A Methodology for the Development of DDI-CDI Implementation Guides

[Glossary](#_44n4t14083gm) **3**

[**Introduction**](#_h7lft3rw8v14) **4**

[Purpose](#_wkwg5zpsxltm) 4

[Sharing and reuse of statistical and research data](#_grrj07y01hsr) 5

[Overview of DDI-CDI](#_xlz4051zxtn2) 5

[Overview of methodology](#_8j0z70fa5t3t) **5**

[Identify (community) functional requirements](#_dvd0jct2osd2) **6**

[High-level data sharing goals](#_o3cwqzmakou6) 6

[Analyze needed description (Process, Data, both?)](#_dh2tkhqoy1ee) 6

[Determine the roles of community actors](#_sxur0jldjahq) 6

[Rights management and data sensitivity](#_y5t8tg5l5ovp) 7

[Determine actions on data](#_trza1lg94rnj) 7

[Determine level of structural granularity of metadataIt](#_2gxglmlv9elj) 7

[**Identify standards, features, and coverage**](#_rljsbjr1fbrq) **7**

[Identify common standards/ontologies used within the community](#_b62ws9sy6p2q) 8

[Identify the existence of DDI-CDI Implementation Guides in other communities](#_9zwf8jdfnzx9) 8

[Which standards support which requirements?](#_y8c6hamo5x76) 8

[DDI-CDI support? Alignment w/ others?](#_b5zbmkkro5wz) 8

[Identify sources of data and metadata within the community](#_j1jwr8kj16to) **9**

[Identify range of available (meta)data types in community](#_xspmi5vvfc5u) 9

[Identify common technology / platforms](#_u55i5pvd2dvy) 9

[Analyze existing community practice for RDM/metadata](#_a237uoo93iiw) 9

[Identify the lack/gaps in community](#_rm8axt7hjmlo) 10

[Consideration of resources and approaches for capture/production of information](#_u1vh6vdas3tz) 10

[Specify DDI-CDI profile/coverage and alignment with other standards](#_szsp5s3h9fw1) **10**

[Determine functional requirements supported by DDI-CDI](#_3oly4i26zjmp) 10

[Identify and analyze samples and consult existing profiles and agreements](#_g5q5rhyo9prz) 10

[Specify community syntax representation for DDI-CDI constructs](#_pfc2c6oi750h) 11

[Integration/alignment with relevant external standards](#_avulnv8mcjun) 11

[Deploy and implement technology and information solutions](#_dueumy6ek5nz) **11**

[Develop software tools to express existing information in DDI-CDI and address gaps in available sources](#_nj1dqxrfnosm) 11

[Creation of metadata as a community resource](#_82ypfjdlcc96) 11

[Capacity building](#_abl3u1xtr3zv) 11

[Specify governance and maintenance versioning](#_u4lg0kiphxfm) **12**

[Maintenance, governance, and conformance](#_6q8xku22yxb6) 12

[Disseminate, promote, and support the IG](#_jphktl875a88) 12

[Appendix](#_p1nfbftquxf3) **12**

[Example implementation: CESSDA (Hypothetical)](#_7xhvvqiyerjl) 13

[Scenario Description](#_x7t3s150aqpr) 13

[Step 1 - Identify (community) functional requirements](#_x9ynxjvq9b3w) 13

[Step 2 - Identify standards features and coverage](#_a5m5zso5hgf1) 14

[Step 3 - Identify sources of data and metadata within the community](#_2u8f1i6q01t7) 14

[Step 4 - Specify DDI-CDI profile/coverage and alignment with other standards](#_yf50o3w6l3na) 14

[Step 5 - Deploy and implement technology and information solutions](#_beqqpa9tpgd9) 15

