

# DDI-Cross-Domain Integration (DDI-CDI): Current Status and Further Developments

Arofan Gregory, CODATA

Chair, DDI-CDI WG

EDDI 2023



License: [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](#)

# Outline

- Overview of DDI-CDI
- Status
- Dagstuhl and Upcoming Events
- Topics for Further Work
- Prototypes
- Prospects
- Summary

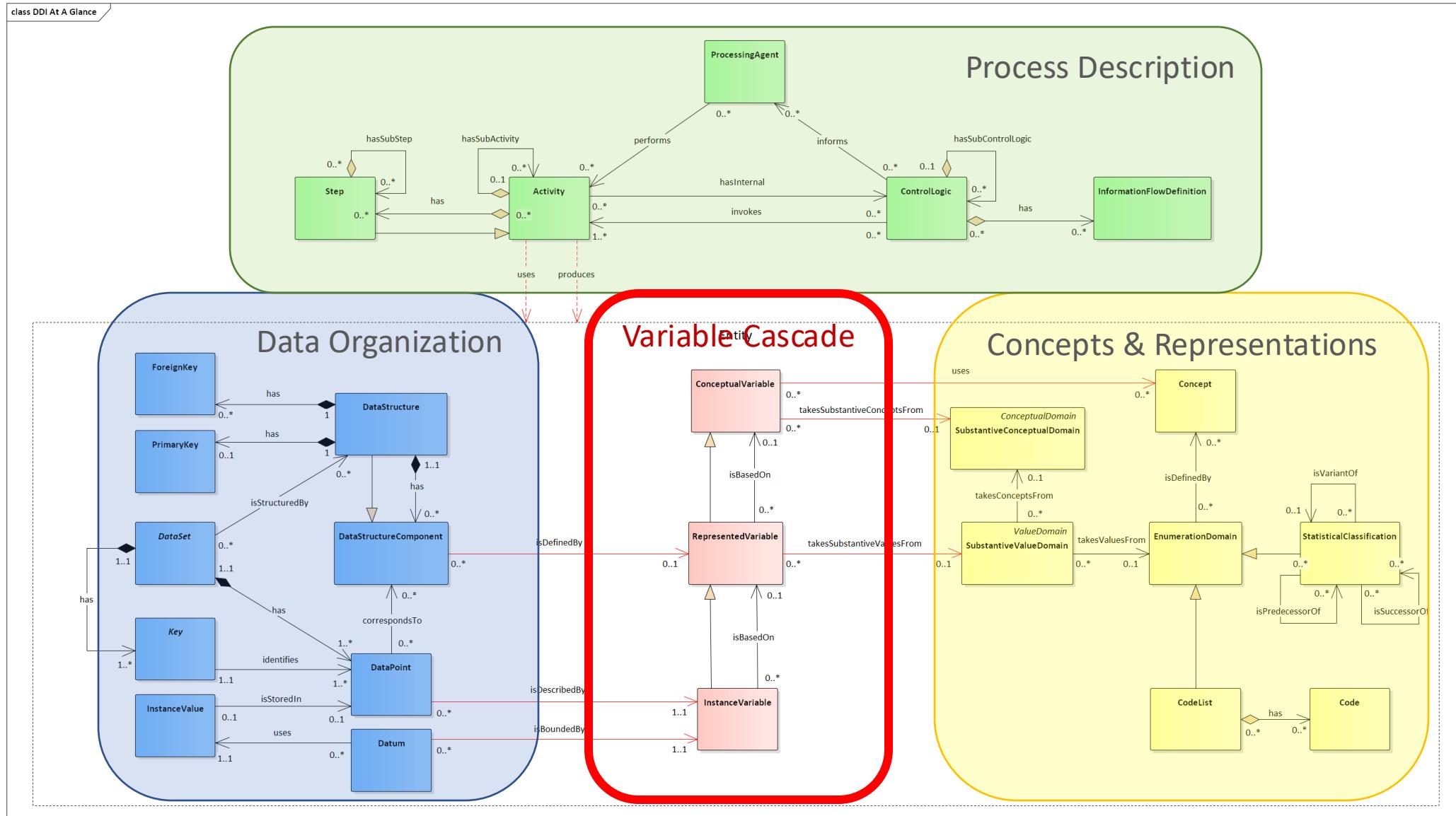
# Overview of DDI-CDI

- Model-driven standard
  - UML model in Canonical XMI is the core specification (UCMIS subset of UML features)
  - XML and RDF syntax representations (OWL, JSON-LD) produced programmatically
  - Other representations in development (Python, ShEx, SHACL, Java, etc.)
- Domain-neutral, but designed to work with DDI-C and DDI-L
- Designed to work with other popular web standards (Schema.org, DCAT, SKOS/XKOS, SDTL, PROV-O, etc.) and domain standards
- Supports machine-actionability for dissemination/reuse
- Based on the DDI 4/"Moving Forward" model

