pseudocode](https://c2metadata.dev.icpsr.umich.edu/c2metadata)
  • Links to resources for developers
  • Sample SPSS and Stata scripts for the supported commands
  • Sample DDI input metadata for testing the scripts
Currently supported commands:

- Recode variable(s)
- Rename variable(s)
- Aggregate variables
- Compute variable(s)
- Delete variable(s)
- Do Repeat Loop
- If – Then Condition
- Add variable label
- Add value labels
- Declare missing values

- Add files
- Match files
- Select cases
- Sort cases
- Sort variables
Documenting data transformations in DDI

• Original dataset:

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Width</th>
<th>Deci</th>
<th>Label</th>
<th>Values</th>
<th>Missing</th>
<th>Columns</th>
<th>Align</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>CaseID</td>
<td>Numeric</td>
<td>4</td>
<td>0</td>
<td>INTERVIEW NUMBER</td>
<td>None</td>
<td>None</td>
<td>9</td>
<td>Right</td>
<td>Scale</td>
</tr>
<tr>
<td>V520041</td>
<td>Numeric</td>
<td>1</td>
<td>0</td>
<td>CARE MUCH WHICH PTY WINS</td>
<td>{1, CARE V... 8 - Hi}</td>
<td>9</td>
<td>Right</td>
<td>Nominal</td>
<td></td>
</tr>
<tr>
<td>V520042</td>
<td>Numeric</td>
<td>1</td>
<td>0</td>
<td>XCARE WHO WINS ST ELCTN</td>
<td>{0, INAP)... 8 - Hi, 0</td>
<td>9</td>
<td>Right</td>
<td>Nominal</td>
<td></td>
</tr>
<tr>
<td>V520043</td>
<td>Numeric</td>
<td>1</td>
<td>0</td>
<td>XCARE WHO WINS LCL ELCTN</td>
<td>{0, INAP)... 8 - Hi, 0</td>
<td>9</td>
<td>Right</td>
<td>Nominal</td>
<td></td>
</tr>
</tbody>
</table>

• SPSS script:

```plaintext
compute Partycare1 = V520041+V520042+V520043.
variable labels Partycare1 'Care who wins elections - Index 1'.
save cutofile='da07213_ComputePlus.sav'.
```

• Revised dataset:

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Width</th>
<th>Deci</th>
<th>Label</th>
<th>Values</th>
<th>Missing</th>
<th>Columns</th>
<th>Align</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>CaseID</td>
<td>Numeric</td>
<td>4</td>
<td>0</td>
<td>INTERVIEW NUMBER</td>
<td>None</td>
<td>None</td>
<td>9</td>
<td>Right</td>
<td>Scale</td>
</tr>
<tr>
<td>V520041</td>
<td>Numeric</td>
<td>1</td>
<td>0</td>
<td>CARE MUCH WHICH PTY WINS</td>
<td>{1, CARE V... 8 - Hi}</td>
<td>9</td>
<td>Right</td>
<td>Nominal</td>
<td></td>
</tr>
<tr>
<td>V520042</td>
<td>Numeric</td>
<td>1</td>
<td>0</td>
<td>XCARE WHO WINS ST ELCTN</td>
<td>{0, INAP)... 8 - Hi, 0</td>
<td>9</td>
<td>Right</td>
<td>Nominal</td>
<td></td>
</tr>
<tr>
<td>V520043</td>
<td>Numeric</td>
<td>1</td>
<td>0</td>
<td>XCARE WHO WINS LCL ELCTN</td>
<td>{0, INAP)... 8 - Hi, 0</td>
<td>9</td>
<td>Right</td>
<td>Nominal</td>
<td></td>
</tr>
<tr>
<td>Partycare1</td>
<td>Numeric</td>
<td>8</td>
<td>2</td>
<td>Care who wins elections - Index 1</td>
<td>None</td>
<td>None</td>
<td>12</td>
<td>Right</td>
<td>Nominal</td>
</tr>
</tbody>
</table>
Documenting data transformations in DDI

Updated DDI:

- Variable level transformations – use derivation field:

```xml
<var ID="V5" name="Partycarel">
  <labl>
    Care who wins elections - Index 1
  </labl>
  <derivation>
    <drvdesc>
      PSEUDOCODE here.
    </drvdesc>
    <drvcmd source="producer" syntax="SPSS">
      compute Partycare1 = V520041+V520042+V520043.</drvcmd>
    <drvcmd source="producer" syntax="SPSS">
      Variable labels Partycare1 'Care who wins elections - Index 1'.<drvcmd>
    <drvcmd syntax="STDIL">
      JSON here
    </drvcmd>
  </derivation>
</var>
```
Documenting data transformations in DDI

Partycare1: Care who wins elections - Index 1

**Derivation**

```
compute Partycare1 = V520041+V520042+V520043.
```

```
Variable labels Partycare1 'Care who wins elections - Index 1'.
```

```
{
  "command" : "compute",
  "variable" : "Partycare1",
  "expression" : {
    "function" : "addition",
    "arguments" : [{
      "variableName" : "V520041"
    }, {
      "variableName" : "V520042"
    }, {
      "variableName" : "V520043"
    }],
    "relativeOf" : false
  },
  "condition" : null,
  "sourceInformation" : {
    "lineNumberStart" : 3,
    "originalSourceText" : "compute Partycare1 = V520041+V520042+V520043"
  },
  "variableName" : "Partycare1",
  "label" : "Care who wins elections - Index 1"
}
```
Documenting data transformations in DDI

Updated DDI:

- File-level transformations: use File Notes with customized content in attributes

```xml
<fileTxt>
  <notes source="archive" subject="derivation" type="pseudocode">
    pseudocode here.
  </notes>
  <notes source="producer" subject="derivation" type="SPSS">
    match files
    /file='da07213_Match_File_A.sav'
    /file='da07213_Match_File_B.sav'
    /by V520002.
    save outfile='da07213_Matched_A+B.sav'.
    execute.
  </notes>
  <notes source="archive" subject="derivation" type="SDTL">
    SDTL .json here
  </notes>
</fileTxt>
```
C2Metadata: End-to-end process and DDI

Thank you!

sandai@umich.edu