[Step 6 - Specify governance and maintenance versioning](#_a9pxkujnbhvu) 15

[Identify range of (meta)data types in community](#_mstqthygwt37) 16

#

# Glossary

* Community of practice: Communities of practice are formed by people who engage in a process of collective learning in a shared domain of human endeavor[[1]](#footnote-1). This document uses the term “community of practice” – often abbreviated as “community” – to refer to a defined group of individuals and/or organizations wishing to exchange and reuse FAIR data.
* Datum: singular instance of *Data*.
* Data: Facts, measurements, recordings, records, or observations about the world,[[2]](#footnote-2).
* DDI: The Data Documentation Initiative (DDI) is a family of international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences[[3]](#footnote-3).
* DDI-CDI: [insert one-line definition][insert url of source]
* DDI Lifecycle: [insert one-line definition][insert url of source]
* EOSC: European Open Science Cloud[[4]](#footnote-4)
* FAIR data: [insert one-line definition][insert url of source]
* Functional requirements: [insert one-line definition][insert url of source]
* Implementation guide: [insert one-line definition][insert url of source]
* Interoperable systems: [insert one-line definition][insert url of source]
* Long dataset: [insert one-line definition][insert url of source]
* NSO: [insert one-line definition][insert url of source]
* Personal data: [insert one-line definition][insert url of source]
* PROV-O: [insert one-line definition][insert url of source]
* Sensitive data: [insert one-line definition][insert url of source]
* Wide dataset: [insert one-line definition][insert url of source]

#

# Introduction



## Purpose

The purpose of this document is to provide a methodology for creating implementation guides (IGs) related to DDI-CDI implementations within a particular community of practice[[5]](#footnote-5). By following the guidelines in the following pages,, the reader (for instance, a metadata manager or data steward) will be better equipped to create an implementation guide for their community.

The Implementation Guide for a community enables application and system developers to design and deploy interoperable systems, based on mutually agreed and properly specified metadata supporting the exchange of data and related resources. In this way, a community’s goals for the sharing of data can be realized, informed by a robust and mutually agreed set of requirements.

This document describes a process of analysis and design, incorporating steps that have a narrative order that is not intended to prescribe the order in which those steps are actually performed: indeed, these steps can be performed in whichever order makes most sense to the reader.

Some working examples of the process in action are provided in Appendix yyyy.

[How many examples? Include Estonian classification example. QUERY: Is it really an example of an IG, or just a description of a complete running of the process described to produce one?]

## Sharing and reuse of statistical and research data

[Describe how DDI-CDI impacts the IR of FAIR data sharing - expressed in generic terms for NSIs as well as FAIR data people. See Webinar Series preso <https://codata.org/initiatives/data-skills/ddi-training-webinars/webinar-on-implementing-fair-what-ddi-can-do-for-you/>.]

With the emphasis on FAIR data sharing has come a strong focus on cataloging metadata sufficient for data portals and queries to locate data sets for reuse. Standards exist for the metadata needed to support these functions - DCAT and Schema.org being perhaps the best-known. DDI-CDI is focused on a different aspect of FAIR - interoperability and reuse.

By providing a more granular understanding of what data sets contain, at a structural level, and in terms of their provenance, DDI-CDI makes it possible to better integrate and reuse data. While no single standard can bridge the semantic divide between domains, DDI-CDI provides a framework for making the challenges of semantic mapping - and ultimately, effective harmonization and reuse of data - more tractable. By providing detailed, machine-actionable metadata about how data sets relate, the full range of functionality implied in the FAIR data-sharing vision is promoted.

## Overview of DDI-CDI

[Add the perspective that DDI-CDI is intended to broaden the capacity for FAIR data sharing, potentially beyond the community which is working to develop an IG.]

[A brief introduction of the capabilities of the standard (so you make sense of the discussion below about selection of classes for data description, process description, classifications, etc.) Reference better documentation on the specification.]

* What is DDI-CDI good for
* What is DDI-CDI not good for, e.g. data catalogues

[What can be done with DDI-CDI which cannot be done in DDI Lifecycle, for example. Describe a long data set usefully, cast data between different structures, etc.]

# Overview of methodology

The main body of this document describes the important steps to think about when creating an implementation guide for a community.

