

The Role of DDI-CDI in EOSC: Report on Activities

Authors

Arofan Gregory, Simon Hodson, and Joachim Wackerow

Please cite this document as: Gregory, A., Hodson, S., Wackerow, J., 2021, 'The Role of DDI-CDI in EOSC: Report on Activities', <https://doi.org/10.5281/zenodo.4707290>

Acknowledgment

This work was supported by the EOSC Secretariat.

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644.

Contents

I.	Overview	2
II.	General Approach	2
III.	Meetings and Events	3
A.	Project Meetings and EOSC Groups.....	3
B.	Focused Meetings for Examples/Use Cases	4
C.	General Events.....	6

I. Overview

This EOSC co-creation project analysed the innovative capabilities of the new domain-agnostic Data Documentation Initiative-Cross Domain Integration (DDI-CDI) specification (currently in review and to be released in a first production version in June 2021). The project examined how the specification could be applied to much-needed search and data integration functions within EOSC. It did so through a structured consultation with a range of experts and initiatives as detailed below.

The main output is a significant report ‘The Role of DDI-CDI in EOSC: Possible Uses and Applications’ <https://doi.org/10.5281/zenodo.4707263> that details the substantive exploration and findings from this process. A number of specific and important issues were explored, and examples given of the uses and applications of the specification for EOSC. The report makes recommendations for EOSC, EOSC stakeholders and for the DDI Alliance as the curator of the DDI-CDI specification.

This document describes the approach taken in the investigation and analysis of how the DDI Cross Domain Integration (DDI-CDI) specification could be used to support the EOSC data-sharing infrastructure. It describes the process for the investigation and identification of recommendations, and enumerates and describes the meetings, interviews, and events in which the authors of the report took part.

These activities fall into three categories:

1. Attendance at project meetings for this effort and related groups within the EOSC frame and within the DDI Alliance
2. Meetings requested with stakeholders in the EOSC community around the examples and topics covered in the report
3. Presentation to and participation in general events related to the topic

The general approach is summarized, and the specific meetings and events in each of the three categories are briefly described.

II. General Approach

Specifications such as the DDI-CDI model are not implemented in isolation: they form part of a system which has many different aspects and will involve a number of other conceptual and technical elements in its architecture and realization. To explore and identify the appropriate use of

a specification such as DDI-CDI within this framework, the relevant requirements of the architecture must be characterized, and other available standards and models must be examined.

The project team was involved in the on-going developments of the DDI-CDI specification, and this provided a substantial understanding of the capabilities of the specification, and the direction in which it is evolving as it approaches publication, anticipated for mid-2021. This also allowed for discussion within the development team of issues which came up during the course of the report, so that all needed clarification regarding the specification and its intended use could be easily obtained.

An understanding of the requirements of the EOSC infrastructure was drawn not only from meetings and discussions, but also consideration of published documents and in-progress drafts. Discussions, presentations, and documents around the FAIR ecosystem were also considered. Through this process, the project team was able to refine and summarise its understanding of the relevant infrastructure requirements from an EOSC perspective.

The DDI-CDI specification and the objectives of the project were presented to a range of experts and EOSC stakeholders through a series of meetings. Most notable were the meetings with the EOSC Interoperability WG, the FAIRsFAIR Metadata Catalogue Group and the dedicated event on 20 November to which a range of experts, stakeholders and EOSC clusters and ERICs were invited.

These events were instrumental in identifying the series of examples/use cases and topics for further exploration. These were not chosen to be a complete representation of all possible use cases, but rather to exemplify the ways in which DDI-CDI could be used to support solutions, which will in many cases also involve other standards, models, and technologies. The topics and examples/use cases were then explored and examined through a series of intensive meetings with the relevant experts and stakeholders, as detailed further in Section III B below.

On the basis of these examples, a series of recommendations and suggested possible next steps were identified. These include activities which will involve identifying EOSC requirements and communicating them to the DDI Alliance for inclusion in the further development of the DDI-CDI specification.

III. Meetings and Events

A. Project Meetings and EOSC Groups

These events involved consideration of the working documents of background material for EOSC broadly, and the specific drafts from groups within EOSC working on topics related to infrastructure and metadata. They include regular project meetings of 1-2 hours held weekly or more often, on an as-needed basis.

1 Sept. 2020: Regular planning meetings, at least weekly, from 1 September.

Regular planning meetings, as needed, 1-2 hours, covering planning and scheduling for other events, and discussion and analysis of specific topics.

