CESSDA Work Plan 2020

CESSDA Metadata Office (MDO) Task 1

Deliverable 3

Introduction to metadata initiatives of interest to CESSDA and Service Providers

Document info

Dissemination Level	PU
Due Date of Deliverable	30/09/2020
Actual Submission Date	14/09/2020
Type	Report
Approval Status	Approved by CESSDA Tools & Services Working Group leader Mari Kleemola and CESSDA Technical Working Group leader John Shepherdson
Version	V1.3
Number of Pages	p.1 – p.81
DOI	10.5281/zenodo.4621051

Version history

Version	Date	Comment	Revised by
0.1	04.03.2020	Agreed on structure	Kerrin Borschewski, Esra Akdeniz
0.2	14.08.2020	Review by author and resulting amendments	All Authors
0.3	26.08.2020	Peer review and resulting amendments	Review by: Tanja Friedrich; Wolfgang Zenk-Moeltgen
0.4	09.09.2020	Review by MDO partners and resulting amendments	Review by: Lorna Balkan and Taina Jääskeläinen
1.0	14.09.2020	Report sent to CESSDA MO	All authors
1.1	29.09.2020	Review by MDO partners and resulting amendments	Review by: Libby Bishop
1.2	01.10.2020	Revision according to feedback from CESSDA MO and feedback from CESSDA WG Leaders	Kerrin Borschewski, Esra Akdeniz,
1.3	10.12.2020	Revision according to feedback from CESSDA MO and feedback from CESSDA WG Leaders	Kerrin Borschewski, Esra Akdeniz

Author List

Organisation	Name	Contact information
GESIS – Leibniz Institute for the Social Sciences	Kerrin Borschewski	kerrin.borschewski@gesis.org
GESIS – Leibniz Institute for the Social Sciences	Esra Akdeniz	esra.akdeniz@gesis.org
GESIS – Leibniz Institute for the Social Sciences	Sophia Piesch	sophia.piesch@gesis.org

Peer-review

Organisation	Name	Contact information
GESIS – Leibniz Institute for the Social Sciences	Tanja Friedrich	tanja.friedrich@gesis.org
GESIS – Leibniz Institute for the Social Sciences	Wolfgang Zenk-Möltgen	wolfgang.zenk-moeltgen@gesis.org
UK Data Archive	Lorna Balkan	balka@essex.ac.uk
GESIS – Leibniz Institute for the Social Sciences	Libby Bishop	ElizabethLea.Bishop@gesis.org

Contents

List of Figures

List of Tables

Executive Summary

High-quality metadata and metadata enrichment are crucial for both data users and data providers. It does not only ensure efficient data access, discovery, and re-use of Social Sciences (meta)data, it also helps researchers who strive for efficient and valuable research, data sharing and citation. From the perspective of data providers, archives, and the CESSDA consortium, data exchange and interoperability of metadata are a crucial aspect for the Consortium of European Social Science Data Archives European Research Infrastructure Consortium (CESSDA ERIC) and especially for CESSDA products and services. For this report, the CESSDA Metadata Office (MDO Task 1) Work Plan Task has selected four different metadata initiatives that already play a significant role for CESSDA or might gain importance in the future. The selection was made based on the questions and inquiries MDO has received in the past from CESSDA SPs and based on the initiatives reflected in CESSDA's policies and the CESSDA strategy 2018-2022. Each of the four initiatives is described and their implications for CESSDA are evaluated. The summarized main results of this evaluation can be found in section 7 of the report. The four initiatives addressed in this report are the Data Documentation Initiative (DDI) standard, the FAIR principles, the Dataverse Network (DVN) project, and the European Open Science Cloud (EOSC) together with the Social Sciences and Humanities Open Cloud (SSHOC) project as part of EOSC.

The content of this report is intended for CESSDA SPs, aspiring CESSDA members, CESSDA MO, and anyone interested in the topics in general and in the metadata work of CESSDA in particular. The report is supposed to provide information for current CESSDA SPs and aspiring CESSDA SPs on the initiatives relevant for CESSDA's metadata work, and guidance for SPs for their general considerations on the quality and possible upgrades of their metadata. Furthermore, the report will provide information for the public and display information on what is essential for CESSDA concerning the metadata work and how CESSDA interacts with and reflects relevant metadata initiatives. Additionally, this report is intended to provide recommendations for strategic decisions on CESSDA's metadata developments and a knowledge base that allows easy and grounded decision-making.

Abbreviations and Acronyms

CDC	CESSDA Data Catalogue
CESSDA ERIC	Consortium of European Social Science Data Archives European Research Infrastructure Consortium
CMM	CESSDA Metadata Model
DC	Dublin Core
DVN	Dataverse Network
DDI	Data Documentation Initiative
DDI-C	DDI Codebook
DDI-CDI	DDI - Cross Domain Integration
DDI-L	DDI Lifecycle
EDDI	European DDI User Conference
EOSC	European Open Science Cloud
EQB	CESSDA Euro Question Bank
FAIR	FAIR principles – Findable, Accessible, Interoperable, Reusable
IQSS	Institute for Quantitative Social Science
MDO	CESSDA Metadata Office Task 1 2020 Work Plan Task
MO	Main Office
NADDI	North American Data Documentation Initiative Conference
PID	Persistent Identifier
RDA	Research Data Alliance
SBE	Social, behavioral, and economic sciences
SDMX	Statistical Data and Metadata Exchange
SSH	Social Sciences and Humanities
SPs	Service Providers
SSHOC	Social Sciences and Humanities Open Cloud project
UML	Unified Modeling Language
W3C	World Wide Web Consortium
XMI	XML Metadata Interchange
XML	Extensible Markup Language

1) Introduction

For scientific research, high-quality metadata (data about data) is an essential part of data exchange and interoperability. With metadata, research data becomes interpretable, useful, and valuable. Researchers profit significantly from well-documented metadata, which make their research visible, replicable, and reusable to other researchers and their science community. In several disciplines, it has become good practice for researchers to share their data, and data sharing requires metadata. Data archives and repositories are the most important focal points for researchers who want to store their data, make them citable, reusable and visible, and open their data for sharing. Currently, the Consortium of European Social Science Data Archives European Research Infrastructure Consortium (CESSDA ERIC) brings together 22 European data archives, which archive and provide data on different scientific subject areas and support researchers. CESSDA has become a success story about the cooperation of archives, providing infrastructures, policies, and participating in internationally relevant projects resulting in significant and relevant services for the scientific community. CESSDA products make it possible for users to find multilingual (meta)data from different social science archives in Europe.

In order to implement those products and ensure the exchange of metadata between the archives, the Service Providers (SPs) of the consortium agreed on one "metadata language" for their metadata. The CESSDA community has decided to use the Data Documentation Initiative (DDI) metadata standard for its study-level documentation. The CESSDA products CESSDA Data Catalogue (CDC), and CESSDA Euro Question Bank (EQB) are all based on DDI, as well as the CESSDA Metadata Model (CMM) (a CESSDA wide metadata schema, which serves as a reference for (rich) metadata). The DDI standard has many benefits, especially the assurance that metadata are well documented and interoperable. Interoperability is also one of the principles of the FAIR (Findable, Accessible, Retrievable, and Interoperable) guidelines, which CESSDA has committed to.

The above-mentioned significance of metadata is reflected in the high number of diverse metadata standards, metadata initiatives, projects, working groups, best practice guidelines, conferences, and innumerable metadata schemas. CESSDA has always kept an eye on various metadata initiatives and has even been partly involved in the ones most relevant for the consortium. CESSDA has, furthermore, implemented principles and strategies of other initiatives, wherever beneficial.

A listing of different standards, initiatives, projects, etc. – for the arts and humanities, social and behavioral sciences – can be found in the appendix. For the core of this report, MDO decided to investigate the metadata initiatives from that list currently most relevant for CESSDA, CESSDA's strategic considerations, and the CESSDA SPs. The choice was made

based on the questions and inquiries MDO has received in the past from CESSDA SPs and based on the initiatives reflected in CESSDA's policies and the CESSDA strategy 2018-2022. The questions MDO received mainly concerned DDI in general, the different DDI specifications, options, or possible imperatives to upgrade the SP's metadata, the relevance of the new DDI specification (DDI – Cross Domain Integration (DDI-CDI)) for the work within CESSDA, and the Dataverse Network (DVN) project. As mentioned above, the use of the DDI metadata standard within CESSDA has a long tradition, and CESSDA's metadata materials are based on different specifications of DDI. Hence, DDI is critical to CESSDA. The DVN and the usage of Dataverses are becoming more and more important within the CESSDA community as well. An increasing number of CESSDA SP's have made or are considering using DVN. The CESSDA strategy 2018-2022, as well as CESSDA policies, refer to the FAIR principles and EOSC. Those initiatives are also extremely relevant for CESSDA. Due to the aforementioned reasons, the four initiatives MDO chose for this report are: the DDI standard, the FAIR principles, the DVN project, and the European Open Science Cloud (EOSC) with a focus on the Social Sciences and Humanities Open Cloud (SSHOC) project.

This report aims to introduce and describe FAIR and EOSC including SSHOC on a general level with the goal of evaluating the implications for CESSDA. As mentioned above, DDI is very important within the CESSDA context and for the social and behavioral sciences in general. Like any other good standard, DDI is constantly being developed, and new specifications of the standard are being created. It is therefore worth looking at the use of DDI for the CESSDA products and CMM and evaluate if the currently used specifications and versions of DDI are still the most suitable for CESSDA needs. This is of particular interest since DDI recently developed a new specification (currently under review) that differs considerably from the existing specifications. The DVN project and use of a Dataverse instance is evaluated in this report, since it is part of the SSHOC project and also since some of the Data Archives in the consortium are using NESSTAR for their documentation (NESSTAR is a product based on one of the earliest versions of DDI and is no longer maintained) and are considering using the DVN software as a "next step possibility". The initiatives reflected in this report are also partly interlinked with each other. EOSC and, hence also SSHOC, adhere to the FAIR principles and consider them for their goals and tasks. Widespread metadata standards such as DDI provide the means to make (meta)data findable, accessible, reusable. Furthermore, DDI metadata is interoperable. Since the metadata covered in the DVN application and within Dataverses are interoperable with systems using DDI, they also adhere to the FAIR principles.

As the extensive list in the appendix shows, many other metadata initiatives for the Social Sciences exist that are not considered in detail in this report. This is because those initiatives are currently less relevant to the consortium than the ones the report addresses. However, it

is not ruled out that they might gain in relevance for CESSDA. Therefore, the consortium and MDO, in particular, will monitor them. Two metadata standards that might be of particular interest in the near future are Dublin Core (DC) and DataCite.[1] The CMM has been checked against DC and DataCite, and the relevant elements in both standards were available in the CMM. MDO will keep track of the work of SSHOC and WP3 (which is working on a service that could help researchers find and use existing online tools for data analysis) in particular. Another initiative that does work on metadata is the Research Data Alliance (RDA)[2]. Two groups within the RDA that are of particular interest for CESSDA's metadata work are the "Research Metadata Schemas WG"[3] and the "Metadata IG"[4]. The RDA Metadata Interest Group addresses in general all aspects for metadata on research data. "In particular, it will attempt to coordinate the efforts of the WGs concerned with metadata to produce a coherent approach to metadata covering metadata modalities of description, restriction, navigation, provenance, preservation and the use of metadata for the purposes discovery, contextualisation, validation, analytical processing, simulation, visualisation and interoperation. It will also liaise with the other WGs especially Data Foundation and Terminology, PIDs, Standardisation of data categories and codes and Data Citation. This IG activity relates to data management policies and plans of research organisations and researchers, and to policies and standards of research funders and of research communities which may or may not be official standards." With schema.org, it is possible to utilize robust and widely known commercial search engines (e.g., Google, Bing, etc.) to aid the discovery of and access to scientific data. However, concerning its application for research data, schema.org has deficiencies, as its development has mainly been based on commercial business use cases. Therefore, the RDA Research Metadata Schemas Working Group has two objectives; namely "1. to identify and bridge gaps in existing schemas commonly used for research data, by bringing together communities who are working with such vocabularies to document research data and related resources; 2. to provide guidelines for those communities whose needs are not addressed by existing metadata schema such as schema.org, and provide guidelines on proposing extensions."[5] MDO has representatives in both the RDA Metadata IG and the RDA Research Metadata Schemas WG and is keeping constant track of the work done in them. This monitoring of the RDA metadata should and will be continued in the future. However, CESSDA might benefit from more active participation in the work of RDA. This could be considered for future strategic evaluations.