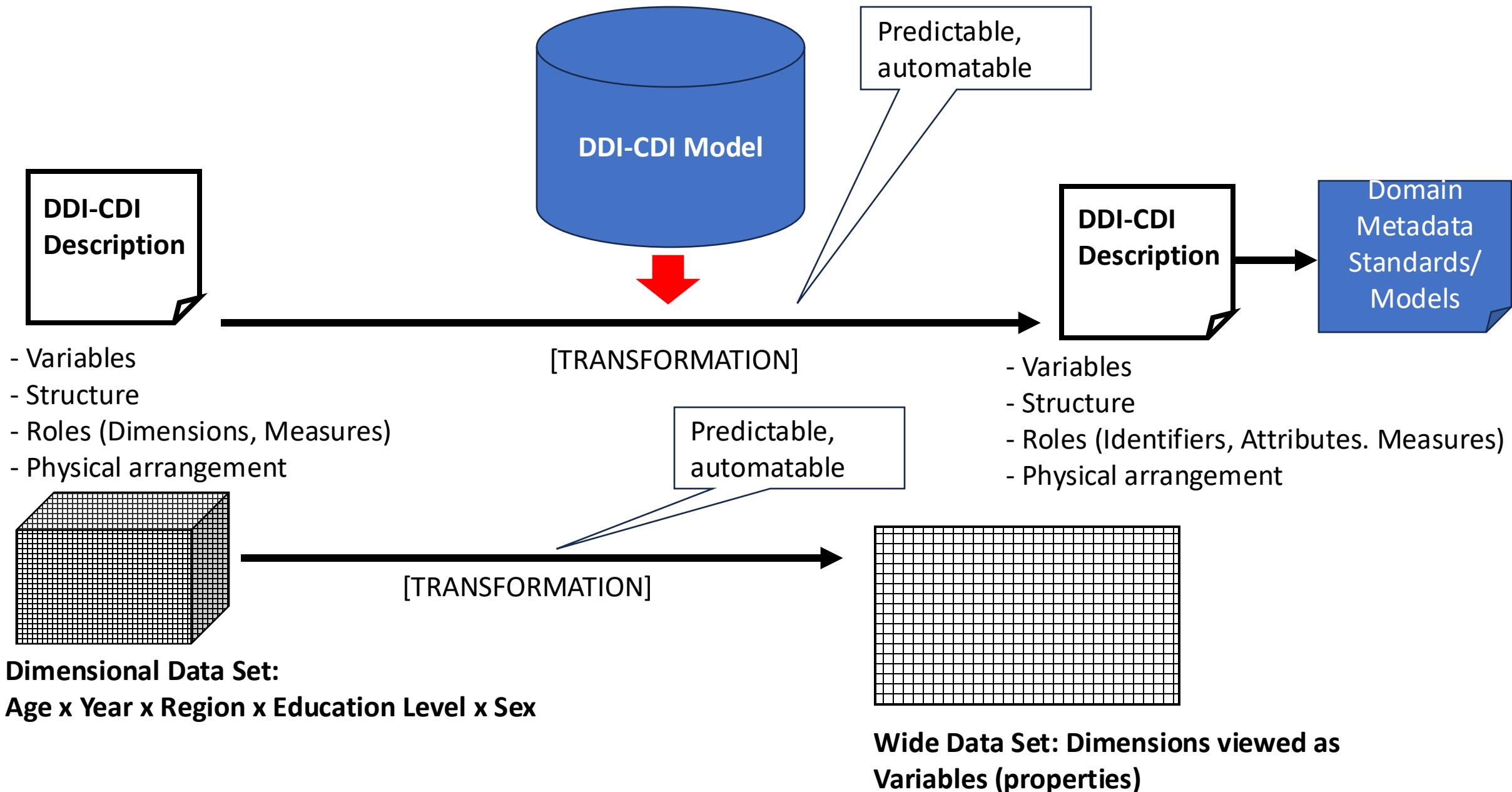
# What Does It Do?

