

# DDI-Codebook & DDI-Lifecycle

Jon Johnson

12 September 2019

CLOSER, UCL Institute of Education



# What is DDI?

- A structure for describing data and its related information
- An open standard designed for data sharing and reuse
- A mechanism for creating consistent, machine-actionable documentation for data
- Supports the entire research data lifecycle



# Historical approach of DDI Versions

- Evolving, non-linear standard
  - DDI Codebook (DDI-C)
  - DDI Lifecycle (DDI-L)
- Designed for different purposes
  - DDI Codebook is intended for documenting a **single study**
  - DDI Lifecycle is intended for documenting **relationships** between **data objects**
- Preparation phase of both standards is different



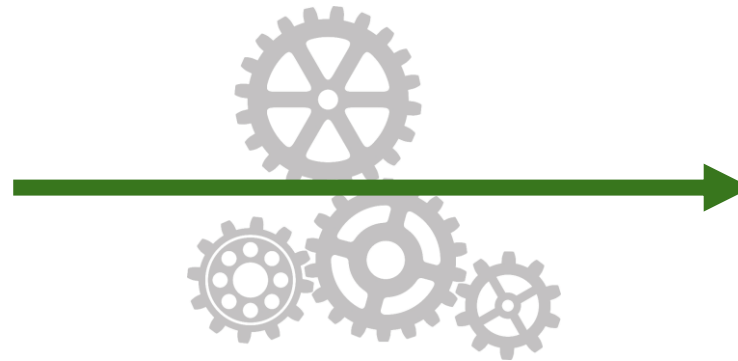
# DDI-Codebook

**Metadata not structured**



Lost Metadata Manager

*DDI Codebook structures*



**Metadata structured by DDI  
Codebook standard**

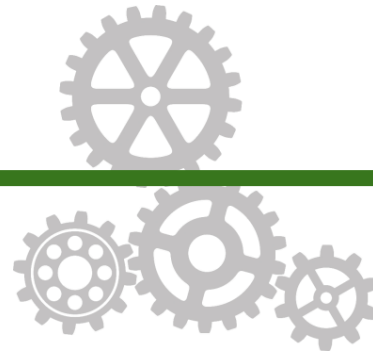




# DDI-Lifecycle



*DDI Lifecycle Structures*



Variable Scheme



Universe Scheme



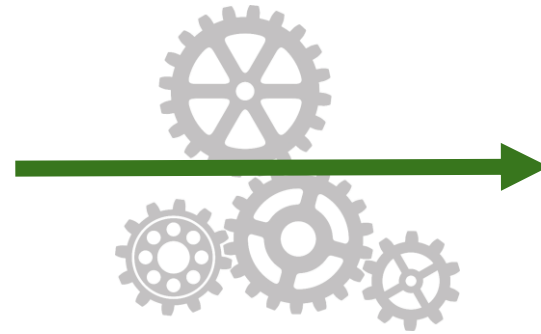
Question Scheme





# Reusable items in Lifecycle

**Metadata structured by  
DDI Lifecycle standard**

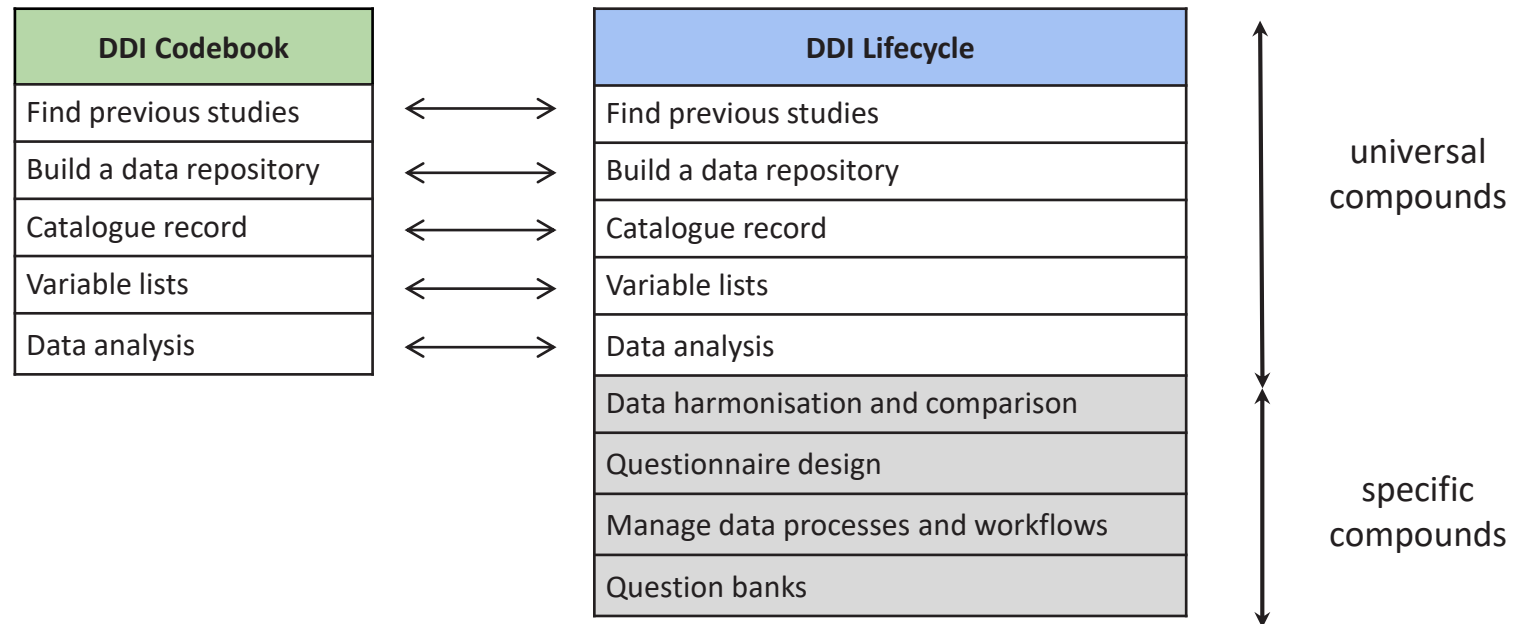


**Metadata specified  
by DDI Lifecycle standard**





# Purposes of DDI Versions





# Codebook or Lifecycle

## DDI-Codebook

```
<qstn>  
<preQTxt>  
CARD 3  
</preQTxt>
```