The following diagram shows a high level view of this process, starting from the point where the decision to use DDI-CDI, typically in combination with other standards, has been taken. The blue circles are groupings of lower-level activities, shown as green boxes. The order does not prescribe a linear flow or process, but the steps will typically be carried out in the order presented. [Alternate text: This methodology specifies an ordering of steps for creating an implementation guide (shown in Figure X). Variations of this ordering are possible and may be more suitable for some implementors (see REF for background).]



[[https://miro.com/app/board/o9J\_lwX77dE=/?fromRedirect=1](https://miro.com/app/board/o9J_lwX77dE%3D/?fromRedirect=1), Frame 2]

The process to be followed shows a number of steps, briefly characterized here. (Each of these steps is described in greater detail below.)

1. **Identify (community) functional requirements**

Each community of practice will have a different understanding of what is required in terms of FAIR data sharing, and what is possible. An agreed set of specific functions should be identified, so that DDI-CDI and other standards can be employed usefully, and implemented in a way which achieves the data-sharing goals of the community.

1. **Identify standards, features, and coverage**

DDI-CDI is designed to work together with other standards used within the community - it does not necessarily support all the functions which may be required. This step involves looking at the functions to be supported, and determining which standards can be used to support the different functional requirements at a high level.

1. **Identify sources of data and metadata within the community**

The patterns of production, management, dissemination, and use of data within communities of practice varies widely across domains and institutions. The relevant sources of both data and reusable metadata (e.g., vocabularies, etc.), as well as any gaps for supporting the desired data-sharing, should be identified.

1. **Specify DDI-CDI profile/coverage and alignment with other standards**

Once the general use of DDI-CDI has been identified (step 2), a more detailed examination of the model should be conducted, and the needed parts of it identified. This forms the basis of the “community profile” or “implementation guide”.

1. **Deploy and implement technology and information solutions**

Standards are useful for describing and exchanging the information needed within a community, but they must be implemented and deployed in order to meet functional requirements. In many cases, such implementations will involve technology solutions, but there are also organizational functions which involve changes to process and practice. This steps covers both, as important aspects of a complete solution. Further, both aspects of a community wide data-sharing solution may influence what parts of the DDI-CDI model can be practically utilized. This step involves taking these considerations into account.

1. **Specify governance and maintenance/versioning**

Effective use of any standard requires that it be practically supported over time. This step involves identifying needed aspects of this activity, including interfacing with and/or supplementing mechanisms for governing and maintenance of the solution.

# Identify (community) functional requirements

Given the range of metadata in the DDI-CDI model, it is important to have a solid basis for selecting the information which is needed for exchange within a community. Further, DDI-CDI is designed to be used in combination with other standards, and may not support all of a communities’ requirements by itself. The first step in producing an IG for a community is to understand the specific functional requirements: what exchange of data and metadata is envisioned, and for what specific business purpose?

Note that this requirement exists for any technology implementation, and this exercise may already have been conducted within a community before the creation of a DDI-CDI profile. The important thing is to understand what specific functionality - and what metadata - can be modeled and supported by the DDI-CDI standard. Such an analysis is the starting point for the selection of a subset of the overall DDI-CDI model for community use.

Implementation guidance for this step should address the following activities.

## High-level data sharing goals

What are the key requirements of data sharing for the community of practice? That is to say, as a first step, in non-technical language, identify what kind of specific data-sharing activities need to be operational following a successful implementation of DDI-CDI. Examples might be “exchanging data at the file level”, “being able to integrate datasets programmatically”, “providing machine-readable variable descriptions” and so on.

## Analyze needed description (Process, Data, both?)

* establish objects e.g. data vs documentation vs metadata
* relations between them
* versioning i.e. will these objects change over time?

## Determine the roles of community actors

* data/metadata producers
* data/metadata publishers
* data/metadata aggregators/portal operator
* consumers (human and machine)
* etc.

## Rights management and data sensitivity

DDI-CDI is *not* intended to handle statements about the rights of community actors to perform particular actions against particular objects (such as, e.g. “as an accredited researcher, I can freely download a dataset; as a non-accredited researcher I have to seek the data producer’s approval first”). Nonetheless, the context of rights management should be considered in any DDI-CDI infrastructure as it may well have structural consequences, and permission, obligations and prohibitions on data can and do change over time which may then introduce unexpected structural constraints.