1 Broeder et al. 2019
2 RDA n.d.c
3 RDA n.d.b
4 RDA n.d.a
5 RDA n.d.a

Once MDO deems this necessary, based on the proceedings and developments in the groups, it will thus inform the consortium.

This report addresses different audiences with varying goals for each audience. The report's content is intended for the CESSDA SPs, aspiring CESSDA members, CESSDA MO, and anyone interested in the initiatives discussed in this report in general, and in the metadata work of CESSDA in particular. With this report, MDO intends to provide a document for CESSDA SPs and aspiring CESSDA SPs that supplies those target groups with the necessary knowledge on the initiatives relevant for CESSDA and the work within CESSDA. It has become apparent that the levels of knowledge on the reflected initiatives differ within the CESSDA community. Therefore, this report shall inform, and allow the whole CESSDA community to come to an equal level of understanding on those initiatives. For those initiatives well-known to the CESSDA community (i.e., DDI Codebook (DDI-C) and DDI Lifecycle (DDI-L), the FAIR principles (on metadata) and ESOC and SSHOC), the "General Description" part is kept brief. More general details on the initiatives can be found in the appendix. The report will, furthermore, provide information for the public and display information on what is essential for CESSDA concerning the metadata work and how CESSDA interacts and reflects relevant (metadata) initiatives. For the CESSDA SPs, the report is supposed to provide some guidance for their general considerations on the quality and possible upgrades of their metadata. In addition, the report is also intended for aspiring CESSDA SPs - to provide some reference on metadata usage within CESSDA: What is the status quo of the metadata in the CESSDA products (as the CDC and EQB), and which level of metadata quality is required in order to have a member's metadata represented in those CESSDA products. Additionally, this report is intended to provide some guidance for strategic decisions on CESSDA's metadata developments and a knowledge base that allows easy and grounded decision-making (on, e.g., possible extensions of metadata and upgrading metadata schemas to a more elaborate DDI specification or version). The report is also supposed to point out some of the metadata needs of the CESSDA SPs.

The report's structure contains seven sections. The following five sections introduce the following four initiatives: DDI, FAIR, EOSC together with SSHOC, and DVN. Since the new specification of DDI (DDI-CDI) is rather different to DDI DDI-C and DDI DDI-L, and therefore deserves special consideration and explanation, the part on DDI is split into two separate sections. The first section is about the DDI-C and DDI-L specifications (Section 2), which are already in use within CESSDA. The second section on DDI addresses the new specification DDI-CDI (Section 3). Section 4 covers the FAIR principles, followed by section 5 that introduces EOSC and SSHOC. Section 6 discusses the Dataverse Network and covered metadata within the application. All sections on the initiatives/standards contain a general

description and the implications for CESSDA, with recommendations (if applicable). The last and final section (Section 7) is the conclusion of this report.

2) DDI Codebook (DDI-C) and DDI Lifecycle (DDI-L)

a) General Description DDI-C and DDI-L

The DDI's metadata standard is a free and international standard to describe data from the social, behavioral, economic (SBE) and health sciences. The scope of the disciplines covered by DDI has recently been extended with the introduction of the DDI-CDI specification. The DDI metadata standard still is and has been for many years the most used and most suitable metadata standard within the social science community. This is because it allows for a detailed, semantically rich documentation of research data. Furthermore, DDI metadata are machine-actionable and promote interoperability, metadata exchange, and reuse of metadata.

Since the beginnings of DDI, the DDI Alliance has produced different specifications – DDI-C, DDI-L, and DDI-CDI – with different versions of the metadata standard.

The standard has been extended and improved with each new specification and version, from the beginnings of DDI-C to the newest DDI specification, DDI-CDI. Since DDI-CDI has only recently been made available for public review (in April 2020) and deserves a closer look for CESSDA purposes, a separate section in this report is dedicated to this specification. The present section focuses on DDI-C and DDI-L.

The first specification published was DDI-C. DDI-C focuses on after the fact documentation and is less extensive than DDI-L and DDI-CDI. DDI-C covers information on document description, study description, variable description, and file description. DDI-L adds complexity and offers more than DDI-C. The DDI-L specification bases documentation on the data lifecycle.[6]

b) DDI-C and DDI-L: Implications for CESSDA

As stated in the "CESSDA Strategy 2018-2022," CESSDA and its member countries "seek to increase the scientific excellence and efficacy of European research in the social sciences, as well as to expand easy access to data and metadata regardless of borders."[7] This mission requires detailed and semantically rich metadata that are machine-actionable and interoperable. Therefore, the DDI metadata standard has always played an essential role in CESSDA. CESSDA supports the work of DDI and is a full member of the DDI Alliance.

[6] See Figure 2.
[7] Dekker 2018, p. 5

CESSDA applies the DDI metadata standard for the CMM – which is a reference model for metadata in the CESSDA context that ensures a common understanding of metadata elements and metadata interoperability of CESSDA products and services – and the CESSDA products CDC and EQB. Among the CESSDA SPs, several DDI specifications are in use. Some use versions of DDI-L, and others use versions of DDI-C. There are two use cases within CESSDA ERIC for which DDI plays a significant role: Those are the CDC and the EQB. For CDC and EQB, DDI sets the basis for the backend development and metadata retrieval. As mentioned above, they also use different specifications of DDI due to their different needs. The CDC currently harvests metadata based on DDI2.5 and DDI1.2.2. The backend components, as well as the metadata schema of the CESSDA EQB, are based on DDI3.2.

In discussions about the DDI specifications, the question of which specification is the best and has the biggest advantages is often posed. This is a very difficult question to answer, and answers vary from case to case. It depends on how many resources there are for documentation and metadata, and which kind of metadata information need to be documented. If the documentation needs to be very extensive and detailed and cover the whole data lifecycle, then DDI-L is the better choice. However, it needs to be kept in mind that DDI-L is much more extensive than DDI-C and hence most likely requires more effort to get acquainted with. The number and complexity of metadata fields in DDI-L pose a challenge for CESSDA archives for how to agree on a common interpretation and application. If documentation that can be covered with DDI-C is sufficient and if the resources are tight, then DDI-C might be the better choice. Therefore, there is no definite answer as to which DDI specification is the best choice since this is dependent on many different factors. This is also reflected in the metadata landscape of the CESSDA SPs and CESSDA products, since CESSDA SPs use different DDI specifications and versions of DDI, according to their needs and available resources.

As for the question of "which specification is the best for CESSDA products," this also depends on the use case. For CDC and EQB, this decision has been made considering the needs of the use cases and the capabilities of the SPs. Hence, for CDC it was decided that currently DDI-C is sufficient to meet the needs of the tool, covering 22 metadata elements on study level. Also, most SPs can provide DDI-C. The choice of DDI-C was not adequate for EQB with its advanced functions and metadata levels, covering over 100 metadata elements on study and question-level metadata. For this reason, the EQB metadata schema is based on DDI-L, allowing to document the data lifecycle. However, since EQB SPs use DDI-C, EQB harvests DDI-C as well and converts DDI-C documentation to DDI-L.

Whether in the future, CESSDA products will move to another specification of DDI or a different version of the specification depends on how they will be developed and the resources available for such a move. For instance, if CDC were to include information on

new data types, switching to DDI-CDI, could be considered.[8] As long as the metadata for CDC will not be extended to variable level or new data types, it might be worth considering to switch to DDI2.6, once it is available. It has certain improvements compared to DDI2.5. One would be that in contrast to DDI2.5, DDI2.6 allows for the differentiation between a PID and other identifiers. Since the use of a PID will be mandatory for CESSDA metadata by the end of 2020, it would be useful to have this distinction in the CESSDA metadata files. A move to the newest DDI-L version 3.3, which was published in April 2020, could also be considered. The majority of changes include new features, and DDI3.3 has improved documentation.[9] It is "more" machine-actionable than DDI3.2 due to strongly typed representations, such as for Categories, Ranking for Response domains, Geographic Structure Codes, etc. DDI3.3 also contains a CV element "GeneralDataFormat", which is not included in DDI3.2, but would be useful for the purposes of CESSDA and the extension of the CMM.

The most important improvements compared to DDI3.2 are new features on classification management, non-survey data collection (with MeasurementItem and MeasurementConstruct), more advanced description on sampling, further weighting information and questionnaire design, usage of property graphs (properties on items and references) as well as more advanced quality statement. Improvements in the classification management are defined by the usage of GSIM (Generic Statistical Information Model) defined property types and additional properties for item references where relationships were described with text. With DDI3.3, it is possible to describe sample frame construction for data collection and sample plans, which enriches the metadata for sampling information. Further weighting information includes metadata on the weighting methodology and process, weighting usage guidelines, and standard weights. With the additional metadata on questionnaire design, DDI3.3 includes the possibility to describe the process for creating and testing a questionnaire, document the development plan, and the results from pre-testing activities. The quality statement improvements that are introduced for DDI3.3 are useful for Eurostat reporting by pairing DDI Quality Statements with definitions of Quality Standards. Additional improvements in DDI3.3 include other materials, which can be versioned, reused, and attached to all items, allowing DDI3.3 to be used as a document content management system.[10] Since DDI3.3 is backwards compatible with DDI3.2 and most changes are additions, existing DDI3.2 software would require only a few changes to support DDI3.3.

[8] See also section 3b "DDI-CDI: Implications for CESSDA."
[9] DDI Alliance 2014
[10] For further details on the specification, please see https://ddialliance.org/Specification/DDI-Lifecycle/3.3/. A comparison between DDI3.3 and the previous version DDI3.2 can be found here: https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/805142559/DDI+3.3+Change+Log.

Additionally, DDI3.3 has a formal model, with which a detailed changelog can be generated between DDI3.2 and DDI3.3.[11] However, since DDI2.5 (for CDC) and DDI3.2 (for EQB) cover the CESSDA SPs' metadata documentation in a sufficient way and switching to a different DDI version requires resources, a DDI version change cannot be recommended at this stage. If users of the tool demand more or different metadata and SPs can deliver the information to EQB and CDC, moving to DDI-L for CDC or moving to DDI3.3 for EQB might be worth considering. In any case, one of the core recommendations of MDO in the "Report on Strategic Metadata Requirements for CESSDA"[12] was to continue collaboration with the DDI Alliance and to stay updated on further developments and improvements. This recommendation continues to apply, especially in respect of the new specification DDI-CDI.[13] In the past, CESSDA has taken the chance to influence the development of the DDI specifications and has communicated the consortium's needs via the DDI issue trackers. MDO is in constant contact with the DDI Alliance to ensure good cooperation between DDI and CESSDA. For CESSDA and the CESSDA SPs, it can be recommended to collaborate even more closely and coordinate and pool efforts to direct the developments of the standard to communicate the needs of the whole consortium concerning DDI standard to the DDI Alliance.

