

# Collaborative Research: Comprehensive Citation Across the Data Life Cycle Using DDI

Larry Hoyle (PI)<sup>1</sup>, Mary Vardigan (Co-PI)<sup>2</sup>, Jay Greenfield<sup>3</sup>, Sam Hume<sup>4</sup>, Sanda Ionescu<sup>2</sup>, Jeremy Iverson<sup>5</sup>, John Kunze<sup>6</sup>, Barry Radler<sup>7</sup>, Wendy Thomas<sup>8</sup>, Stuart Weibel<sup>9</sup>, Michael Witt<sup>10</sup>

NSF Award # 1448107, meeting at Schloss Dagstuhl\*, event 14432

Project archive:

<http://kuscholarworks.ku.edu/handle/1808/15746>



*Q: What objects documented by DDI should be citable?*

All versionable objects, some may not be used

*Q: What elements are needed in DDI and CDISC to support citation and describe data sources in a comprehensive way across the lifecycle?*

See Proposed DDI4 modeling at right

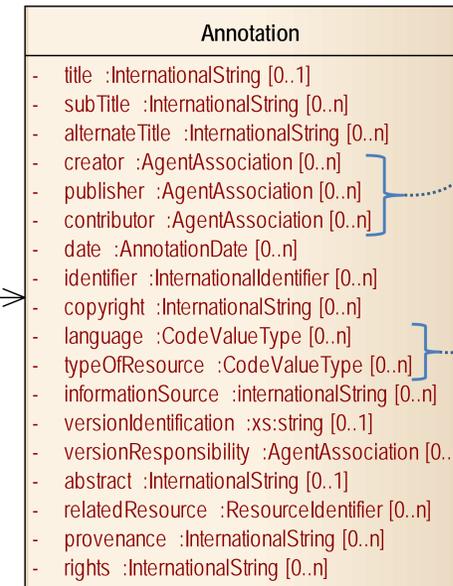
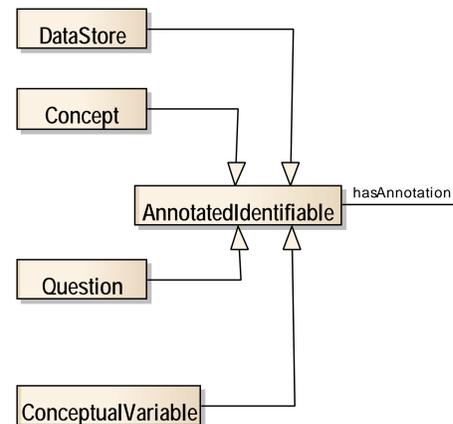
*Q: Given the lifecycle focus of DDI, how can we support broad attribution for contributions, and how can we describe the level of contribution?*

Follow and extend the CRediT Taxonomy

*Q: Other citation-like metadata?*

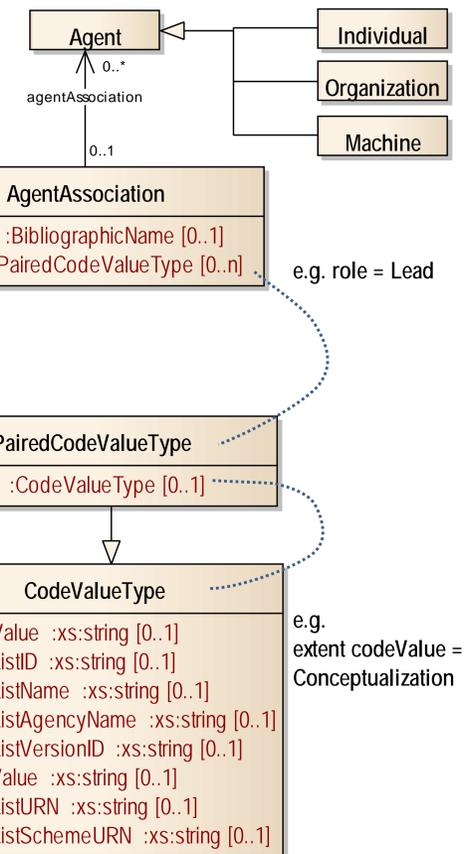
Administrative and characterization metadata needing an unpredictable vocabulary – Examples: question sourcing in support of OMB reviews; the identification of manufacturers of medical devices used in the collection and management of biospecimens

Most objects inherit from AnnotatedIdentifiable. Examples:



The Annotation object will also have an additional property capable of containing administrative, characterizing, and other information structured by an external vocabulary

DDI4 model extract from: <http://lion.ddialliance.org/>



## DDI Moving Forward

[https://dditools.atlassian.net/wiki/display/DDI4/\\*Moving+Forward+Project](https://dditools.atlassian.net/wiki/display/DDI4/*Moving+Forward+Project)

The DDI initiative, established in 1995 with NSF support, aims to document research datasets and processes thoroughly so that data are independently understandable. DDI covers the research data life cycle from conceptualization to collection and processing, to data publication and beyond. Advantages of the DDI approach are that metadata are machine-actionable and reusable. Originally expressed in XML schemas, DDI is now evolving as a model-based specification (DDI4). This effort is proceeding through a series of in person sprints and online meetings.

## 2014 Dagstuhl Sprint

<http://www.dagstuhl.de/14432>

The 2014 Dagstuhl Sprint included DDI experts along with representatives from the Dublin Core Metadata Initiative and CDISC, the Clinical Data Interchange Standards Consortium funded by NSF, to focus on adding structure to DDI4 to support enhanced citation of data. The sprint produced the additions to the DDI4 model shown above along with proposed additions to CDISC (not shown).

## The CRediT Taxonomy

<http://credit.casrai.org/proposed-taxonomy/>

The working group at Dagstuhl proposed adopting the CRediT taxonomy as the top-level vocabulary of a potentially extensible categorization scheme for contributorship. The taxonomy includes 14 categories of contribution: conceptualization, methodology, software, validation, analysis, investigation, resources, curation, writing, review and editing, visualization, supervision, administration, and funding acquisition. Each category can be ascribed with one of three degrees of contribution: lead, equal, and supporting. Example: Ann A Researcher – Conceptualization, lead; Formal Analysis, equal; Writing – original draft, equal; Funding acquisition, lead.



<sup>1</sup>IPSR, University of Kansas; <sup>2</sup>ICPSR, University of Michigan; <sup>3</sup>Booz Allen Hamilton; <sup>4</sup>Clinical Data Interchange Standards Consortium (CDISC); <sup>5</sup>Colectica; <sup>6</sup>California Digital Library, University of California; <sup>7</sup>Institute on Aging, University of Wisconsin-Madison; <sup>8</sup>Minnesota Population Center, University of Minnesota; <sup>9</sup>OCLC Research; <sup>10</sup>Distributed Data Curation Center, Purdue University

\*Schloss Dagstuhl (Leibniz Center for Informatics) organizes and supports weekly workshops on diverse topics related to computer science where scientists from all over the world come to do collaborative research. The center is installed in a very remote and relaxed location in the countryside of Saarland in Germany.