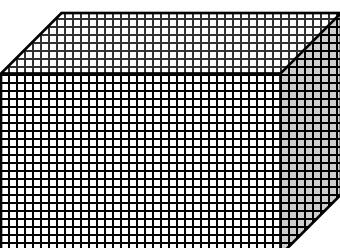
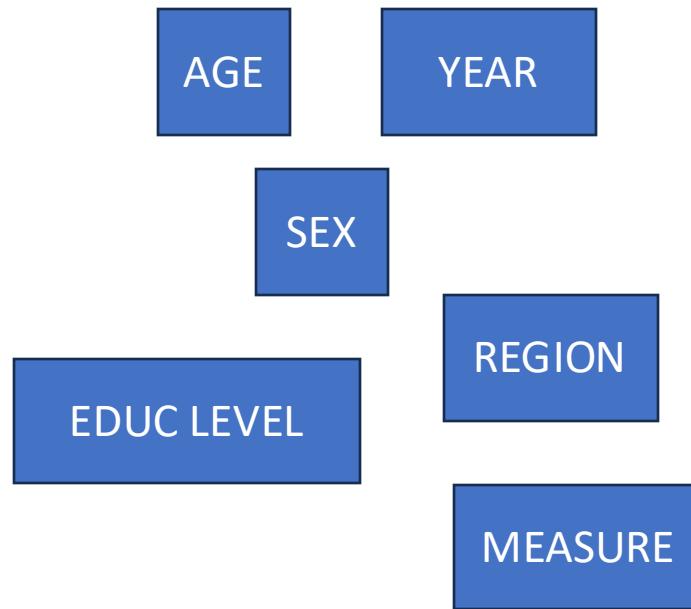
- Describes data in new formats for easier integration with traditional data
  - Wide (traditional) data
  - Long data (sensor/event)
  - Key-value (big) data
  - Multi-dimensional (cube) data
- Exposes data resources for easier integration by users from *any domain*
  - Describes the variables/datums in a data set
  - Describes how semantics (codelists/ontologies/concepts) are attached to the data and the *roles* they play
  - Supports re-assembly of variables/datums into any needed data structure
- Describes the *process* of data integration so that users can understand unfamiliar data
- DDI-CDI does not cover study-level metadata: it relies on other standards to provide this

# DDI-CDI at-a-glance



Courtesy of the amazing Flavio Rizzolo, Statistics Canada





A diagram illustrating the target format. A large red arrow points from the initial state to this table. The table has six columns with gray headers: "AGE", "YEAR", "SEX", "REGION", "EDUC LEVEL", and "MEASURE". Each column contains three white rectangular boxes, representing data records.

AGE	YEAR	SEX	REGION	EDUC LEVEL	MEASURE

The target format is wide, with the dimensions treated as variables (properties): at least one must be a place where integration can be performed with another data set (typically time and geography). Other variables may be combined into compound identifiers, or may be treated as additional descriptors or measures, depending on what they are. The roles of variables Change according to the structures – the meanings/values do not.

Aggregate values would be repeated to align with micro-data records as needed to provide the “complete” records for analysis.

Recoding/semantic transformations are handled separately as appropriate to the structural re-arrangement.

# Status

- Candidate release turned over to DDI Technical Committee in May 2023
  - Download: <https://bitbucket.org/ddi-alliance/ddi-cdi/>
- Revised candidate release now being finalized after implementation/testing
  - Finalize XML schema optimization for generation of syntax representations
  - Updated examples of XML and RDF syntaxes
- Focused review of syntax representations
  - Detailed notes on generation of syntaxes prepared
- Vote for release as specification
  - Webinar for DDI community

# Integrated Field-Level Documentation

UML Model: DDI Cross Domain Integration (DDI-CDI 1.0) » DDICDILibrary » Classes » Conceptual » CategorySet previous | next | index

**DDI CDI**

Quick search  Go

**Table of Contents**

- Context
- DDICDILibrary
  - Classes
    - Agents
    - Conceptual
      - Category
      - CategoryPosition
      - CategoryRelationship
      - CategoryRelationStructure
      - CategorySet
        - Definition
        - Examples
        - Explanatory notes
      - Concept
      - ConceptMap
      - ConceptRelationship
      - ConceptStructure
      - ConceptSystem
      - ConceptSystemCorrespondence
      - ConceptualDomain
      - ConceptualVariable
      - InstanceVariable
      - Population
      - RepresentedVariable
      - SentinelConceptualDomain
      - SubstantiveConceptualDomain
      - Unit
      - UnitType
      - Universe
      - VariableCollection
      - VariablePosition
      - VariableRelationship
      - VariableStructure
    - DataDescription
    - FormatDescription
    - Process
    - Representations
  - DataTypes
  - DesignPatterns

## CategorySet

**Fully qualified class name:** DDICDILibrary::Classes::Conceptual::CategorySet

### Definition

Concept system where the underlying concepts are categories.

### Examples

"Male" and "Female" categories in a category set named "Gender".

### Explanatory notes

The categories in a category set help define the meaning of the category set. Gender can be defined as "male or female" - see example above. A category set can be used directly by questions to express a set of response choices.