3) DDI - Cross Domain Integration (DDI-CDI)

a) General Description DDI-CDI

DDI-CDI is a new specification developed by the DDI Alliance. It was opened for public review in April 2020. DDI-CDI is a useful and practical application of the DDI model, which emerged from many years of work on a "next-generation" DDI specification, known as DDI 4. DDI-CDI is a core of DDI 4.[14] The new specification, DDI-CDI, is aligned with DDI-C and (especially) DDI-L to enable the integration of external systems using DDI. The new specification is not a continuation or replacement of the other DDI specifications, but an "extension to the suite of DDI work products which will help those in the SBE domains and outside of them integrate the expanding range of data required by today's research."[15] DDI-CDI makes it possible to make optimal use of modern systems, by using a variety of models and complying with a range of related specifications.[16] DDI-CDI is geared towards the description of new forms of data (e.g., NoSQL/big data, event data, multi-dimensional

[11] DDI Alliance 2017
[12] Förster et al. 2020b
[13] For further information, see also section 3b "DDI-CDI: Implications for CESSDA."
[14] DDI Alliance 2020g, 2020i
[15] DDI Alliance 2020g, 2020d
[16] DDI Alliance 2020g, 2020d, 2020b

data), which are not adequately covered by the former DDI specifications, DDI-C and DDI-L.[17] DDI-CDI and DDI-L have, however, many features in common.[18] "The intention is that DDI-CDI be a tool which can supplement systems using earlier versions of DDI, enabling them to better handle new types of data."[19] With DDI-CDI it is possible to document a variety of research data in different formats coming from different sources, independent of their scientific and policy domain.[20] To enable this, DDI-CDI provides a flexible means of describing data, which represent connections between identical data existing in different formats. Furthermore, DDI-CDI makes it possible to describe the provenance of data in detail.[21] DDI-CDI can be extended, to add information not covered by other specifications, and to prepare the specification best for the future. Furthermore, it is not only aligned to DDI-C and DDI-L, but is designed to align with other relevant standards and models (including DCAT (Data Catalog Vocabulary), Schema.org, various data-related W3C vocabularies (including DataCube, but also others such as SOSA/SSN (Sensor, Observation, Sample, and Actuator/Semantic Sensor Network Ontology), CSV on the Web, etc.), SDMX, JSON-based standards around data, PROV, and others).[22]

For DDI-CDI, the DDI Alliance went back to a more conceptual model compared to DDI-L. The conceptual model of DDI-CDI is captured using (a limited subset of the) Unified Modeling Language (UML) formalization, whereas DDI-C and DDI-L are captured in XML. Using UML makes it easier to map DDI-CDI to models within systems. Since the focus of DDI-CDI is on UML, XML is less relevant to DDI-CDI than to DDI-L and DDI-C. However, as well as the UML model there is a syntax representation in XML of DDI-CDI, which enables the direct implementation of the model.[23] The UML model of DDI-CDI is provided in the form of Canonical XML Metadata Interchange (XMI), which is an interchange format for UML models supported by many different modeling and development tools. XMI allows for the exchange of UML models between various tools, making it easy to have implementations in many software environments.[24]

The functional goal of DDI-CDI has shifted compared to the other DDI specifications: "where DDI-C was an XML representation of a data dictionary, and DDI-L a more complex model designed to support metadata from data conception and capture through publication and reuse, DDI-CDI is an attempt to describe data and its provenance independent of these

17 DDI Alliance 2020i, 2020d
18 DDI Alliance 2020a
19 DDI Alliance 2020a, p. 5
20 DDI Alliance 2020a
21 DDI Alliance 2020a, 2020d
22 DDI Alliance 2020i, 2020a
23 DDI Alliance 2020a
24 DDI Alliance 2020a

contexts."[25] Hence, DDI-CDI supports the description of processes as the documentation of provenance.[26] Provenance means the processes used for the data's compilation, to prepare them for their use. The need to describe the provenance of data in detail stems from the growing demand for data from different sources and external domains. Traditionally, many social science data were derived from surveys using questionnaires or written protocols. Today various alternate sources are used to create (new types of) data, making it necessary to understand those data. DDI-C and DDI-L describe structures (e.g., a table or a cube) and meanings, where a variable is synonymous with a column. By contrast, DDI-CDI separates the structural description of the meaning and enables the description of structural forms as tables or key-value stores. "Because the way in which such a model can be implemented is more variable than it is for traditional SBE data management systems, the emphasis in DDI-CDI is on a model, formalized in UML."[27] The four significant data structures that DDI-CDI can be used for are wide data, long data, dimensional data, and key-value data.[28]

Another aspect of the new DDI specification is the datum-oriented approach. DDI-CDI uses a datum-oriented approach to support the best description of the provenance of data and the different data structures. With this approach, DDI-CDI recognizes that an individual datum can play different roles and have different functions in different structures (e.g., as a measure, a descriptor, an identifier). "The general idea in DDI-CDI is to be able to attach all necessary metadata to the single datum so that it can be 'followed' across different data structures. This differs from some other approaches used in other DDI products (DDI-C, DDI-L), where some of this information was attached at a higher level (e.g., the data set or record)."[29] Using DDI-CDI, the identity of a datum can be maintained, and the different roles of a datum and its exact transformations may be described in detail. With DDI-C and DDI-L, it is already possible to describe specific processes applied to data (e.g., cleaning, derivations, etc.) as granular descriptions of specific programmatic functions. With DDI-CDI, higher-level processes can be described to provide an account of the provenance of data. DDI-CDI is a significant extension of the DDI model compared to the capabilities of DDI-L and DDI-C.[30] The DDI Alliance describes possible use cases for DDI-CDI as follows: "Envisioned implementations include the addition of new types of data into management systems on the basis of descriptions using the DDI-CDI model, where the existing system may already use DDI-C or DDI-L, and the transformation into DDI-CDI of existing DDI Lifecycle/Codebook metadata for the purpose of integrating data with other data that might

[25] DDI Alliance 2020a, p. 5
[26] DDI Alliance 2020d
[27] DDI Alliance 2020a, p. 5
[28] DDI Alliance 2020a
[29] DDI Alliance 2020d, p. 27
[30] DDI Alliance 2020i, 2020d

only be described using the DDI-CDI model. This last case would apply where non-traditional data sources (big data, sensors, etc.) were being integrated with more traditional ones."[31]

b) DDI-CDI: Implications for CESSDA

DDI-CDI provides powerful and valuable capabilities that add to those of the previous DDI specifications. As the DDI Alliance has described, one envisioned implementation for DDI-CDI is adding DDI-CDI-based metadata on new data types for systems that use DDI-C or DDI-L. Another one is the transformation of existing metadata, captured in DDI-C or DDI-L, into DDI-CDI, to integrate these descriptions with metadata using the DDI-CDI-model.

When asking if or when DDI-CDI needs to be adopted by CESSDA, one must consider the use cases of CESSDA and check what is needed accordingly. Those use cases are currently CDC and EQB. Furthermore, there is a difference between the current needs and requirements of CDC and EQB and what they might need in, e.g., five years, which depends on how they will develop. For the current needs of CDC metadata, DDI-C metadata suffices, and most SPs are capable of providing DDI-C metadata. By contrast, EQB's features require rich metadata and the functions of DDI-L; hence, DDI-C would not be enough for EQB. As for looking ahead in the case of the CDC, there are several possibilities to consider once the current version of DDI-C no longer meets its needs. Those could include, for example, updating to DDI-L, using an updated version (if one exists) of DDI-C, or working with the DDI Alliance to develop an updated version, etc. Due to the internal referencing in DDI-L, tools (like an editor) would be needed to exhaust and make use of the possibilities DDI-L offers within the CESSDA context to the fullest extent. In case it is decided to move the CDC metadata to DDI-L, CESSDA should keep in mind that an editor would be useful for SPs, especially for those with fewer resources. However, it needs to be considered that the creation and maintenance of such tools requires a lot of effort. For EQB, a move to an updated version of DDI could be considered, when future metadata requirements are no longer met by the DDI standard currently used. As for the future, DDI-CDI itself will become important for CESSDA products and services, as soon as the SPs archive data that cannot be adequately documented with DDI-C or DDI-L (such as new data types, or very detailed information on dataset level not provided in DDI-L), and the documentation of those data is to be displayed in CESSDA products (such as the CDC and EQB). This depends on the future strategic decisions of CESSDA. The "CESSDA Strategy 2018-2022" acknowledges new data types, such as registries and social media data as a possible scenario for CESSDA. Therefore, CESSDA is aware of a possible shift concerning metadata standards. One of

[31] DDI Alliance 2020i, p. 6

CESSDA's strategic goals is to support standards on metadata. The consortium is prepared to meet the challenges that might arise with new data types.[32] Once such data need to be documented, using DDI-CDI would be the logical step. Since DDI-CDI can be used in addition to metadata captured in DDI-C or DDI-L, the existing documentation in DDI within the CESSDA context could remain. If additional documentation were needed, DDI-CDI documents could be created that reference the corresponding documentation files in the older DDI formats. This feature makes the DDI-CDI model an extremely attractive option, especially since a complete replacement of the existing documentation with metadata in DDI-CDI would not be necessary in any case. However, the DDI-CDI model is not trivial and requires substantial knowledge and work (even for experienced DDI users and CESSDA tool developers) for its application. This needs to be kept in mind when asking for additional metadata in DDI-CDI. Earlier discussions with the CESSDA SPs suggest that they currently have different levels of maturity as far as metadata are concerned. Moreover, since some SPs struggle to make the switch from NESSTAR to DDI-C, DDI-L, or Dataverse, the more urgent task within the CESSDA context should be to get the SPs to the same level, which would allow an upgrade with additional DDI-CDI metadata in the future.[33] At the moment, there is no up-to-date data available on the specifications or versions of DDI currently used by each SP, nor of their future plans concerning upgrades of their metadata. An online survey with current and aspiring SPs on their metadata practices would be a possibility for the future, to get more information on this and use this information as a basis for further strategic considerations. In any case, a good, constant, and thorough communication between CESSDA products, SPs, Main Office (MO), external initiatives, and MDO is needed to meet the challenges that will arise in the future and to ensure the best possible service that CESSDA can offer.

The fact that the CMM is already quite extensive is another issue that should be noted when considering the extension of CESSDA metadata with information (e.g., on new data types) that needs to be captured in DDI-CDI. The CESSDA SPs have expressed they were facing problems with the scope of the CMM and the many numbers of elements. Of course, the CMM is a very useful reference for good and rich metadata in the Social Sciences, and the SPs are not asked to adopt the complete CMM, but only the parts that are needed for the CDC and EBQ. However, the SPs might be alarmed if CESSDA's metadata schema added a lot of additional and new metadata elements that need to be captured in DDI-CDI, especially if the requirements of CDC and EQB do not reflect this need. Extending the scope of CESSDA's data holdings to new data types will be unavoidable in the future. But as soon as

[32] Dekker 2018

[33] See also conclusion in CESSDA internal report on the MDO webinar; Deliverable 2b, MDO Task1 WP 2020.

CESSDA decides to include elements for new data types in CMM, using the DDI-CDI model, the reasons and implications for this need to be communicated properly to the CESSDA SPs.

