

# Integrating DDI Codebook with PROV

## The purpose of the document

The purpose of this document is to concisely describe the changes made to integrate the DDI 2.5-NCRN xml schema with the PROV-XML schema.

## Namespace Declaration

The Prov-XML schema is aliased in the namespace declarations, and added to the import section. The schema location is set to the local directory to be consistent with the current codebook schema, but could be referenced by URL as well. Having a local schema location is suitable for installation within a restricted environment.

```
<xs:schema xmlns:xhtml="http://www.w3.org/1999/xhtml"
  xmlns:fn="http://www.w3.org/2005/xpath-functions" xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:saxon="http://xml.apache.org/xslt" xmlns="ddi:codebook:2_5"
  xmlns:dc="http://purl.org/dc/terms/" xmlns:prov="http://www.w3.org/ns/prov#" targetNamespace="ddi:codebook:2_5"
  elementFormDefault="qualified" attributeFormDefault="unqualified">
  <xs:annotation>
    <xs:documentation>
      This is a w3c Schema "Technical Implementation" of the DDI Conceptual Specification.
      This schema is intended for use in producing electronic versions of codebooks for quantitative social scienc
      Please note that the attribute xml-lang in the a.globals group is an error that was persisted to retain back
    </xs:documentation>
  </xs:annotation>

  <xs:import namespace="http://www.w3.org/XML/1998/namespace" schemaLocation="xml.xsd"/>
  <xs:import namespace="http://www.w3.org/1999/xhtml" schemaLocation="ddi-xhtml11.xsd"/>
  <xs:import namespace="http://purl.org/dc/terms/" schemaLocation="dcterms.xsd"/>
  <xs:import namespace="http://www.w3.org/ns/prov#" schemaLocation="prov.xsd"/>
```

## Study Level Provenance

This section describes the changes made integrating prov:Document into the DDI Codebook-NCRN schema.

## Embed within RelStdy

### Step 1 - Material Reference Complex Type with Prov

To include prov:Document within the relStdy element, a new complex type called 'materialReferenceWithProvType', which inherits from materialReferenceType can be introduced as follows:

```

<xs:complexType name="materialReferenceWithProvType" mixed="true">
  <xs:complexContent>
    <xs:extension base="materialReferenceType">
      <xs:sequence>
        <xs:element ref="prov:document" minOccurs="0" maxOccurs="1" />
      </xs:sequence>
      <xs:attribute name="provDocURI" type="xs:anyURI" use="optional"/>
    </xs:extension>
  </xs:complexContent>
</xs:complexType>

```

This allows prov document to be embedded or referenced by URI.

Step 2 - Modify the type of relStdy to the new complex type

The relStdy element is changed to inherit from this 'materialReferenceWithProvType', and examples are given

```

<xs:element name="relStdy" type="materialReferenceWithProvType">
  <xs:annotation>
    <xs:documentation>
      <html:div>
        <html:h1 class="element_title">Related Studies</html:h1>
        <html:div>
          <html:h2 class="section_header">Description</html:h2>
          <html:div class="description">Information on the relationship of the current data collection to others [4 lines]
        </html:div>
        <html:div> [8 lines]
        <html:div> [3 lines]
        <html:div>
          <html:h2 class="section_header">Prov Example - Embedded Provenance Document</html:h2>
          <html:div class="example">
            <html:samp class="xml_sample"><![CDATA[
              <relStdy>ICPSR distributes a companion study to this collection titled FEMALE LABOR
              FORCE PARTICIPATION AND MARITAL INSTABILITY, 1980: [UNITED STATES] (ICPSR 9199).
              <prov:document>...</prov:document></relStdy>
            ]]></html:samp>
          </html:div>
        </html:div>
        <html:div>
          <html:h2 class="section_header">Prov Example - Referenced Provenance Document</html:h2>
          <html:div class="example">
            <html:samp class="xml_sample"><![CDATA[
              <relStdy provDocURI="http://dx.doi.org/10.1234/exProv123">ICPSR distributes a companion
              study to this collection titled FEMALE LABOR FORCE PARTICIPATION AND MARITAL INSTABILITY,
              1980: [UNITED STATES] (ICPSR 9199).</relStdy>
            ]]></html:samp>
          </html:div>
        </html:div>
      </xs:documentation>
    </xs:annotation>
  </xs:element>

```

## Variable Level Provenance

4. A single attribute is added to the variable type

```

</xs:attribute>
<xs:attribute name="geoVocab" type="xs:string"/>
<xs:attribute name="catQnty" type="xs:string"/>
<xs:attribute name="representationType">
  <xs:simpleType>
    <xs:restriction base="xs:NMTOKEN">
      <xs:enumeration value="text"/>
      <xs:enumeration value="numeric"/>
      <xs:enumeration value="code"/>
      <xs:enumeration value="datetime"/>
      <xs:enumeration value="other"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute ref="prov:ref" use="optional"/>
<xs:attribute name="otherRepresentationType" type="xs:NMTOKEN" use="optional"/>
</xs:extension>
</xs:complexContent>
</xs:complexType>

```

Per the PROV-XML identifier conventions, “PROV-XML uses the prov:id xml-attribute to denote PROV Identifier and the prov:ref xml-attribute to denote reference-by-id to an instance of a prov-”type” or prov-”relation” which has a matching PROV Identifier. The value of prov:ref must be a valid PROV Identifier, **but there is no requirement that for every prov:ref a corresponding prov:id must be known to exist.**”<sup>1</sup>

In our implementation, a variable would reference a prov:Bundle that would be found within embedded prov:Document. Here is an example of a prov:Bundle:

---

<sup>1</sup> <http://www.w3.org/TR/2013/NOTE-prov-xml-20130430/#identifier-conventions>

```

<prov:bundle prov:id="ex:bundle1">
  <ex:version>1</ex:version>
</prov:bundle>

<prov:bundleContent prov:id="ex:bundle1">
  <prov:entity prov:id="ex:report1" />

  <prov:entity prov:id="ex:report2">
    <prov:type xsi:type="xsd:QName">report</prov:type>
    <ex:version>2</ex:version>
  </prov:entity>

  <prov:wasGeneratedBy>
    <prov:entity prov:ref="ex:report2" />
    <prov:time>2012-05-25T11:00:01</prov:time>
  </prov:wasGeneratedBy>

  <prov:wasDerivedFrom>
    <prov:generatedEntity prov:ref="ex:report2" />
    <prov:usedEntity prov:ref="ex:report1" />
  </prov:wasDerivedFrom>
</prov:bundleContent>

```

To reference this for a given variable, you would simply add the optional element, prov:ref="ex:bundle1".

For complete documentation, please see

<http://ncrn.cornell.edu/docs/ddi/2.5.NCRN.P/docs/codebook.html>