```
<qstnLit>  
A8. Using this card, generally speaking, would  
you say that most people can be trusted, or  
that you can't be too careful in dealing with  
people? Please tell me on a score of 0 to 10,  
where 0 means you can't be too careful and 10  
means that most people can be trusted.
```

```
</qstnLit>  
</qstn>
```

## DDI-Lifecycle

```
<QuestionItem scopeOfUniqueness="Agency">  
  <URN>urn:ddi:it.istat:B2:1.0.0</URN>  
  <QuestionText>  
    <LiteralText><Content>Sempre nella settimana che  
    va "DA LUNEDI' ... A DOMENICA..." Lei aveva  
    comunque un lavoro che non ha svolto, ad esempio:  
    per un periodo limitato di ridotta attività, per  
    malattia, per vacanza, per cassa integrazione  
    guadagni, etc.?</Content></LiteralText>  
  </QuestionText>  
  <CodeRepresentation blankIsMissingValue="true">  
    <CodeSchemeReference>  
  
    <URN>urn:ddi:it.istat:CodeListYesNo:1.0.0</URN>  
    </CodeSchemeReference>  
  </CodeRepresentation>  
</QuestionItem>
```



# DDI-Lifecycle coverage





# Benefits of DDI

- Interoperability
- Rich content
  - Granular
  - Expansive
- Increased search capability
  - Precision in searching
- International community



# Available Standards

## DDI Codebook 2.5

DDI-Codebook is a more light-weight version of the standard, intended primarily to document simple survey data. Originally DTD-based, DDI-C is now available as an XML Schema.