4) FAIR Principles (on Metadata)

a) General Description FAIR Principles (on Metadata)

The FAIR principles for scientific data management and stewardship are based on four goals; namely, to make scientific digital data more Findable, Accessible, Interoperable, and Reusable. For each goal, the authors defined more specific guidelines, resulting in a total of 15 principles. The FAIR Guiding principles do not refer to any specific standard or specification but rather act as best practices to data producers and publishers in dealing with data and their associated metadata. They apply to data (or any digital object), metadata (information about that digital object), and infrastructure (FAIR principles). Since the first publication of the FAIR guidelines in 2016, the principles have become widely popular. Numerous organizations, projects, and initiatives concerned with data infrastructures, data sharing, and metadata have adopted the FAIR principles as guidelines.[34] In the following, the principles referring to metadata will be listed and described in more detail.

Table 1: FAIR Principles.[35]

FINDABLE	
The findability of metadata is a pre-condition for (re)using them. Metadata should, therefore, be easy to find, both for humans and computers. Metadata need to be machine-readable to enable automatic discovery of datasets and services.	
F1. (Meta)data are assigned a globally unique and persistent identifier	• Principle F1 is the most important one, as other aspects of FAIR require the use of globally unique and persistent identifiers. • Globally unique and persistent identifiers provide a unique identification (identifier) to every element of metadata. They consist of a persistent internet link (e.g., a URL) and assist computers to search, interpret, or automatically integrate the data. "Identifiers are essential to the human-machine interoperation that is key to the vision of Open Science."[36] Identifiers also support researchers to cite correctly, when reusing data and metadata. • Identifiers must be: o globally unique (i.e., once a set of data or metadata is assigned an identifier, this identifier cannot be reassigned to another dataset). Identifiers can be obtained from registry services "that use algorithms guaranteeing the uniqueness of newly minted identifiers."[37] o persistent. A registry service guarantees the resolvability of the identifier in the future.

[34] Wittenburg et al. 2020
[35] Source: https://www.go-fair.org/fair-principles/. *GO FAIR (CC BY 4.0).*
[36] GO FAIR n.d.b
[37] GO FAIR n.d.b

F2. Data are described with rich metadata (defined by R1 below)	• Data should be findable, without the help of an identifier, but simply by the information provided in their metadata. Compliance with F2 simplifies the location of data and increases reuse and citation of said data. • Metadata should be exhaustive and contain information about "the context, quality and condition, or characteristics of the data."[38] • "Rich metadata allow a computer to automatically accomplish routine and tedious sorting and prioritizing tasks that currently demand a lot of attention from researchers."[39] • Metadata authors should be as generous as possible in providing metadata.
F3. Metadata clearly and explicitly include the identifier of the data they describe	• As the metadata and the dataset usually come in separate files, their relationship should be made clear by including the dataset's globally unique and persistent identifier in the metadata. • Most repositories will automatically assign globally unique and persistent identifiers to their research datasets.
F4. (Meta)data are registered or indexed in a searchable resource	• (Meta)data are searchable and findable on the Internet – they need to be discoverable. • Discoverability of digital resources can be ensured, e.g., by indexing (core elements of indexing are described by F1-F3).
ACCESSIBLE	
After finding the required data, the user needs to know how it can be accessed. This may include authentication and authorization.	
A1. (Meta)data are retrievable by their identifier using a standardized communications protocol	• (Meta)data should be retrievable by their identifier "without specialized or proprietary tools or communication methods"[40] via standardized communications protocols (e.g., http(s) or ftp). • Standardized communications protocols enable the retrieval of (meta)data by simply clicking on a link.
A1.1 The protocol is open, free, and universally implementable	• Access to the metadata should be free (no-cost) and open (-sourced) to anyone with a computer and an internet connection, thus ensuring that they are globally implementable to enable data retrieval.
A1.2 The protocol allows for an authentication and authorization procedure, where necessary	• Providing clear information concerning the conditions for accessibility of data is central in order to allow machines to automatically understand and then execute or inform the user about accessibility requirements. (This does not necessarily imply the data must be "open" or "free." Even private data or protected data can be FAIR). • Repositories often require the creation of a user account to potentially set user-rights permissions for controlled data access.
A2. Metadata are accessible, even when the data are no longer available	• Even when the dataset is no longer available, the metadata should continue to persist. • A2 corresponds to the aspects of registration and indexing specified in F4.
INTEROPERABLE	

[38] GO FAIR n.d.c
[39] GO FAIR n.d.c
[40] GO FAIR n.d.a

Usually, data integration - combining data from different sources – is needed. In addition, the data need to "interoperate with applications or workflows for analysis, storage, and processing."[41]	
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	• Humans, as well as machines, should be able to exchange, read, and interpret (meta)data. This requires interoperability, meaning that computer systems have at least fundamental knowledge of each other's system's data exchange formats "without the need for specialized or ad hoc algorithms, translators, or mappings."[42] In order to make sure that datasets are automatically findable and interoperable, "(1) commonly used controlled vocabularies, ontologies and thesauri (along with resolvable globally unique persistent identifiers, see F1), and (2) a good data model (a well-defined framework to describe and structure (meta)data)"[43] ought to be used.
I2. (Meta)data use vocabularies that follow FAIR principles	• The controlled vocabularies used in the metadata have globally unique and persistent identifiers to document and easily resolve these controlled vocabularies. This documentation should be easy to find and access for (meta)data users.
I3. (Meta)data include qualified references to other (meta)data	• "A qualified reference is a cross-reference that explains its intent."[44] • (Meta)data resources should be referred to one another by as many meaningful links as possible. These links need to be described in a way that allows understanding the relations between data (e.g., "builds upon", "relates to", "is supplement to"). • The datasets need to be properly cited, including their globally unique and persistent identifiers.
REUSABLE	
To optimize the reuse of data, metadata need to be well-described to foster replication or combination of metadata and data.	
R1. (Meta)data are richly described with a plurality of accurate and relevant attributes	• Users (machine or human) should be able to decide easily if data is useful in a particular context. • This implies "plurality," meaning that as much metadata as possible should be included, independent of its relevance to the publisher. Metadata should provide information about the context under which the data was generated in addition to discovery metadata (e.g., experimental protocols, type of data, limitations about the data, software).
R1.1. (Meta)data are released with a clear and accessible data usage license	• In contrast to (*technical*) **Interoperability,** R1.1 refers to *legal* interoperability. It needs to be clear whether and how the data are licensed. The conditions of use need to be explicit for machines and humans. This is especially important when automated searches involve more licensing considerations.
R1.2. (Meta)data are associated with detailed provenance	• To increase (re)usability, the data origin/history,[45] citation, and workflow need to be clear. • The provenance information should ideally be described in a machine-readable format.

[41] GO FAIR n.d.d
[42] GO FAIR n.d.e
[43] GO FAIR n.d.e
[44] GO FAIR n.d.f
[45] See R1.

R1.3. (Meta)data meet domain-relevant community standards	• Existing community standards and best practices for data archiving and data sharing should be followed since standards facilitate reuse of metadata and data. When there are valid reasons to deviate from standards and best practices, these should be clearly stated in the metadata.

The question of the benefits of FAIR and if they are worth the effort for their compliance has been raised several times. This is a valid question that needs to be evaluated. A cost-benefit analysis of the FAIR principles, by PricewaterhouseCoopers EU Services (2018)[46], shows that compliance with the FAIR principles has many advantages for all research stakeholders, such as researchers, archives, funders, and the whole research community. While some investments are necessary to make (meta)data FAIR, those investments are nothing compared to the benefits that come with the application of FAIR and to the costs that arise when not complying with FAIR.

Given the relevance of the FAIR principles (that stems from their manifold benefits of the principles described above) - not just within CESSDA, but for science in general – this report evaluates their implications for CESSDA.

b) FAIR Principles (on metadata): Implications for CESSDA

As explicitly stated in the "CESSDA strategy 2018-2022," CESSDA will comply with the FAIR principles in order to reach the goal of data and metadata access beyond any boundaries.[47] Self-assessment on CESSDA's implementation status of the FAIR policies revealed that "Concepts for FAIR implementation" and "Investment in FAIR" are already put into practice.[48] CESSDA is performing well on "Skills for FAIR," that is providing training and exchange of knowledge (e.g., a mentorship program for new SPs). The realization of a "FAIR culture" within CESSDA has started. CESSDA is working on interoperability frameworks that rely on such a shared culture. There is still some work to be done regarding a "FAIR ecosystem" and especially on "Incentives and metrics for FAIR data and services." To enable interaction between machines, automated processing should be supported. For this to be achieved, "CESSDA needs to optimize its workflows within the 'FAIR ecosystem'."[49] Furthermore, CESSDA should develop and implement metrics for FAIR Digital Objects. The consortium is highly active in developing internal assessment frameworks and certifying FAIR services (e.g., CoreTrustSeal for data repositories).[50] A range of CESSDA SPs already achieved such trust badges. However, all SPs are encouraged to obtain them.[51]

[46] PwC EU Services 2018a, 2018b
[47] Förster et al. 2020b
[48] Self-assessment is based on the FAIR Action Plan recommendations. For further information, see Dekker 2020.
[49] Dekker 2020, p. 225
[50] Dekker 2020
[51] Förster et al. 2020b

The extent to which the CESSDA SPs reference the FAIR principles on their websites, reports, and projects varies greatly. An important step towards open access to research data is to increase the awareness and visibility of FAIR data management. Therefore, efforts to promote awareness should be increased.

CESSDA sequentially implements the FAIR principles, currently accomplishing **F**indability (e.g., implementation of CDC & EQB[52]) while working on the other principles.[53] One of the MDO's core recommendations for metadata strategies focuses on the use of metadata among CESSDA SPs, "especially on the level of 'survey instruments,' 'variables,' and 'question and answers'."[54] [55] Furthermore, the metadata provided should adhere to the FAIR principles.[56] CESSDA focuses on developing uniform criteria for the description of different access levels (Accessibility) and Interoperability frameworks (e.g., EOSC and SSHOC). As Reusability follows from the other principles, a comprehensive understanding of the FAIR principles is needed.[57] CESSDA metadata materials should not only be beneficial for the social sciences but research domains beyond this scope. In today's information society, data and interdisciplinary research are growing in importance in solving complex problems. Hence, data are increasingly being linked, with different data types and content being combined. Therefore, working to achieve the greatest degree of FAIR is a priority.[58]

The following table lists the FAIR principles discussed above that concern metadata, and the extent to which CDC and EQB currently meet the requirements of those principles. It is important to note that the evaluation refers to the current status of the products. Where possible, future objectives are mentioned. However, they were not included in determining if a FAIR principle is fully met. Only what is in place and implemented at present is included in the evaluation.

Table 2: CESSDA products and the FAIR Principles.

