DDI Discovery: An Overview of Current RDF Vocabularies

Arofan Gregory
Metadata Technologies NA
Joachim Wackerow
GESIS
Background

• For many years, interest has been increasing in the set of standards-based technologies from W3C around “Linked Data”
  – (aka “The Semantic Web”, RDF, etc.)
  – RDF stands for the “Resource Description Framework” – a way of making the Web semantically richer and machine-actionable
  – In the recent past, these technologies have shown increasing promise

• DDI has always been an “XML standard”
  – But this has changed!
  – DDI now does both RDF and XML as underlying syntaxes for implementation
  – RDF is based on the model *implicit* in DDI Codebook and DDI Lifecycle XML
A Tale of Two Ontologies

• This presentation is not a modeler’s view of the DDI RDF ontologies
  – Thomas Bosch of GESIS (and others) have done many good presentations in this style
• This is more of a general overview, to explain what has been developed and published, and how it might broaden the use of DDI
• DDI RDF Discovery Vocabulary (studies & research data)
• XKOS Vocabulary (statistical classifications)
Well, Maybe Three Ontologies...

• In the recent past, as a result of pressure from the Open Data movement in the UK, an ontology had been developed for publishing aggregate statistics
  – Data Cube
  – Now part of the W3C Open Data suite of standards

• This was based on the Statistical Data and Metadata Exchange (SDMX) standard
  – Also designed to be used as a description of tabulations in a DDI-based vocabulary for describing microdata
DDI RDF Discovery Vocabulary ("Disco")

• The initial application of DDI to RDF was to handle the majority use-case we see with RDF
  – Publication on the Web!
  – Describes conceptual metadata about studies and data sets for finding them across the Web of Linked Data
  – Both at the study level and at the variable level
• For research data, this is a function most often seen after data has been produced and disseminated
  – Data archives
  – Data libraries
  – Data producers
  – Statistical and government agencies
XKOS

• The second most-used RDF vocabulary is the Simple Knowledge Organization System (SKOS) vocabulary
• XKOS is a set of extensions to SKOS to support formal statistical classifications
• Useful as a stand-alone vocabulary, and in combination with Disco and Data Cube
DDI Discovery
The Full Picture
DDI Discovery Publication

• Available on the DDI Alliance website (www.ddialliance.org)
• As an HTML specification, but also available in machine-processable formats (TURTLE, N3, RDF XML)
• “Unofficial Draft” status
  – Still responding to implementer’s issues and comments
  – Relatively stable
XKOS Publication

• Available at the DDI Alliance website
• Published as an HTML specification
• Also published in machine-actionable formats (TURTLE, N3, RDF XML)
• Also with an “Unofficial Draft” status
  – More stable than the Disco vocabulary
Looking Forward

• There is one more RDF vocabulary being worked on in this series...

• Achim Wackerow will present this in the next session...

• The “next generation” DDI 4 will provide broader support for RDF
  – Every part of the DDI Model will be expressed both as a set of XML schemas and as a set of RDF vocabularies
DDI 4 as a “Model-Based” Standard

DDI 4 Model

DDI XML Schemas

Semantically Equivalent

DDI RDF Vocabularioes
Questions?