11 Sept. 2020: First FAIRsFAIR workshop - Metadata catalogue integration for interdisciplinary research

The FAIRsFAIR work on integrating metadata catalogues provides a significant point of contact between metadata used for discovery, and that used for data integration and reuse. DDI-CDI was introduced at this initial meeting and it was agreed with the group that the role of

DDI-CDI in EOSC and its relationship with other specifications (including DCAT) would be an important topic of consideration.

5 Oct. 2020: Meeting with EOSC Interoperability Task Force.

The EOSC Interoperability Task Force addresses issues which lay the groundwork for the overall architecture and the way in which EOSC will implement the FAIR ecosystem. The DDI-CDI specification and its role were discussed. The EOSC Interoperability Framework provides a preliminary, but incomplete, view of the role that DDI-CDI can play providing structural metadata. This is discussed further in the report that is an output of the present project.

9 Oct. 2020: Second FAIRsFAIR workshop - Metadata catalogue integration for interdisciplinary research

This meeting was a continuation of the earlier workshop, looking at how DDI-CDI fit into the broader metadata picture. The workshop identified alignment with DCAT-AP as a key example in illustrating how catalogue metadata and more detailed metadata might interact.

21 Oct. 2020: Lightning talk on the project to EOSC Symposium

A brief presentation of the work of the group to other EOSC efforts.

20 Nov. 2020: DDI-CDI and EOSC Workshop

This was the key workshop to present the project to a wide range of EOSC stakeholders. Among the 47 participants were representatives from the FAIR Interoperability Working Group, the FAIRsFAIR Metadata Catalogue Group, EOSC Clusters and ERICs and a range of metadata experts. The workshop featured an extensive presentation of the DDI-CDI specification and the objectives of the project. The event was instrumental in identifying a number of the examples/use cases which form the core of the report and are further detailed below in section B. The recording and presentation are available on the [event page](#).

30 March 2021: FAIRsFAIR Metadata catalogue workshop – Pilot 1

This meeting focused on how the FAIRsFAIR metadata catalogue pilot project might be framed. The draft report from this project was circulated in advance and a presentation on the report and its findings and recommendations was given. It was agreed that the FAIRsFAIR Metadata Catalogue Group would further develop examples, focusing on the interaction of DCAT-AP and DDI-CDI and building on the examples presented in the report.

B. Focused Meetings for Examples/Use Cases

This category includes those events which took place in order to address specific topics for the report. Initially, this included identifying appropriate examples/use cases, and areas for investigation. As these were identified, meetings with specific people who had the needed expertise were arranged. We are particularly grateful to the experts named in the section below for their engagement with the topics, for helping elucidate and refine the examples/use cases and for their willingness to share their insights and knowledge and to provide feedback on presentations, diagrams and draft sections of the report which were shared with them.

28 Oct. 2020: meeting with CESSDA ERIC Representatives, Ron Dekker (director) and Carsten Thiel (CTO)

This was a meeting to help identify examples/use cases and examples for the project, and to get a better sense of how DDI-CDI might fit into a SSHOC/CESSDA frame.

4 Nov. 2020: follow up meeting on CESSDA ERIC

Further discussions based on the above.

16 Dec. 2020: First discussion of DDI-CDI and DCAT with Alejandra Gonzalez Beltran (one of the editors of DCAT Version 2)

The alignment between DDI-CDI and DCAT had been identified with the FAIRsFAIR Metadata Catalogue Group and through the key workshop of 20 November, and was explored in this initial discussion to determine how to proceed.

16 Dec. 2020: First discussion of the UKDA case study with Darren Bell (director of technical services - UK Data Service)

The SERL laboratory work had been suggested as a potential use case, and the UKDA, which handles the data management aspects of that project, provided basic information and sample data.

17 Dec. 2020: Meeting with Dataverse variable metadata working group.

Initial presentation of DDI-CDI to the Dataverse community, Mercè Crosas - co-principal investigator, Danny Brooke - program manager, Gustavo Durand - technical lead and architect, James Myers - senior developer and architect for GDCC, Global Dataverse Community Consortium.

16 Feb. 2021: Dataverse Meeting

Possible topics for a use case were presented and discussed.

17 Feb. 2021: Second UKDA Discussion, Darren Bell

Follow-up with UKDA regarding the SERL use case.

24 Feb. 2021: European Social Survey (ESS) Case Study meeting with Knut Kalgraff Skjåk (deputy director general), Bodil Agasøster (head of section survey & data), Hilde Orten (section for interviews and computer services)

Meeting with representatives of the team at the Norwegian Centre for Research (NSD) Data who are responsible for the ESS Multi-Level Application to obtain data samples and discuss functionality.