FAIR principles	CDC	EQB
FINDABLE		
F1. (Meta)data are assigned a globally unique and persistent identifier	Yes/No. Some SPs assign persistent identifiers (PID) to their metadata records. The responsibility for this lies with the SPs.	Yes/No. Not every metadata element contains a PID. However, Study PID is part of the EQB metadata schema, implemented and shown in the EQB User

[52] See also Table 2.
[53] Dekker 2020
[54] See F2.
[55] Förster et al. 2020b, p. 1
[56] See Table 1.
[57] Dekker 2020
[58] Förster et al. 2020b

		Interface. The responsibility for this lies with the SPs.
F2. Data are described with rich metadata (defined by R1 below)	No. The current CDC metadata schema contains 22 elements on study level. However, the CDC links to the SPs pages, where more detailed information on the data can be found. However, an extension of the CDC metadata schema is planned for the future.	Yes. EQB metadata schema provides about 100 metadata elements. Users profit from a wide range of information on questions, study and basic variable information.
F3. Metadata clearly and explicitly include the identifier of the data they describe	Yes/No. The current version of the CDC metadata schema includes the element "PID Study" which is not yet mandatory (This is also due to the fact that according to the CESSDA Data Access Policy, the use of PIDs for studies becomes mandatory at the end of 2020). However, there is additionally a mandatory element "Study number," which refers to the ID provided by the SP for each study.	Yes/No. The dataset itself is not included in EQB. However, the users can access the dataset and dataset ID on the SP's homepage. The link to further study description and dataset is provided in EQB.
F4. (Meta)data are registered or indexed in a searchable resource	Yes. The metadata held in CDC are browsable and searchable. Indexing in other catalogues is planned for 2021.	Yes. The metadata are browsable and searchable.
ACCESSIBLE		
A1. (Meta)data are retrievable by their identifier using a standardized communications protocol	No. The CDC itself has no machine-actionable means of retrieving the metadata held in CDC. However, the responsibility for the data and metadata holdings lies with the SPs.	Yes/No. The EQB (meta)data are reusable, however questions regarding the Terms of Use for Users and a Service Level Agreement between CESSDA SPs and CESSDA (EQB), which defines under which license the metadata should be provided, needs to be clarified.
A1.1 The protocol is open, free, and universally implementable	Yes/No. CDC metadata can be openly accessed via the internet. Anyone with a computer and an internet connection can access the metadata of CDC. However, CDC has no standardized communications protocol for accessing the metadata. CDC uses OAI-PMH and will be retrievable from CESSDA in 2021. It is already retrievable from data holding SPs.	Yes. EQB is open-access and is retrievable via the internet.
A1.2 The protocol allows for an authentication and authorization procedure, where necessary	N/A[59] – is not necessary for metadata.	Yes. For the option to compare and download metadata, the user must register with an email address, because the metadata on questions will be browsable and

[59] See A1.

		downloadable and users will be allowed to come back to their choices.
A2. Metadata are accessible, even when the data are no longer available	Yes/No. The metadata are accessible, even when the data are no longer available, unless the SP adds deletion notices for one or more study records, or the CDC indices are reset and a full re-harvest takes place.	Yes/No. The metadata are currently not accessible when the data is no longer available. However, EQB keeps a copy of the harvested metadata. It is possible in the future to offer the original and converted DDI metadata as a download link or offer the export either as PDF or DDI.
INTEROPERABLE		
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	Yes. The CDC metadata retrieval is based on the DDI (Codebook) metadata standard. This is converted to CMM v1 within CDC.	Yes. The EQB metadata retrieval is based on the DDI-L metadata standard.
I2. (Meta)data use vocabularies that follow FAIR principles	Yes/No. Some SPs use controlled vocabularies when creating the metadata records (CESSDA topic classification, ELSST (European Language Social Science Thesaurus), DDI AnalysisUnit, DDI TimeMethod, DDI SamplingProcedure, DDI ModeOfCollection, ISO 3166-2) which comply with the FAIR principles.	Yes. The controlled vocabularies are the CESSDA topic classification, DDI TimeMethod, DDI SamplingProcedure, DDI Mode of Collection, which all comply with the FAIR principles.
I3. (Meta)data include qualified references to other (meta)data	No, CDC metadata currently don't include qualified references to other (meta)data. Currently, some SPs include such references in their more detailed documentation on their websites for (at least some) of their data. However, qualified references could be something that might be added in the future.	Yes/No. EQB provides a link to all other questions used in the questionnaire of a study for each of the question items. It also provides a URL on study level linking to the questionnaire and variables (dataset) the question belongs to. Citation information is provided as well.
REUSABLE		
R1. (Meta)data are richly described with a plurality of accurate and relevant attributes	Yes/No. CDC does not include the actual data but provides information on where the data can be accessed. The metadata provided by CDC itself are metadata on study level, captured in 22 elements.	Yes. EQB does not include the actual data however, it provides information on where the data can be accessed. However, it includes metadata on question level, study-level, variables and documentation.
R1.1. (Meta)data are released with a clear and accessible data usage license	Yes/No. The metadata schema of CDC contains an element on data access. This element contains information relating to the access and use conditions of data. However, the element is currently not mandatory.	Yes. The website includes information about the Privacy Policy, the Acceptable Use Policy and Users need to accept the Terms of Use before downloading question information for reuse.

R1.2. (Meta)data are associated with detailed provenance	Yes. CDC provides information about the metadata origin (publisher) and data collection information (time/mode of collection).	Yes. EQB provides information about the metadata origin (publisher), citation information, data collection information (time/method).
R1.3. (Meta)data meet domain-relevant community standards	Yes. The CMM[60], which was used as a basis for the CDC metadata schema is published (in Zenodo) and is closely modelled according to DDI-C.	Yes. The metadata provided to the community is compliant to CMM and DDI.

As Table 2 displays, both EQB and CDC are doing alright in terms of meeting the FAIR principles on metadata. Most of them are mostly met. When it comes to the first principle (Findability), both EQB and CDC,

- require (for CDC starting 2021) persistent and globally unique identifier
- set the basis for rich metadata, for EQB with over 100 metadata fields, or plan to enrich the metadata (CDC)
- allow users to search and browse the metadata

For the other principles (Accessible and Interoperable), good reasons exist, why they are not met (yet) or only partly met. Firstly, both EQB and CDC are only metadata providers for the CESSDA SPs, and data access is the responsibility of the SPs. Secondly, for some of the principles, plans to make the products compliant to the FAIR principles are underway. And lastly, some of the principles are not relevant for the CESSDA products, and they don't need to comply with them.

The last principle, "Reusable" is mostly met by both EQB and CDC:

- Information on the data is available, and/or links are provided.
- Access and usage information for the users are provided.
- Information on the metadata origin can be found.
- The metadata meet domain-relevant standards, which is DDI for the Social Sciences.

The CESSDA CDC is among the repositories that have been evaluated by the EOSC Nordic project on their "FAIRness".[61] "The project's landscaping initiative is intended to measure the state of FAIR uptake among scientific data repositories in the Nordic and Baltic region and to monitor the development of FAIRness during the project period."[62] The score of the CDC in this evaluation is not very high. The FAIR metrics score of CDC could be improved by adding linked data according to schema.org and JSON-LD into catalog records.

[60] Borschewski et al. 2019
[61] Jaunsen et al. 2020
[62] Jaunsen et al. 2020, p. 6

This leads to the conclusion that the CESSDA products are acceptably positioned concerning their adherence to the FAIR principles on metadata. However, there is room for improvement. The future plans of both CDC and EQB should improve the FAIRness of the products, even more so when, e.g., for EQB, enhancing linked data according to Schema.org and JSON-LD as mentioned above.

Adhering to the FAIR principles is significant for the consortium to remain a leading player in the international scientific community. CESSDA is aware of this relevance and has included the FAIR principles explicitly in its strategy[63], its policy considerations (e.g., for the Data Access policy[64], and the CESSDA PID policy[65]), and its general agenda and considerations. Additionally, the consortium evaluates its performance on the FAIR principles providing a realistic assessment of the adherence to the FAIR principles and prospects for further efforts of alignments.[66] It can be concluded that CESSDA is currently doing well concerning the compliance to the FAIR principles, especially when taking into account its contribution to the EOSC.[67] CESSDA and CESSDA SPs should permanently assess, uphold, and improve their compliance with FAIR.

5) European Open Science Cloud (EOSC) and the Social Sciences and Humanities Open Cloud (SSHOC) project

a) General Description EOSC and SSHOC

The EOSC has been proposed and is financially supported by the European Commission. EOSC is part of the European Cloud Initiative and supports the "vision of global Open Science as a driver for enabling a new paradigm of transparent, data-driven science and accelerating innovation."[68]

"The initiative reinforces Open Science, Open Innovation and Open to the world policies. It will

- foster best practices of global data findability and accessibility;
- help researchers get their data skills recognized and rewarded;
- help address issues of access and copyright and data subject privacy;
- allow easier replicability of results and limit data wastage;

[63] Dekker 2018
[64] Woollard et al. 2016
[65] Hausstein and Horton 2020
[66] Dekker 2020
[67] See also section 5 below "European Open Science Cloud (EOSC) and the Social sciences and Humanities Open Cloud (SSHOC) project."
[68] European Open Science Cloud 2020

- contribute to clarification of the funding model for data generation and preservation, reducing rent-seeking and priming the market for innovative research services."[69]

When discussing EOSC, there is one part most relevant for CESSDA and the Social Sciences, namely the SSHOC project. As a thematic cluster project for the Social Sciences and Humanities (SSH), the SSHOC project contributes to the realization of the EOSC.[70] [71] With SSHOC, a "full-fledged Social Sciences and Humanities Open Cloud (SSHOC) where data, tools, and training are available and accessible for users of SSH data"[72] shall be provided. SSHOC sets out to develop and provide EOSC services for the SSH community by interconnecting existing and new infrastructures. The project adheres to the FAIR principles[73] and the principles of Open Science to maximize data reuse.[74] It provides resources, activities and a network for training and prepares an appropriate governance model for SSH-EOSC.[75]

b) EOSC and SSHOC: Implications for CESSDA

Within the "CESSDA Strategy for 2018-2022," EOSC and the intention of being part of EOSC (especially on social science data) are mentioned multiple times, as well as the purpose of aligning with the European Science Agenda.[76] In this strategy, CESSDA's aim "towards a mature high-quality social science infrastructure, well embedded in the EOSC, with CESSDA's SPs performing as trusted repositories, and heavily used by researchers and professionals in their roles of data producers and data users"[77] is declared. As a matter of fact, CESSDA decided to actively take part in shaping the EOSC. To reach this goal, CESSDA has joined the coalition of doers and is a coordinator of the SSHOC project.[78] For the coalition of doers of EOSC, "CESSDA will:

- provide a one-stop shop for search and discovery of datasets relevant to social science research through a joint product and service data catalogue;

---

69 EOSC Secretariat 2019
70 Broeder et al. 2019
71 Social Sciences & Humanities Open Cloud 2020a, 2020b; Ilijašić Veršić and Dekker 2020; Tóth-Czifra 2020; Wittenberg et al. 2020; Ilijašić Veršić and Ausserhofer 2019; Social Sciences & Humanities Open Cloud 2020c
72 CESSDA 2020
73 For more information, see section 4 in this report "FAIR principles (on Metadata)."
74 Broeder et al. 2019
75 Social Sciences & Humanities Open Cloud 2020a, 2020b; Ilijašić Veršić and Dekker 2020; Tóth-Czifra 2020; Wittenberg et al. 2020; Ilijašić Veršić and Ausserhofer 2019; Social Sciences & Humanities Open Cloud 2020c
76 Dekker 2018; CESSDA 2017
77 Dekker 2018, p. 3
78 CESSDA 2017, 2018; Tóth-Czifra 2020

- develop and provide tools for submitting and using data in a safe and secure way;
- train data owners and data users, including supporting training activities across European Research Infrastructure Consortia (ERIC);
- contribute to the implementation of FAIR principles – especially with regard to sensitive and confidential data;
- ensure trusted data repositories – by ensuring the acquirement of seals for data as well as data service organisations (such as the Data Seal of Approval/CoreTrustSeal);
- be a hub for its SPs and other professional organisations to share expertise and best practices."[79]

The integration of CESSDA assets in EOSC is, without a doubt, essential for CESSDA. Being part of EOSC and participating in the process of its creation has considerable and valuable advantages for CESSDA, such as:

- expanding its reach,
- determining and following best practices,
- fostering open science and innovation,
- following the FAIR principles,
- boosting efficient research,
- facilitating reuse of data,
- helping Europe to attain a leading role in scientific data infrastructures,
- being able to shape EOSC, and
- making existing infrastructures more powerful by interconnecting them and adding new infrastructures.