Key topics that may be considered among others are:

* Rights management, described formally with [ODRL](https://www.w3.org/TR/odrl-model/) (Open Digital Rights Language) which models how (meta)data and documentation at all levels of granularity can be accessed by whom and under what conditions.
* Data sensitivity / personal data i.e. which data might need to be made available in different security contexts, should it contain sensitive variables such as gender identity or income. The Data Privacy Vocabulary ([DPV](https://dpvcg.github.io/dpv/)) is an emerging W3C standard, developing a taxonomy of privacy and data protection related terms.
* Disclosure control i.e. to what extent data is inherently disclosive i.e data which can identify a specific individual without their consent or secondly, how data might become disclosive in combination with other data.

## Determine actions on data

* interchange
* harmonize
* integrate
* disseminate

## Determine level of structural granularity of metadataIt

* collection e.g. time series or a study
* dataset
* variable
* datum
* high level process
* for example, [PROV-O](https://www.w3.org/TR/2013/REC-prov-o-20130430/) applications

# Identify standards, features, and coverage

Based on the functional areas and actions identified, the ability of the community to provide and exchange data and metadata will depend on agreed, standard expressions. DDI-CDI will cover some important aspects of this information exchange, but will typically need to be combined with other standards to support functional requirements. This stage of the analysis identifies which standards are potential candidates for use in the modeling and exchange of information to support specific functions and actions.

Accordingly, implementation guidance for this step should address the following activities.

## Identify common standards and ontologies used within the community

* Survey common standards implemented widely within the domain
* Determine how prevalent they are
* Mention examples: domain ontologies and vocabularies, [PROV-O](https://www.w3.org/TR/2013/REC-prov-o-20130430/), [DCAT](https://www.w3.org/TR/vocab-dcat-2/), [Schema.org](https://schema.org/), [DDI Codebook](https://ddialliance.org/Specification/DDI-Codebook/2.5/), [DDI Lifecycle](https://ddialliance.org/Specification/DDI-Lifecycle/3.3/), [OpenAire](https://guidelines.openaire.eu/en/latest/data/introduction.html), [OAI-PMH](https://www.openarchives.org/pmh/), [ResourceSync](https://www.openarchives.org/rs/toc), [OMOP CDM](https://www.ohdsi.org/data-standardization/the-common-data-model/), etc.

## Identify the existence of DDI-CDI Implementation Guides in other communities

* Are there DDI-CDI IGs in communities from which you wish to re-use data? Within communities that are likely to want your data?
* Determine if these IGs can become inputs to your own community IG

## Which standards support which requirements?

* Take the list of standards and identify which functional requirement they support. Note that some standards may support more than a single functional requirement, and some functional requirements may be met by more than one standard commonly used within the community.
* The point of this activity is to establish a list of possibilities, not to select specific standards for particular requirements
* Consider that DDI-CDI is not a domain-specific standard but is generic across domains, and there may be standards within the domain for describing data or processes which can also be described in DDI-CDI. This should be noted.

## DDI-CDI support? Alignment w/ others?

* Identify which functional requirements are being supported with DDI-CDI
* Determine which other standards will potentially need to be aligned/mapped to DDI-CDI
* Example of this type of decision: Using DCAT-AP (for cataloguing) with DDI-CDI (for describing long and wide data sets)

#  Identify sources of data and metadata within the community

Independent of the standards and models used, the requisite information to support all identified functional requirements identified in (1) must already exist or be produced by community actors. Needed information may already exist in various tools and applications employed by the community, or may be managed and made available in standard or documentary form. This step involves identifying the sources of needed information of all types, as determined by the functional requirements.

Implementation guidance for this step should address the following activities.