25 Feb. 2021: Dataverse Meeting

Discussion about the existing capabilities and plans for the Dataverse platform, in light of potential future implementations/features.

25 Feb. 2021: DDI-CDI and FIPs Meeting

Meeting with Erik Schultes (GO FAIR International Support and Coordination Office, science coordinator) and Barbara Magagna (chair of RDA WG on Interoperable Descriptions of Observable Property Terminology (I-ADOPT)) regarding FAIR Implementation Profiles (FIPs) and the overall FAIR ecosystem.

26 Feb. 2021: DDI-CDI and FDO Meeting

Meeting with Luiz Bonino (GO FAIR International Support and Coordination Office) to discuss current proposals and thinking around FAIR Digital Objects and upcoming activities for further discussion/specification.

5 March 2021: DDI-CDI and DCAT second meeting

Follow-up with Alejandra Gonzalez Beltran regarding alignment between DDI-CDI and DCAT.

9 March 2021: Research Data Alliance Interoperable Descriptions of Observable Property Terminology (I-ADOPT) workshop

Following the discussion with Barbara Magagna regarding FIPs, we attended this workshop to further explore developments in I-ADOPT and determine what relationship this work might have on the use of DDI-CDI from a FAIR ecosystem perspective.

23 March 2021: Dataverse meeting

Presentation of the draft example/use case on Dataverse for the report, discussion of recommendations and planning of next steps.

C. General Events

These activities included substantive contributions to and presentations at events which are relevant for the topic of the report, and which helped bring some different perspectives to the work. Although they were not directly focused on EOSC or the DDI-CDI application within that framework, they were instrumental in building an understanding of how DDI-CDI would fit into a broader ecosystem. The discussions prior to and during the FAIR Convergence Symposium were particularly important in this regard

7 Oct. 2020: Meeting between DDI MRT (DDI-CDI development team) and DRUM (CODATA Digital Representation of Units of Measure) Task Group

In looking at cross-domain data sharing, a standard approach to unit of measure is important. This meeting introduced DDI-CDI to the DRUM effort, to begin a discussion about how DDI-CDI could support the application of approaches emerging from the DRUM work.

8 Oct. 2020: DDI-CDI and DRUM Workshop

This meeting focused on validating DDI-CDI approaches to referencing standard units of measure.

4 Nov. 2020: Discussion of ‘Convergence’ Session for the International FAIR Convergence Symposium (IFCS), including consideration of DDI-CDI in the FAIR ecosystem

This was a planning meeting for an event in which DDI-CDI was to be presented alongside other elements of the FAIR ecosystem to help ground discussions about how all of the related standards and technologies could work together to support data sharing. This discussion was instrumental in preparing and framing the further discussions relating to FIPs and FDOs and those examples in the report (see other listings, below).

12 Nov. 2020: presentation on DDI-CDI, ALPHA Network and INSPIRE at RDA Plenary Epidemiology session

The use of DDI-CDI and approaches taken in the ALPHA Network and INSPIRE project were presented to a group looking at standards for data supporting epidemiological monitoring and surveillance.

30 Nov. 2020: IFCS Symposium Plenary Session, including *inter alia* discusses relationship of DDI-CDI with the FAIR ecosystem

This was the event referenced above, in which DDI-CDI was positioned relative to FIPs, FDOs, and other elements of the FAIR ecosystem.

2 Dec. 2020: IFCS FAIR provenance session, including *inter alia* discussion of role of DDI-CDI

Different standards and implementations for describing provenance, process, and workflows were presented from different domains. This served as an initial workshop in a continuing informal discussion between relevant groups.

3 Dec. 2020: IFCS Infectious disease sessions, including discussion of DDI-CDI, ALPHA and INSPIRE

The approach to describing provenance and the integration of data across domain boundaries (population data and clinical data) as prototyped in the ALPHA Network and INSPIRE projects was presented, alongside work on data sharing for researching infectious disease.

4 Dec. 2020: IFCS DDI Session, includes discussion of relation of DDI Codebook, DDI Lifecycle and DDI-CDI

This was a workshop for introducing all of the DDI Alliance specifications for metadata, for data sharing and integration across domains (DDI-CDI) and for data production and management within the Social, Behavioural, and Economic (SBE) sciences (DDI Codebook, DDI Lifecycle).

9 Dec. 2020: DDI-CDI Provenance Workshop

Following on from the IFCS session, this was a further meeting for exploring the ways in which process and provenance is modeled in different domains. The aim is to identify requirements and solutions which can be practically implemented across scientific domain boundaries.