Since the EOSC principles are well aligned with the strategic ambitions of CESSDA, EOSC is considered in CESSDA's strategy, and CESSDA is a coordinating partner in SSHOC, the consortium is very well positioned in this aspect. Since CESSDA is doing so well in this area, there is not much further to recommend. In general, it can be advised that CESSDA keeps an eye out for such far-reaching mega-projects in the future, monitor and actively participate in them. In terms of metadata, CESSDA needs to evaluate the results of the work of SSHOC WP3. They have done an analysis of the interoperability of the metadata landscape within the SSH. This analysis concluded that metadata practices within SSH are very heterogenous and that the organizations included in the analysis apply a wide variety of metadata standards. However, diversity is needed to meet the needs for metadata of the different research domains, and no single standard would be able to support all kinds of use cases. It is recommended by WP3 that the different domains should use specific metadata standards, but that for the need of building conversion services, DC and DataCite are to be

[79] CESSDA 2017

used by all domains.[80] The CMM has been checked against DC and DataCite, and the relevant elements in both standards were available in the CMM as well. MDO will monitor the developments of SSHOC and the work WP3 in particular and communicate on these issues with CESSDA MO, when relevant.

One initiative whose work is relevant for CESSDA in the EOSC aspect is FAIRsFAIR. The role of FAIRsFAIR in the EOSC ecosystem is "particularly to foster the creation and interconnection of metadata catalogues in order to facilitate and incentivise sharing and finding interdisciplinary data for a common scientific performance among disciplines."[81] In July 2020, the Scientific Research and Innovation Agenda (SRIA) for EOSC "stressed the need for a multi-stakeholder European partnership to enhance the circulation of research data and knowledge in digital form across borders and disciplines, and to allow scientists and machines to collaborate in creating, storing, processing, finding, accessing and reusing scientific data."[82] FAIRsFAIR's efforts are directed at the realization of an integrated approach that allows EOSC to deal with cross-disciplinary, global collaborative research, which facilitates sharing and reusing (meta)data across all scientific disciplines. In the FAIRsFAIR deliverable "D3.6 Proposal on integration of metadata catalogues to support cross-disciplinary FAIR uptake", the initiative focused on metadata interoperability at syntactic and an element set level. To foster interdisciplinary research with a common discoverability framework, FAIRsFAIR proposes to use an agreed upon high-level metadata element set that supports data integration of different research data types and research domains. For its evaluations concerning the Social Sciences and Humanities metadata and (meta)data catalogs, FAIRsFAIR built on the experiences and the work of SSHOC WP3. Contrary to the SSHOC WP3 work, FAIRsFAIR also included schema.org and DDI-CDI in their evaluations due to the growing interest in them.[83]

FAIRsFAIR will carry out a pilot (end of 2020 until mid-2021) to support the discoverability of (meta)data catalogs by aggregators. The integration service used for this pilot will be the discovery portal B2FIND that covers all domains and harvests various metadata schemas and standards which are mapped to the B2FIND schema. The B2FIND schema is similar to the DataCite metadata schema, but a bit more elaborate. In order to prepare the CESSDA metadata assets for harvesting by B2FIND, MDO will work on a mapping between the CDC and B2FIND in 2021. The pilot of FAIRsFAIR will furthermore trial the use of DCAT/DCAT-AP

[80] Broeder et al. 2019
[81] Mendez et al. 2020, p. 7
[82] Mendez et al. 2020, p. 9
[83] Mendez et al. 2020

and DDI-CDI for metadata catalogs. Those metadata standards were also recommended in the FAIRsFAIR report on FAIR requirements for persistence and interoperability.[84] CESSDA and MDO are keeping an eye on and partly keep involved in the development of DDI-CDI.[85] Depending on the outcomes of FAIRsFAIR and the resulting changes in EOSC, CESSDA might need to map or enhance their metadata with DDI-CDI. Also, depending on the further evaluations of FAIRsFAIR and their outcomes, CESSDA might have to take a look into DCAT and consider its application for CESSDA metadata.

6) Dataverse

a) General Description Dataverse

The Dataverse Network (DVN) software development started in 2006 at the Institute for Quantitative Social Science (IQSS) at Harvard University.[86] The DVN is an open-source application for publishing, referencing, extracting, and analyzing research data and offers data sharing solutions.[87] The software is structured in a way that a Dataverse installation consists of several sections, so called "Dataverses." Future public datasets are published in a Dataverse, which contains other Dataverses and in which each dataset contains descriptive metadata and data files[88].[89]

The DVN project offers a central repository infrastructure, thus requires no local storage, hardware, software, and IT staff on the part of the data provider.[90] The software maintenance of the DVN application is coordinated at IQSS, which supports data archives to acquire, manage, preserve, and distribute a collection of data sets along with the associated documentation and metadata. It facilitates data storage, archiving, cataloging, and preservation formatting, as well as backups, recovery, remote repository caching, virtual collections of remote objects, persistent identifiers, and citations.[91]

b) Implications for CESSDA

The advantages of Dataverse for CESSDA SPs can be summarized as follows:

[84] Riungu-Kalliosaari et al. 2020

[85] Mendez et al. 2020

[86] Crosas 2011
[87] King 2007: http://thedata.org
[88] E.g., https://dataverse.nl/.
[89] Crosas 2011
[90] King 2007
[91] Crosas 2011; King 2007

1) The possibility to apply own Dataverses: To install the DVN software is free of charge and accessible for institutions and archives who want to share, preserve, and cite data. For those CESSDA SPs who do not have their own data infrastructure, the DVN application is an easy and affordable opportunity to build such an infrastructure according to their requirements. Since 2019, CESSDA has been working with DARIAH (Digital Research Infrastructure for the Arts and Humanities), CLARIN (Common Language Resources and Technology Infrastructure), and ERIHS (European Research Infrastructure for Heritage Science) in a DataverseSSHOC project.[92] The goal is to build a reliable and production-ready Open Source data infrastructure based on sustainable microservices that organizations can install and reuse. The project aims to provide a data repository service running on EOSC by February 2022 and a report on the principles of governance and sustainability of the service by April 2022.[93]
2) The possibility to enrich the metadata and improve the metadata processing: The DVN application is an opportunity for CESSDA SPs to enrich their metadata. The SSHOC Dataverse project aims to establish a NESSTAR DDI migration tool (completion planned for June 2021 in CESSDA Cloud): a DDI converter that will automatically migrate study metadata to Dataverse metadata.[94] Especially for CESSDA SPs who are currently using NESSTAR, this will be an opportunity to improve the quality and quantity of the stored and provided metadata. The advantages of Dataverse compared to NESSTAR are that Dataverse:
    - is suitable for all kinds of data, not only survey data
    - has a persistent identifier (study level, and possible for file level)
    - can keep track of versions of the dataset
    - does not face slowness of server like NESSTAR
    - is still being developed and offers support compared to NESSTAR
    - has a variety of user permissions[95]
3) The possibility to allow the integration with other data schemas such as DDI, OpenAire, DataCite, Dublin Core, and Schema.org: The inclusion of different metadata standards ensures that DVN metadata can be mapped easily to standard metadata schemas and be exported into JSON format (XML for tabular file metadata) for preservation and interoperability. The metadata covered in the metadata schema of the DVN application that are relevant for CESSDA products and SPs are:

[92] Social Sciences & Humanities Open Cloud 2019; Wittenberg et al. 2020
[93] See also section 5 "European Open Science Cloud (EOSC) and the Social sciences and Humanities Open Cloud (SSHOC) project)."
[94] More information: https://github.com/IQSS/dataverse-ddi-converter-tool/.
[95] Wittenberg et al. 2020

- Citation Metadata[96]: compliant with DDI Lite[97], DDI2.5 (Codebook)[98], DataCite3.1[99], and Dublin Core's DCMI Metadata Terms[100]. Language field uses ISO 639-1[101] controlled vocabulary.
- Geospatial Metadata[102]: compliant with DDI Lite, DDI2.5 (Codebook), DataCite, and Dublin Core. Country / Nation field uses ISO 3166-1[103] controlled vocabulary.
- SSH Metadata[104]: compliant with DDI Lite, DDI2.5 (Codebook), and Dublin Core.[105]

All in all, the DVN application and the usage of Dataverses can be considered a good alternative and opportunity for archives who want to improve their metadata interoperability and consider moving from NESSTAR to the Dataverse software. However, the question remains: Is Dataverse also a good alternative for CESSDA SPs? To answer this question, it is necessary to have a closer look at the metadata covered in the DVN metadata schema. The user guides page[106] shows the types of metadata the Dataverse software comes with and how it is organized. The DataverseEU project prepared a mapping of CMM and Dataverse metadata and has extended the metadata with the CMM metadata schema. In preparation of this report, it was decided to prepare our own mapping, taking into account the Citation section, Geospatial section, and Social Sciences section.[107] The MDO mapping has shown the following results: All mandatory, upper elements on study level that are set as "mandatory" within CMM are covered within the DVN metadata schema. Elements that could not be found in the existing schema are marked in red in the table in the appendix. However, the metadata schema of the DVN covers information on titles, funding, distribution, contributors, classifications, keywords, methodology, access, citation, publications, documents, and study series. These fields are provided with only a few metadata elements, which can, however, be sufficient if an alternative for NESSTAR is needed and (meta)data documentation on a basic level is necessary. For the purposes of CDC and currently 22 metadata fields in CDC, the usage of the DVN software can be worth considering. However, for EQB, further communication with IQSS and the DVN community is needed. A first evaluation showed that

[96] Google Docs n.d.b
[97] DDI Alliance 2020c
[98] DDI Alliance 2020e
[99] DataCite Metadata Working Group 2015
[100] DCMI 2020
[101] Library of Congress 2013
[102] Google Docs n.d.a
[103] Wikipedia 2020
[104] Google Docs n.d.c
[105] Dataverse 2020
[106] At http://guides.Dataverse.org/en/latest/user/appendix.html#metadata-references.
[107] See Table 3 in Appendix for the mapping between Dataverse and CMM.

the DVN metadata schema supports only metadata on study level (only in DDI-C, not in DDI-L), even though the API is able to export variable metadata (here again only in DDI-C, not in DDI-L). Question level metadata such as Question Item Text, Question Item ID, and Response Categories/Codes, which are crucial for EQB, can currently not be documented within the DVN application or Dataverses.

Even though the DVN software is currently not very suitable for publishing survey data in a very granular way and question-level metadata, this does not mean that it will be unusable for CESSDA SPs and products in the future. There is an ongoing DataverseEU project that will offer a data repository service for data archives with limited technical resources. This service can be used in various ways. GESIS, along with DANS (Data Archiving and Networked Services), ADP (Social Science Data Archives) and SND (Swedish National Data Service), will adopt and extend the metadata within the DVN. Further adaptations like multilingualism of the user interface, harvesting of the CESSDA Data Catalogue, and integration with the CESSDA CVS will be made.[108] Additionally, MDO got in touch with the Dataverse community and IQSS and received the information that they are open to include missing metadata and that it would take some design and development work. There are already some discussions in the community[109] and developments that started to address the requirements to add questions to the metadata fields, e.g., the Data Curation Tool, that generate additional metadata fields mentions the "qstn" element from DDI2.5 on the web.[110] With the special requirements of the EQB and CESSDA to include question metadata, further cooperation with IQSS, together with SPs using NESSTAR and other projects (e.g., DataverseEU), should be continued. MDO will uphold the communication with the Dataverse community on the needs of the CESSDA community.