## Identify range of available (meta)data types in community

1. Describe the data and the associated metadata at a high level (identify types of data)
* structural: wide or long, etc.?
* coverage: topical, temporal, spatial
* Identify stewards of different data and metadata types - which member performs which functions? (collection, editing, management, dissemination, etc.)
* Identify relationships between types of data
1. Associate types of data with the sources, processors, managers, and consumers (etc.)
* Which community members produce/consume which types?

## Identify common technology / platforms

[The standard is platform-independent, but it is necessary to address existing solutions.

Common applications in the Social Sciences are Colectica and Dataverse. Additionally, legacy Nesstar installation may still need to be supported or a migration should precede the implementation of CDI.]

* Consider not only what tools are used, but how they are used
* Consider the extent of use (all actors of a particular type in the community, or only some of them)
* Consider the information which exists but is not necessarily easily exported from tools as well (metadata embedded in stats packages, for example)

## Analyze existing community practice for RDM/metadata

* If there is a culture of metadata management within the community, even if not relying on specific technologies, standards, or tools, then this should be identified.
* Often, Excel spreadsheets contain a huge amount of metadata which is not actively managed, but may represent a resource to be mined (as one example). Other similar resources may exist.

## Identify the lack/gaps in community

Consider the desired actions and set of information resulting from the analysis of functional requirements in step 1.

Compare the information needed to support identified actions with the available data and metadata, Identify any gaps between the two.

Identify any functional gaps between the desired actions and current community practice. Identify the needs which cannot be met.

## Consideration of resources and approaches for capture/production of information

* Consider possible approaches to filling the gaps (technological and at the level of practice)
* Consideration of the cost of extracting and transforming metadata and other information
* Consideration of the costs of producing new metadata and other needed information
* Consider short-, medium-, and long-term actions and the overall plan for the community

# Specify DDI-CDI profile/coverage and alignment with other standards

## Determine functional requirements supported by DDI-CDI

* Look at specific actions and types of data identified above
* Compare with the available classes in DDI-CDI to determine extent of coverage
* Consider which other standards might be used to address any gaps

## Identify and analyze samples and consult existing profiles and agreements

[Can we reference a detailed document from the Data Structures group]

[What about process description? Is it the same?]

* Comprehensive set of examples
* Existing metadata profiles may be a better source for identifying what is needed in DDI-CDI

[Multi-step process for analyzing data examples:

1. Identify classes
2. Identify properties, associations, and limits on multiplicity
3. Identify which vocabularies will be supported for which purposes]
4. Consider mapping domain terminology to DDI-CDI terminology

## Specify community syntax representation for DDI-CDI constructs

choose between: JSON, Turtle, XML

* may be multiple syntax bindings
* may use different syntaxes for specific types of information
* should cover not only syntax expression of DDI-CDI metadata, but also agreement on how data is physically formatted/expressed (as CSV files, as an XML format, as a form of RDF, etc.)

## Integration/alignment with relevant external standards

Identify points of contact between functional requirements covered by DDI-CDI and those covered by other standards, and determine what is needed in DDI-CDI to support integration. May happen at the level of the model/profile, or at the level of the syntax representation. In some cases, there may be no contact at all.

# Deploy and implement technology and information solutions

## Develop software tools to express existing information in DDI-CDI and address gaps in available sources

Select the approaches to be used in creating metadata and data according to the agreed profile. This will include migration from existing stores, mapping and transforming metadata from other current formats, and developing tools for filling gaps in the needed metadata.

## Creation of metadata as a community resource

[Example is the production of ELSST within the CESSDA community to facilitate classification of data across the community and enable multi-lingual search]

## Capacity building

Training for technologists and implementers, RDM practitioners, and other users (etc., etc.)

Publishing documentation and supporting resources

Operation of community services

Provision of community metadata resources

# Specify governance and maintenance/ versioning

## Maintenance, governance, and conformance

* What is the governance of the IG within the community of practice? Who does an implementer contact if they have issues? What is the expected process/timeline?
* How is the IG maintained and versioned?
* How does another community which is interested in sharing data engage with the community for which the IG is being written? This can turn into harmonizing IGs of two communities, etc.
* Any conformance or certification mechanisms in force within the community should be described, including self-conformance, testing/validation tools, certifications, etc.