[Read more about DDI 2.5...](#)

Online field level documentation

XSD Schema entry point

Download documentation

## DDI Lifecycle 3.2

DDI-Lifecycle is designed to document and manage data across the entire life cycle, from conceptualization to data publication and analysis and beyond. It encompasses all of the DDI-Codebook specification and extends it. Based on XML Schemas, DDI-Lifecycle is modular and extensible.

**Implementation Guide: Best Practices for Usage of DDI 3.2 and Future Versions**



[Read more about DDI 3.2...](#)

Online field level documentation

XSD Schema entry point

Download documentation

## Controlled Vocabularies - Overview Table of Latest Versions

What is a controlled vocabulary?

The DDI Controlled Vocabularies Group (CVG) has created a set of controlled vocabularies that can be used with DDI as well as for other purposes and applications. Select DDI Alliance vocabularies are already in use at organizations like the Finnish Social Science Data Archive (FSD), the GESIS - Leibniz Institute for the Social Sciences, the Inter-university Consortium for Political and Social Science (ICPSR), Mathematica Policy Research, the UK Data Archive (UKDA), and the University at Bielefeld, Germany. Nesstar Publisher (<http://www.nesstar.com/>) now incorporates the controlled vocabularies for Analysis Unit and Time Method.

A paper on "Controlled Vocabularies for DDI 3: Enhancing Machine-Actionability" provides additional background on this effort.

### Download

Download package with latest versions of all DDI Controlled Vocabularies: DDI-CV\_2018-07-18.zip

- + Available Formats
- + Usage
- + Translations
- + Publication, Maintenance, and Management
- + Versioning Policy

## DDI RDF Vocabularies

**Current status: The vocabularies underwent public review in 2014. XKOS was officially published in June 2019. Disco has been updated for public review and publication in 2019, and PHDD has been postponed due to similar development in related vocabularies like CSV on the Web.**

The DDI Alliance has supported work on three RDF vocabularies, including: XKOS, an RDF vocabulary for describing statistical classifications, which is an extension of the popular SKOS vocabulary; the DDI-RDF Discovery vocabulary for publishing metadata about datasets into the Web of Linked Data (Disco); and PHDD, a vocabulary for describing existing data in rectangular format.

- [XKOS - Extended Knowledge Organization System](#)
- [Disco - DDI-RDF Discovery Vocabulary](#)
- [PHDD - Physical Data Description](#)
- [Mailing list](#)
- [Development history](#)



# Challenges of DDI

- Complexity
- Level of buy-in
- Involves initial investment but saves costs in the long-term
  - May involve changes to processes and systems
  - Investment in technology tools may be considerable
- Legacy metadata may require updating
  - Consistency issues
  - Format transformations
- Training is required



# Summary

- DDI *facilitates* the creation of metadata
- **Expressed in XML**
  - The XML schema is a way of tagging text for meaning, not appearance
- DDI allows you to have a clear boundary between the information required to define and understand data and the data itself
- Understanding the differences between DDI Codebook and Lifecycle allows you to make the appropriate choice for your use case
- Part of the decision about which version to implement should be informed by your existing levels of information and your available skills and technology stack



Steine unsortiert

woodleywonderworks,  
<https://www.flickr.com/photos/wwwworks/2472232245> (CC-BY)



Steine sortiert

Windell Oskay,  
<https://www.flickr.com/photos/17425845@N00/2156888497> (CC-BY)



Bagger (3516880947\_of44a89c1c\_z.jpg)

Stephen Edmonds:  
<https://www.flickr.com/photos/popcorncx/3516880947/> (CC-BY-SA)



Bagger (lego-717196\_960\_720.png)

Desktopnexus.com:  
<https://abstract.desktopnexus.com/wallpaper/2425074/>



Bagger (3514881626\_be3e87cc58\_o.jpg)

Stephen Edmonds:  
<https://www.flickr.com/photos/popcorncx/3514881626/> (CC-BY-SA)



Bagger (4485538519\_d4ef5e284b\_o.jpg)

Stephen Edmonds:  
<https://www.flickr.com/photos/popcorncx/4485538519/in/album-72157617802609935/> (CC-BY-SA)