7) Conclusion

In this "Report on other initiatives/projects on metadata outside of CESSDA," MDO has described four different metadata initiatives that play a significant role for CESSDA in general and, more specifically, for its work, services, products, and strategic considerations. Those initiatives are the DDI standard (with DDI-C, DDI-L, and the newest specification of the standard – DDI-CDI), the FAIR guidelines, EOSC and the SSHOC project, and the Dataverse Network project. Each of those initiatives has been evaluated according to its implications for CESSDA. The report addresses CESSDA SPs, aspiring CESSDA members,

[108] GESIS - Leibniz-Institute for the Social Sciences n.d.
[109] See https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/70652970/Use+Cases for further information.
[110] Gautier 2020

CESSDA MO, and anyone interested in the topics in general and in the metadata work of CESSDA in particular. It is supposed to provide information and guidance for strategic decisions. The main results of the observations covered in this report are summarized in the following.

Section 2 discussed the DDI-C and DDI-L specifications of the DDI standard. The DDI standard is a free, international metadata standard for the social behavioral, economic, and health sciences (at least this is true for DDI-C and DDI-L, since DDI-CDI has significantly widened the scope of DDI to make it domain-independent). DDI-C is the more basic specification, which provides metadata elements to document information that could be found in a traditional codebook. DDI-L is the more elaborate specification that bases documentation on the DDI data lifecycle and emphasizes the reuse of metadata.

Section 3 discusses the new DDI-specification DDI-CDI, which is currently opened for public review. DDI-CDI is aligned with DDI-C and (especially) DDI-L. It is not intended to replace DDI-C or DDI-L, but more to be an add-on that can be used together with DDI-C or DDI-L. DDI-CDI is geared towards the description of new forms of data (e.g., NoSQL/big data, event data, multi-dimensional data) and allows for the documentation of a variety of research data, independent of the research domain, in different formats, coming from different sources, with regard to provenance.

CESSDA declares in its strategy (for 2018-2022) that it aims to improve the scientific excellence and efficacy for the social sciences in Europe and to break down barriers to data imposed by geographical borders or metadata access. This objective cannot be achieved without semantically rich and high-quality metadata, which are machine-actionable and interoperable. This is why CESSDA has adopted DDI for its purposes. Within the CESSDA community, the SPs use different specifications – and for those different specifications various versions for – their metadata that match their needs and available resources. CESSDA products are also using different DDI specifications, according to their requirements and needs. The metadata of the CDC are currently based on DDI2.5, while the metadata schemas of EQB and CMM are both based on DDI3.2. For the metadata currently harvested by CDC, DDI2.5 suffices. The big advantage of using DDI-C for CDC metadata is that most CESSDA SPs can deliver DDI-C. This is not the case for DDI-L at the moment. Using DDI-C for EQB, however, was not an option, since EQB needs rich metadata on studies and questions. Which DDI specifications or versions of the specification will be used in the future for the products and CMM depends on how the needs of CESSDA will change, on how the products will be developed, and which further metadata shall be included in the different CESSDA products. If, for example, users of EQB require more metadata on weighting and sampling, and the SPs can deliver metadata in DDI3.3 (the latest version of the DDI-L specification), upgrading EQB and the CMM might be worth considering. Another possibility

would be that the CDC shall display metadata elements, e.g., on question and variable level. In that case the metadata of CDC might need to be upgraded to DDI-L. If the development goes even further and metadata on new data types need to be displayed by CESSDA products, the extension of the metadata and (additional) use of DDI-CDI need to be considered. However, when upgrading metadata to new specifications or versions, it must always be kept in mind that such an upgrade requires time and resources from CESSDA as well as the CESSDA SPs. In order to provide metadata in a new form, the CESSDA SPs need to be able to make such a switch. At the moment, some SPs still struggle to break away from NESSTAR. It is possible that those archives might not have the means to move to one of the elaborate specifications of DDI. In general, it can be recommended that CESSDA continues to use DDI for its metadata, stays up to date with the developments of DDI, and continues its collaboration with the DDI Alliance in the future. Furthermore, it can be advised that CESSDA SPs should collaborate even more closely to coordinate and pool all efforts to direct the development of the standard to communicate the whole consortium's needs concerning DDI. As for the CESSDA community, good and constant exchange and communication between CESSDA MO, products, SPs, and MDO will help meet the challenges CESSDA might face in the future, regarding metadata.

Section 4 is about the FAIR principles. The FAIR principles represent best practices for data producers, publishers, and archives to make scientific digital data, metadata, and infrastructures more Findable, Accessible, Interoperable, and Reusable. As introduced in section 4 ("FAIR principles (on Metadata)"), following FAIR principles requires certain resources. However, the costs for not following FAIR are many times higher than complying with FAIR. In general, not following FAIR has a negative impact on the quality of research, the economy, and the readability of research data. The FAIR principles help to reduce duplications in research, which minimizes the time, efforts, and funds needed for studies. Furthermore, the FAIR principles help researchers meet the requirements of their funding agencies. The principles foster machine-readability in standardized formats of (meta)data and scale up research findings. All the positive effects of FAIR make it possible to focus the research on the activities that add value to the actual research, such as analysis or interpretation, since it reduces the time needed for searching, collecting of (meta)data, etc. Furthermore, going FAIR has a positive influence on scientific infrastructure, enhancing knowledge discovery and innovation. Even though FAIR data doesn't need to be open access, FAIR fosters the openness of data. For all the reasons above, the benefit of applying the FAIR principles exceeds the costs of FAIR. This fact pertains to archives, researchers, funders, and the whole research community.

CESSDA acknowledges the importance and the advantages of FAIR and is strongly committed to following the FAIR principles. Considerations on FAIR are included in the

"CESSDA strategy 2018-2022" as well as in CESSDA's policies (e.g., the CESSDA Data Access Policy and the CESSDA PID Policy). The application of the FAIR principles is relevant for any scientific institution that wants to be a player on the global scale. Furthermore, the EOSC is also committed to the FAIR principles. Since CESSDA is and wants to remain a global player in the social science context and wants to be part of EOSC, the FAIR principles are an important basis for the work of the consortium. For this report, MDO has evaluated the performance of EQB and CDC concerning the FAIR principles. The result of this evaluation is that both products are mostly alright positioned when it comes to their compliance with the principles, especially when considering the plans for the future of the products. However, there is still space left for improvement. Also, in an evaluation by the EOSC Nordic project on the FAIRness of the CDC, the catalog attained one of the lower scores. Therefore, it is recommended to continually reflect upon and improve the FAIRness of data and metadata materials within CESSDA. CESSDA has evaluated the general performance of the consortium concerning the adherence to the FAIR principle. This self-evaluation proves how committed CESSDA is to FAIR. It showed that CESSDA is doing very well in realizing Findability and successively extending its compliance to the other principles. The consortium should continue to follow this path in the future. In its evaluation, MDO found that even though many institutions and initiatives are aware of the need for data to be FAIR, some overlook the need for the metadata to be FAIR. It is recommended to apply the FAIR principles not only to data but also to metadata.

Section 5 concerns the EOSC and the SSHOC project. EOSC is part of the Open Cloud Initiative and shall foster open science and open innovation, remove their policy and technical barriers by bringing together existing scientific data infrastructures that are currently spread across countries and disciplines and are not at all or not sufficiently connected. SSHOC covers the part of EOSC for the SSH. The result of the SSHOC project will be a mature Social Sciences and Humanities Open Cloud, where users of SSH data can find data, products, and training. CESSDA has been aware of EOSC and is involved in its development as well as the SSHOC project from the beginning. EOSC is also mentioned in the CESSDA strategy, as are the FAIR principles. Since it is crucial for CESSDA to have its data assets represented in EOSC, the active involvement in EOSC and SSHOC has been the right decision. CESSDA's strategy is well aligned to the EOSC principles, and CESSDA is the coordinator of the SSHOC project. The consortium is, therefore, very well positioned as related to EOSC. A general advice is that CESSDA shall keep an open eye for far-reaching projects as EOSC in the future and repeat what has been done concerning EOSC; namely, monitor those projects and actively participate. Furthermore, in particular, CESSDA and MDO need to monitor the work and results of the SSHOC WP3. Their evaluation of the metadata landscape in the SSH and the interoperability within SSHOC has resulted in the

recommendation that for building conversion services, DC and DataCite are to be used by all domains. CESSDA needs to include this recommendation in its strategic decisions.

Section 6 covers Dataverse and the relevance of Dataverse for CESSDA SPs and products. Dataverse offers the possibility of creating so-called dataverses and facilitates interoperability with CESSDA products due to its coverage with various metadata standards. Especially with regard to studies, citation, and geography metadata, Dataverse is very well positioned and offers a good alternative to NESSTAR, whereby the switch from NESSTAR to Dataverse also depends on the success of current projects (e.g., SSHOC Dataverse) and the participation of MDO in further developments within Dataverse. Furthermore, there are still some requirements and challenges that must first be overcome in order to be useful for CESSDA's metadata requirements. Particularly at the question level, some development and the use of resources are still required to expand the Dataverse metadata and make it usable for CESSDA EQB. Discussion and developments have already started and MDO will take part in giving advice to Dataverse on the metadata requirements that are needed to ensure that the developments that are being made in this area also meet the requirements of CMM and help CESSDA SPs who might consider using Dataverse to be able to integrate their metadata into CESSDA CDC and/or EQB. MDO recommends keeping track of the developments of Dataverse since Dataverse has the potential to become relevant for CESSDA products (and is already used by some SPs). Furthermore, CESSDA should continue to cooperate with the Dataverse Network project, and MDO will uphold the communication with DVN on the CESSDA community's needs.

This report focused on the four currently most relevant metadata initiatives for CESSDA (the DDI standard, the FAIR guidelines, EOSC and SSHOC, and Dataverse). Other metadata initiatives exist (see also appendix) for the SSH that currently are not as relevant for CESSDA as those evaluated in detail, but that might gain more relevance in the future (for example, the RDA with the "Research Metadata Schemas WG" and the "Metadata IG"). Therefore, MDO recommends keeping an eye on those initiatives and continuously evaluating their meaning for CESSDA. As for the RDA in particular, CESSDA might profit from a more active engagement in the initiative. This is something that could be evaluated for future strategic decisions.

In general, a good, constant, and thorough communication between CESSDA products, SPs, CESSDA MO, external initiatives, and MDO on any metadata issues (e.g., possible improvements, possible developments, possible new initiatives, metadata needs of SPs, etc.) is recommended, to meet the metadata challenges that will arise in the future and to ensure the best possible service that CESSDA can offer.

CESSDA's efforts to be an international and leading player for the European social science community have borne fruits. The consortium complies with the relevant international standards and best practices on metadata. Furthermore, it contributes to (pan)European projects, and keeps updated on and plays a part in the relevant scientific communities. The path CESSDA is on holds promises for the future relevance of CESSDA and will certainly further serve the social science research community.

To sum up the results of this report, the following list provides the most important recommendations that have resulted from MDO's evaluation.