## Disseminate, promote, and support the IG

[Look at Fairsharing.org]

* The IG should be published in a fashion which provides for maximum visibility not only for the members of the community, but also for external parties which might be interested in sharing data or metadata with the community. It serves both as a guide for community implementers, and as a guide to the specific “dialect” of DDI-CDI used by that community in its systems, and specified in their IG.
* Repositories and other highly visible centralized resources for describing FAIR sharing within communities should be leveraged to the extent possible. This will include putting menmtion of a community IG in the appropriate FAIR Implementation Profile (if one exists). Sites such as Fairsharing.org should be leveraged where possible.
* If members of the community are encouraged or expected to expose data and/or metadata for reuse in DDI-CDI, then appropriate catalogs and portals - both within the community and at a more generic level - should be identified. (For example, an NSO in Europe might be able to inform possible users through repositories operated by Eurostat of the European Commission, etc.)

# Appendix

## Example implementation: CESSDA (Hypothetical)

### Scenario Description

[Insert one-line description of CESSDA, e.g. “CESSDA stands for Consortium of European Social Science Data Archives. CESSDA provides large-scale, integrated and sustainable data services to the social sciences. It brings together social science data archives across Europe.”

To enable data sharing across domain boundaries within the EOSC, CESSDA wants to provide access to their study data in DDI-CDI. This will enable a researcher, independent of their scientific domain, to load the data into a suitable environment, e.g. a Jupyter Notebook Service provided by an e-Infrastructure or running it locally on their local machine. For the purpose of this hypothetical use case, we assume that all data can be shared freely and that there exist libraries or bindings for DDI-CDI in the favoured programming language.

#### Step 1 - Identify (community) functional requirements

 **High-level data sharing goals**

CESSDA’s data holdings are spread across 23 national data archives in as many European countries and a wide variety of languages. CESSDA exists to share data primarily within the social sciences, and uses metadata standards widely accepted within that community. For traditional social science data sets, the ability to reuse data already exists. The idea behind using DDI-CDI is to broaden the range of potential users of the archives’ holdings beyond the social sciences, and to make it easier for non-traditional data to be integrated with existing holdings.

Research around topics like COVID-19 demand that data from many domains be integrated in single research projects, and the study of populations is often a key element in this type of research. In addition to supporting these new types of users, there is an increasing need for the CESSDA archives to handle data which is non-traditional in form (e.g., sensor data, administrative data, social media data).

**Analyze needed description (Process, Data, both?)**

The primary need is to be able to describe data sets of different types. The description of processing is an additional type of metadata which could be valuable in explaining the context and trustworthiness of reused data for researchers outside the social sciences. This is a lower priority than the description of data structures, but will be included for those cases where it is available.

**Determine the roles of community actors**

There are several actors within this scenario, including the archives which constitute the membership of the CESSDA network, the CESSDA main office responsible for the centralized services and standards which allow coordinated operations within the network (i.e., the CESSDA Data Catalogue for discovery across the entire network, metadata specifications which will be shared, etc.). The users of data are also actors, and they include both social science researchers and those from outside the domain. We will also consider machine consumption of data to indicate that these are also actors in this scenario.

**Rights management and data sensitivity**

Access conditions and similar information are shared between the actors in this hypothetical scenario. Although the disclosure control and sensitivity issues are delegated to the archives, for discovery purposes this information needs to be exchanged.

**Determine actions on data**

 Actions required on the data include:

* Indexing variables to support search across data of all forms
* Providing information about the relationships between concepts (e.g., variables and vocabularies) in disparate data sets to support both discovery and harmonization and integration by users
* The responsibility for transforming data structures and other types of integration/harmonization based on information provided by CESSDA would be placed on the researchers, particularly those coming from other domains. (It may be the case that individual archives choose to provide services in this area, but these are not considered actions for the community in this scenario.)
* Advertising the availability of metadata in machine-actionable form is a function of the CESSDA catalogue; providing that metadata in those forms is only possible at the individual archive sites. The CESSDA catalogue is a consumer of such metadata, but does not manage or disseminate it.

