DDI Example of Use:

Unit Record Data in CSV

Contributors: Daniel Gillman (Bureau of Labor Statistics), Arofan Gregory (Aeon Technologies), Larry Hoyle (University of Kansas, Institute for Policy and Social Research), Knut Wenzig (DIW Berlin)

# Business Case

The Australian Election Survey for 2013 is encoded in a comma-delimited format. The Australian Data Archive provides access to this data set. It is described in two ways – as a minimal data description to support processing of the data, and in a more complete fashion. Both are provided in this example.

# Relevant Classes from the Model

In this section we will apply the Data Dictionary view, as it combines both logical and physical description. Metadata which is not related to the description of the data, but instead is related to the study is not covered.

|  |  |  |
| --- | --- | --- |
| **Item** | **DDI 4 Construct** | **Notes** |
| Variable name | InstanceVariable.name |  |
| Variable label | InstanceVariable.displayLable |  |
| Variable type | InstanceVariable.hasIntendedDataType |  |
| Variable value format | ValueMapping.physicalDataTytpe |  |
| Variable value range | ValueMapping.physicalDataTypeInstanceVariable.takesSubstantiveValuesFrom.DescribedValueDomain.minimumValueInclusive/minimumValueExclusive/maximumValueInclusive/maximumValueExclusive | Broken? Could be modeled as a class which allows for repeating segments within a range. Could change the SubstantiveValueDomain.DescribedValueDomain cardinality to 0..n 0..n. Same for SentinelValueDomain |
| Missing value | InstanceVariable.SentinelValueDomain |  |
| Statistics | Not in the model, other than the StatisticalSummary class in the “Keep” package | Not needed for minimum descriptor |
| Code scheme | Substantive Value Domain (CodeList, StatisticalClassification, etc.) |  |
| Sequence of variables | RectangularLayout.PhysicalLayoutOrder.PhysicalLayoutOrderedPair.ValueMapping | Pairwise ordering of Variables - transitive |
| File name | Need to add – do we have a class representing the physical file? A DataStore is a logical construct (which oh by the way has a character set property – move to a class representing the physical instance?) |  |
| File 1st line variable names | RectangularLayout.hasHeader/headerRowCount | CSVW has both, even though somewhat duplicative |
| Delimiter | RectangularLayout.delimiter/isDelimited |  |
| Quote character | RectangularLayout.quoteCharacter |  |
| Escape character | RectangularLayout.escapeCharacter |  |
| File encoding | RectangularLayout.encoding |  |
| End-of-line character | RectangularLayout.lineTerminator |  |

# Examples – Object Instances

Below is the CSV file containing a set of variables from the AES:



And as opened in Microsoft Excel



File-Level Metadata:

|  |  |
| --- | --- |
| **DDI 4 Construct** | **Values** |
| Need to add – do we have a class representing the physical file? A DataStore is a logical construct (which oh by the way has a character set property – move to a class representing the physical instance?) |  |
| RectangularLayout.hasHeader/headerRowCount | hasHeader = “true”/headerRowCount = 1 |
| RectangularLayout.delimiter/isDelimited | Delimiter = “,”/isDelimited=”true” |
| RectangularLayout.quoteCharacter | Not applicable |
| RectangularLayout.escapeCharacter | Not applicable |
| RectangularLayout.encoding | ASCII |
| RectangularLayout.lineTerminator | \n (line feed) |
| CodeList.contains.CodeItem.contains.Code  | [Not in XML – we have a serious problem here. Model is way too deep. Simplify to agree with the pattern.] |
| CodeList.contains.CodeItem.contains.Code.denotes.Category.descriptiveText.content | [Category should be a specialization of Concept] |
|  | Note: Signifier is abstract and cannot be used directly as a datatype. Fix where needed.] |

The values at the variable level:

Column Order

|  |  |
| --- | --- |
| **DDI 4 Construct** | **Value** |
| RectangularLayout.PhysicalLayoutOrder.PhysicalLayoutOrderedPair.source/target.ValueMapping.formatsInstanceVariable [Is ValueMapping needed? Direct relationship between PhysicalLayoutOrderedPair and InstanceVariable.] | Source = DivisNumTarget = UniqueID |
| RectangularLayout.PhysicalLayoutOrder.PhysicalLayoutOrderedPair.source/target.ValueMapping.formatsInstanceVariable  | Source = UniqueIDTarget = DATECOMP |
| RectangularLayout.PhysicalLayoutOrder.PhysicalLayoutOrderedPair.source/target.ValueMapping.formatsInstanceVariable  | Source = DATECOMPTarget = State |
| RectangularLayout.PhysicalLayoutOrder.PhysicalLayoutOrderedPair.source/target.ValueMapping.formatsInstanceVariable  | Source = StateTarget = Division |
| RectangularLayout.PhysicalLayoutOrder.PhysicalLayoutOrderedPair.source/target.ValueMapping.formatsInstanceVariable  | Source = DivisionTarget = A4 |
| RectangularLayout.PhysicalLayoutOrder.PhysicalLayoutOrderedPair.source/target.ValueMapping.formatsInstanceVariable  | Source = A4Target = G1age |
| RectangularLayout.PhysicalLayoutOrder.PhysicalLayoutOrderedPair.source/target.ValueMapping.formatsInstanceVariable  | Source = G1ageTarget = XG5 |
| RectangularLayout.PhysicalLayoutOrder.PhysicalLayoutOrderedPair.source/target.ValueMapping.formatsInstanceVariable  | Source = XG5Target= weight |
| RectangularLayout.PhysicalLayoutOrder.PhysicalLayoutOrderedPair.source/target.ValueMapping.formatsInstanceVariable  | Source = weightTarget = PARTYABY |
| RectangularLayout.PhysicalLayoutOrder.PhysicalLayoutOrderedPair.source/target.ValueMapping.formatsInstanceVariable  | Source = PARTYABYTarget = SWINGN |

Column 1:

|  |  |
| --- | --- |
| **DDI 4 Construct** | **Values** |
| InstanceVariable.name | DivisNum |
| InstanceVariable.displayLable | Electoral Division – AES Numeric Code |
| InstanceVariable.hasIntendedDataType | Nominal |
| ValueMapping.physicalDataType | Numeric Code |
| ValueMapping.physicalDataTypeInstanceVariable.takesSubstantiveValuesFrom.DescribedValueDomain.minimumValueInclusive/minimumValueExclusive/maximumValueInclusive/maximumValueExclusive | Not applicable |
| InstanceVariable.SentinelValueDomain | Missing = “-1” |
| Not in the model, other than the StatisticalSummary class in the “Keep” package | [table of summary statistics] |
| Substantive Value Domain (CodeList, StatisticalClassification, etc.) | AES Numeric Electoral Division Codes (examples):101 Canberra 102 Fraser 103 Banks 104 Barton105 Bennelong106 Berowra |

# Relationship to Other Standards/Vocabularies

[Describe any useful relationships with other models/standards/namespaces as appropriate, and at a fine level of detail if appropriate.]

# XML Example

[Provide an example of the DDI 4 XML binding for the appropriate part of the model]

# Adherence to Design Principles

[Look at the [DDI 4 Design Principles](https://ddi-alliance.atlassian.net/wiki/download/attachments/37552132/Design_Principles.pdf?version=1&modificationDate=1466520975681&api=v2) and indicate rationale/discuss each in light of this example.]

#