### Diagram

[Open diagram in additional window](#)

**UML Diagram: Class CategorySet in Context**

```
graph TD; CS[Conceptual::ConceptSystem] -- "allowsDuplicates: Boolean" --> CS; CS -- "catalogDetails: CatalogDetails" --> CS; CS -- "externalDefinition: Reference" --> CS; CS -- "identifier: Identifier" --> CS; CS -- "name: ObjectName" --> CS; CS -- "purpose: InternationalString" --> CS; CS -- "isDefinedBy" --> C[Conceptual::Concept]; CS -- "has" --> CS_S[Conceptual::ConceptStructure];
```

**Hints**  
- Move the mouse cursor over a name to see more information.  
- Click on a name to go to the corresponding page.  
- The arrows of the inheritance tree are colored.

Class has 5 associations  
Only 3 are shown here.

# Dagstuhl and Upcoming Events

- Workshop in September 2023 at Schloss Dagstuhl
- Groups:
  - Implementation guides/supporting implementation
  - Non-numeric, non-coded datums (“qualitative”)
  - Using DDI-CDI: documentation on what the standard is for and how it can be used by organizations
  - Dependencies between variables (O-ADOPT, O&M, etc.)
  - Syntax representations
- Informal face-to-face at EDDI 2023
- Face-to-face in Paris, April 2024 (margins of COSMOS)
- Dagstuhl 2024

# Topics for Further Work

- Finalize “Using DDI-CDI” documentation (explanatory)
- Implementation Guides: creating community profiles for use
  - Subset of model
  - Syntax representations
  - Other standards
- Tools
  - Improve community subset documentation tool
  - DDI transformations (DDI-C, DDI-L)
  - Collaborate with Developer’s Group
- Non-numeric, non-coded datums (N3CD) – qualitative sub-group?
- Further syntax representations
- “Beta” release process to support implementers
  - Coordinate with TC
- Refinements to model (Complex Units of Measure, etc.)

# Prototypes

- UKDA Data Product Builder
  - Variables
  - Wide, long, multi-dimensional data structures
- EOSC Future Science Project “Climate Neutral and Smart Cities” (Sikt)
  - DDI-CDI Process description
  - Integration with DDI Lifecycle/Colectica
- Transformation tools
  - CSV-to-CDI (JSON-LD) : <https://ddialliance.github.io/ddi-cdi-sample-generator>
  - STATA/SPSS Sample Generator (Sikt): <https://bbeuster.pythonanywhere.com/>
  - Bureau of Labor Statistics
  - Indicator model

# Prospects

- FAIR Implementation offers new users
  - WorldFAIR “Cross-Domain Interoperability Framework” (CDIF)
  - Other EOSC Implementations (FAIR Impact, EOSC Interoperability Framework)
- Cross-domain FAIR implementations within DDI Community
  - Combined with DDI-C/DDI-L

# Cross-Domain Interoperability Framework (CDIF)

- Guidelines for the coordinated use of a set of cross-domain standards as a “lingua franca” for FAIR data exchange and use
  - Specific profiles of DDI-CDI for data description
- CDIF has gathered 30 experts from different standards communities
  - Working Group: Meets bi-weekly to draft CDIF recommendations
  - Advisory Group: Reviews drafts and identifies direction/advises
- Major inputs:
  - WorldFAIR FAIR Implementation Profiles and 11 Domain Use Cases, from social science to bio-diversity to nano-technology to geo-chemistry...
  - FAIR domain and cross-domain metadata standards
  - FAIR initiatives (EOSC IF, FAIR Impact, EOSC Future, GO FAIR, DRUM, etc.)
- Initial draft of CDIF to be delivered as part of WorldFAIR outputs in Summer 2024



# FAIR Activities and Standards: CDIF

- Foundational: FAIR Digital Object Framework (**FDOF**)
- Find
  - Discover FAIR resources and explore/evaluate their utility prior to access (coverage, etc.)
  - **Schema.org, DCAT. DDI-CDI (variables)**
- Access
  - Negotiate access to non-public data
  - Enhance efficiencies through automation
  - **ODRL, DPV**
- Assess/Integrate
  - Understand data structure (**DDI-CDI**)
  - Understand semantics (**SKOS/XKOS, OWL, SSSOM**)
  - Determine origination/context (**PROV-O, I-ADOPT/O&M**)

# Summary

- DDI-CDI extends existing DDI specifications to use different types of data from other domains
- It is a vehicle for cross-domain FAIR implementation
- Version 1.0 in coming months
  - Well-documented
  - Well-tested
  - Builds on existing specifications/processes
- Helps bring DDI products in line with current technologies and modular standards approaches

Questions?