**Determine level of structural granularity of metadata**

It is expected that services will be provided at different levels by individual archives, and that these levels will correspond to collections of data (e.g., longitudinal studies), individual studies potentially with multiple data sets of different types, individual data sets, and variable-level information. This granularity will be reflected in the information shared between actors of different types within the network, and will need to be addressed.

#### Step 2 - Identify standards features and coverage

Identify common standards/ontologies used within the community

Identify the existence of DDI-CDI Implementation Guides in external communities

Which standards support which requirements?

DDI-CDI support? Alignment w/ others?

#### Step 3 - Identify sources of data and metadata within the community

Identify range of available (meta)data types in community

* DDI Codebook

Identify common technology / platforms

* Nesstar - being phased out
* Dataverse
* Colectica

Analyze existing community practice for RDM/metadata

Identify the lack/gaps in community

Consideration of resources and approaches for capture/production of information

#### Step 4 - Specify DDI-CDI profile/coverage and alignment with other standards

Determine functional requirements supported by DDI-CDI

Identify and analyze samples and consult existing profiles and agreements

Specify community syntax representation for DDI-CDI constructs

Integration/alignment with relevant external standards

#### Step 5 - Deploy and implement technology and information solutions

Develop software tools to express existing information in DDI-CDI and address gaps in available sources

Creation of metadata as a community resource

Capacity building

#### Step 6 - Specify governance and maintenance versioning

Maintenance, governance, and conformance

Disseminate, promote, and support the IG

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* Case study CESSDA:
	+ Given an existing DDI Codebook 2.5 file, how to turn it into a CDI file
	+ <https://dbkapps.gesis.org/dbkoai/?verb=GetRecord&identifier=oai%3Adbk.gesis.org%3ADBK%2FZA1298&metadataPrefix=oai_ddi25>
	+ <https://snd.gu.se/en/oai-pmh?verb=GetRecord&identifier=2020-59&metadataPrefix=oai_ddi25>
	+ How can someone else re-use that file?
* Given a census file, turn it into a CDI file.
* Case study Stats Estonia:
	+ Classifications and concepts in DDI Lifecycle
	+ Publishing metadata for reuse (FAIR sharing)

----

Implementation guide for applying CDI <-> implementing a whole system and figuring out whether/how CDI fits in

Questions:

* Why implement CDI?
	+ Is it easy if you already have DDI Codebook/GSIM files?
	+ Is it easy to turn CDI into DC?

FAIR -> CDI is about the I and the R -> to define Functional Requirements

Requirements of researchers:

* Can I use this? – Assessment = Exploration
	+ taking a first look at a dataset
	+ figure out what it is, reading description/metadata
	+ identify variables, CVs used etc
	+ playing with the data in a statistics software, identify frequencies etc
	+ provenance (e.g. curation level)
	+ ⇒ Interoperability
* How do I use it with R/python or stata? Jupyter Notebooks?

provenance vs processing

* process vs data descriptions

## Identify range of (meta)data types in community

Interoperability with existing standards

* DDI Lifecycle
* DDI Codebook
* SDMX
* DCAT-AP
* Schema.org
* Dublin Core
* OpenAIRE / B2Find
* SKOS
* Neuchâtel
* GSIM/GSBPM/GLBPM
* PROV
* BPMN
* Domain ontologies
* Domain standards
* (OAI-PMH)

Contributors so far:

Pier Luigi Buttigieg

Laura Molloy

Marta Limmert

Darren Bell

Hilde Orten

Carsten Thiel

Knut Wenzig

Arofan Gregory

(Check Edourd, Flavio, etc.)

1. <https://wenger-trayner.com/introduction-to-communities-of-practice/> [↑](#footnote-ref-1)
2. <https://casrai.org/term/data/> [↑](#footnote-ref-2)
3. [Adapted from] <https://ddialliance.org/> [↑](#footnote-ref-3)
4. <https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en> [↑](#footnote-ref-4)
5. [↑](#footnote-ref-5)