

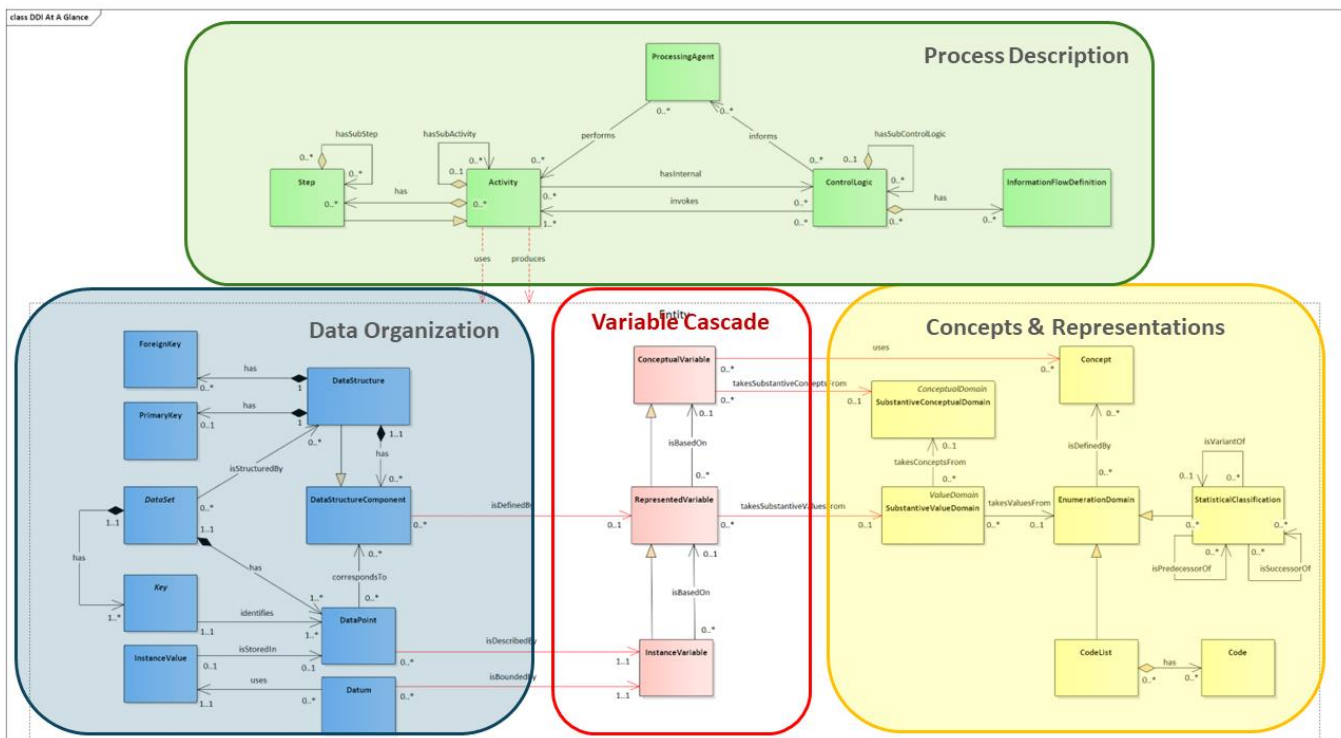


CDI

Data Documentation Initiative: Cross Domain Integration

What is DDI-CDI?

DDI-CDI is a new metadata specification from the DDI Alliance, which focuses on areas often not addressed, but critical for data integration: granular variable-level descriptions, and the full processing context which has produced the data. It is designed to describe the integration of data coming from different sources, infrastructures, or disciplines/domains. It describes not only wide data sets, but also long, multi-dimensional, and key-value types. It is designed to work with other existing domain and web standards, extending them to support the combination and reuse of data more easily. It works with other DDI specifications (DDI Codebook and DDI Lifecycle) as well as popular vocabularies such as DCAT¹, SKOS², PROV³, SDMX⁴, and many others. It can be used to implement such common models as the GSBPM⁵ and GSIM⁶.



The DDI-CDI Model "At a Glance"

Why Now?

Today, there is a growing demand for more data coming from many different sources in many different structures and formats. In the age of FAIR data-sharing, there is a need to be able to easily combine these diverse data, and to document how it is done. This requires a fine-grained description of data in a single form, allowing the data to be processed within a single system. Further, the processing of data is critical to allow for understanding, transparency, and accurate use. DDI-CDI is designed to meet this multi-faceted demand.

What Technology is It Implemented in?

Unlike other DDI specifications which used XML, DDI-CDI is a “model-driven” standard: it is a UML model, optimized for interoperability,⁷ which has standard syntax representations in XML and different RDF serializations (Turtle and JSON-LD). Because it is not specific to any one syntax, the model can work to promote interoperability using any number of different representations, including Python (and other object-oriented languages), R, SHACL, etc.

Who is Using It?

The specification is only now becoming available, but has already been implemented for different purposes by many organizations which use other flavors of DDI. These include the European Social Survey for describing cross-domain integration processing, the U.S. Bureau of Labor Statistics for managing indicators and time series, and UK Data Archive for their emerging data product builder, supporting real-time negotiation of access to confidential data. DDI-CDI is being adopted as part of the WorldFAIR Cross Domain Interoperability Framework for describing data for use in FAIR implementation globally. It is expected to see widespread adoption within open science clouds and similar large-scale infrastructures.

How Can I Learn More?

The DDI Alliance website provides a good place to start: <https://ddialliance.org/Specification/DDI-CDI/>

¹ W3C Data Catalog Vocabulary

² W3C Simple Knowledge Organization System

³ W3C PROV (Provenance) Ontology

⁴ ISO 17369 Statistical Data and Metadata Exchange

⁵ UN Economic Commission for Europe - Generic Statistical Business Process Model

⁶ UN Economic Commission for Europe – Generic Statistical Information Model

⁷ The model is expressed in Canonical XMI according to the UML Class Model Interoperable Subset (UCMIS).