Recommendations concerning DDI:

- CESSDA: In case information on new data types shall be included in CESSDA products, the extension of the metadata using DDI-CDI as an addon would be a good solution.
- CESSDA MO: Keep the needs and the available resources of CESSDA SPs in mind for strategic decisions concerning metadata (for example, an extension of CMM, a switch to a different standard, or the upgrade of CESSDA's metadata materials to another DDI specification or a version of a specification).
- CESSDA MO: Provide sufficient resources (including time) for MDO, in case of a switch to a different standard, or the upgrade of CESSDA's metadata materials to another DDI specification or version of a specification.
- CESSDA: Continue close collaboration with DDI Alliance and stay updated on further developments and improvements of the standard.
- CESSDA: SPs should collaborate even more closely and coordinate and pool all efforts to direct the development of the DDI standard to communicate the whole consortium's needs concerning DDI.

Recommendations concerning FAIR

- ALL: Data and metadata need to adhere to FAIR principles. For CESSDA, it is especially important to remain a leading player in the international scientific community. CESSDA and CESSDA SPs should permanently assess, uphold, and improve their compliance with FAIR.
- ALL: Apply FAIR principles not only to data but also to metadata.
- CESSDA: Improve FAIRness of CESSDA products (e.g., by adding linked data according to schema.org and JSON-LD into catalog records), both for data and metadata.

Recommendations concerning EOSC/SSHOC

- MDO/CESSDA: Monitor the work and results of SSHOC (especially) WP3, in case they have consequences for CESSDA's metadata work/schemas. Close contact between MDO and SSHOC WP3 is currently constituted by the fact that MDO partners also work on SSHOC WP3.
- MDO/CESSDA: Continue to monitor the metadata work of RDA (especially of the "Research Metadata Schemas WG" and the "Metadata IG").
- CESSDA MO: Consider a more active role in the work of the RDA in future strategic evaluations. E.g., by having more representatives in RDA groups and by extending the CESSDA participation and contribution at RDA conferences.

Recommendations concerning the Dataverse Network (DVN):

- CESSDA: Keep track of the DVN development, since DVN has the potential to become relevant for CESSDA products (and is already used by some SPs).
- CESSDA: Continue further cooperation with the Dataverse community and the DVN project (IQSS).
- MDO: Uphold communication with DVN on the needs of the CESSDA community.

General recommendations:

- CESSDA: Keep an eye out for any far-reaching mega-projects (as, for example, EOSC) in the future, monitor, and actively participate in them.
- CESSDA/MDO: A systematic survey among CESSDA SPs asking which level of quality their metadata currently has, which standard (and which version of it) SPs currently use, what the SPs plans in terms of metadata developments are, etc., would be a possibility to receive a clear image of the current status of metadata quality and metadata potentials among the SPs and would provide clear and up-to-date footing for further strategic considerations on metadata development.
- MDO/CESSDA: Monitor metadata initiatives currently not relevant for CESSDA, as they might become relevant in the future.
- MDO/CESSDA: Good, constant, and thorough communication between CESSDA products, SPs, CESSDA MO, MDO, and also external initiatives on any metadata issues (e.g., possible improvements, possible developments, possible new initiatives, metadata needs of SPs, etc.), to meet the metadata challenges that will arise in the future and to ensure the best possible service that CESSDA can offer.

Publication bibliography

Blank, Grant; Rasmussen, Karsten Boye (2016): The Data Documentation Initiative. The Value and Significance of a Worldwide Standard. In *Social Science Computer Review* 22 (3), pp. 307–318. DOI: 10.1177/0894439304263144.

Blomberg, Niklas; Petzold, Andreas (2020): ESFRI thematic cluster view on EOSC. Zenodo. Available online at http://doi.org/10.5281/zenodo.3631247, checked on 9/3/2020.

Borschewski, Kerrin; Kerrin; Förster, André; Friedrich, Tanja; Zenk-Möltgen, Wolfgang; Miranda, Patrícia et al. (2019): CMM CESSDA Metadata Model (Version 1.0). Zenodo. Available online at http://dx.doi.org/10.5281/zenodo.3543756, checked on 9/3/2020.

Broeder, Daan; Trippel, Thorsten; Degl'Innocenti, Emiliano; Giacomi, Roberta; Sanesi, Maurizio; Kleemola, Mari et al. (2019): SSHOC D3.1 Report on SSHOC (meta)data interoperability problems (Version v1.0).

Campus Research Computing Consortium (2020): About. What is CaRCC? Available online at https://carcc.org/about/, checked on 8/27/2020.

CESSDA (2017): CESSDA ERIC commits to European open science by joining the 'coalition of the doers'. Available online at https://www.cessda.eu/News-Events/News/CESSDA/CESSDA-ERIC-commits-to-European-open-science-by-joining-the-coalition-of-the-doers, checked on 9/3/2020.

CESSDA (2018): Our new website focuses on users and tools & services. Available online at https://www.cessda.eu/News-Events/News/CESSDA/Our-new-website-focuses-on-users-and-tools-services, checked on 9/3/2020.

CESSDA (2020): Social Sciences & Humanities Open Cloud (SSHOC). Available online at https://www.cessda.eu/About/Projects/Current-projects/SSHOC, checked on 9/3/2020.

CORDIS (2020): Social Sciences & Humanities Open Cloud. Available online at https://cordis.europa.eu/project/id/823782, checked on 9/3/2020.

Crosas, Mercè (2011): The Dataverse Network®: An Open-Source Application for Sharing, Discovering and Preserving Data. D-Lib Magazine (1/2). Available online at http://www.dlib.org/dlib/january11/crosas/01crosas.html, checked on 8/27/2020.

DataCite Metadata Working Group (2015): DataCite Metadata Schema for the Publication and Citation of Research Data. Available online at http://schema.datacite.org/meta/kernel-3.1/doc/DataCite-MetadataKernel_v3.1.pdf, checked on 9/24/2020.

Dataverse (n.d.): About. The Project. Available online at https://dataverse.org/about, checked on 8/27/2020.

Dataverse (2020): Appendix. Additional documentation complementary to the User Guide. Available online at http://guides.dataverse.org/en/latest/user/appendix.html#metadata-references, checked on 9/1/2020.

DCMI (2020): DCMI Metadata Terms. Available online at https://www.dublincore.org/specifications/dublin-core/dcmi-terms/, checked on 9/24/2020.

DDI Alliance (2012): Updates to DDI-Codebook. DDI Version 2.5, pp. 1–4. Available online at https://ddialliance.org/Specification/DDI-Codebook/2.5/updates_to_ddi-2.pdf, checked on 9/3/2020.

DDI Alliance (2014): DDI Lifecycle 3.2 (2014-02-05) [HTML corrected 2014-05-15]. Available online at https://ddialliance.org/Specification/DDI-Lifecycle/3.2/XMLSchema/FieldLevelDocumentation/, checked on 9/1/2020.

DDI Alliance (2017): Implementation Guide. Best Practices for Usage of DDI 3.2 and Future Versions. Increasing Interoperability and Ease of Adoption. Available online at https://ddialliance.github.io/ddimodel-web/static/DDI-Lifecycle-Best-Practices.pdf, checked on 9/1/2020.

DDI Alliance (2020a): DDI - Cross Domain Integration: Introduction. Available online at https://ddi-alliance.bitbucket.io/DDI-CDI/DDI-CDI_Public_Review_1.zip, checked on 7/20/2020.

DDI Alliance (2020b): DDI Cross Domain Integration: Architecture and Alignment with Other Standards. Available online at https://ddi-alliance.bitbucket.io/DDI-CDI/DDI-CDI_Public_Review_1.zip, checked on 7/20/2020.

DDI Alliance (2020c): DDI Lite (Recommended Elements). Available online at http://www.ddialliance.org/specification/ddi2.1/lite/index.html, checked on 9/24/2020.

DDI Alliance (2020d): DDI-Cross Domain Integration: Detailed Model. Available online at https://ddi-alliance.bitbucket.io/DDI-CDI/DDI-CDI_Public_Review_1.zip, checked on 7/20/2020.

DDI Alliance (2020e): Document, Discover and Interoperate. Available online at https://ddialliance.org/, checked on 8/27/2020.

DDI Alliance (2020f): Membership List. Available online at https://ddialliance.org/ddi-membership, checked on 9/24/2020.

DDI Alliance (2020g): Public Review: DDI - Cross Domain Integration (DDI-CDI). Available online at https://ddialliance.org/announcement/public-review-ddi-cross-domain-integration-ddi-cdi, checked on 7/20/2020.

DDI Alliance (2020h): Wall of Adopters. Available online at https://ddialliance.org/ddi-adopters, checked on 9/24/2020.

DDI Alliance (2020i): MRT Report: DDI-CDI - Cross Domain Integration (DDI-CDI) Features and Status.

Dekker, Ron (2018): CESSDA Strategy 2018-2022. CESSDA, checked on 9/1/2020.

Dekker, Ron (2020): Social Data: CESSDA Best Practices. In *Data Intelligence* 2, pp. 220–229. DOI: 10.1162/dint_a_00044.

Digital Curation Centre (2020a): CERIF - Common European Research Information Format. Available online at https://www.dcc.ac.uk/resources/metadata-standards/cerif-common-european-research-information-format, checked on 8/27/2020.

Digital Curation Centre (2020b): Data Package. Available online at https://www.dcc.ac.uk/resources/metadata-standards/data-package, checked on 8/27/2020.

Digital Curation Centre (2020c): DataCite Metadata Schema. Available online at https://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema, checked on 8/27/2020.

Digital Curation Centre (2020d): List of Metadata Standards. Available online at https://www.dcc.ac.uk/guidance/standards/metadata/list?page=1, checked on 8/27/2020.

Digital Curation Centre (2020e): MIDAS-Heritage. Available online at https://www.dcc.ac.uk/resources/metadata-standards/midas-heritage, checked on 8/27/2020.

Digital Curation Centre (2020f): PREMIS. Available online at https://www.dcc.ac.uk/resources/metadata-standards/premis, checked on 8/27/2020.

Digital Curation Centre (2020g): QuDEx - Qualitative Data Exchange Format. Available online at

https://www.dcc.ac.uk/resources/metadata-standards/qudex-qualitative-data-exchange-format, checked on 8/27/2020.

Digital Curation Centre (2020h): SDMX - Statistical Data and Metadata Exchange. Available online at https://www.dcc.ac.uk/resources/metadata-standards/sdmx-statistical-data-and-metadata-exchange, checked on 8/27/2020.

EDDI (n.d.): EDDI 2020 – Virtual Conference, 1-2 December. Available online at https://www.eddi-conferences.eu/, checked on 9/24/2020.

EOSC Secretariat (2019): European Open Science Cloud: What Is It About? Available online at https://www.eoscsecretariat.eu/news-opinion/european-open-science-cloud-what-it-about, checked on 9/3/2020.

European Data Portal (n.d.): DCAT. Available online at https://www.europeandataportal.eu/data/datasets/dcat?locale=en, checked on 8/27/2020.

European Open Science Cloud (2020): EOSC. Available online at https://www.eosc-portal.eu/about/eosc, checked on 9/3/2020.

Förster, André; Borschewski, Kerrin; Bolton, Sharon; Jääskeläinen, Taina (2020a): The matter of meta in research data management: Introducing the CESSDA Metadata Office Project. In *IASSIST Quarterly* 44 (3), pp. 1–17. DOI: 10.29173/iq970.

Förster, André; Borschewski, Kerrin; Zenk-Möltgen, Wolfgang (2020b): Report on Strategic Metadata Requirements for CESSDA, 6/26/2020. Available online at https://public.3.basecamp.com/p/aLLLuG1jaXkiETH8Vuv9GuMS, checked on 8/5/2020.

Gautier, Julian (2020): Adjust how new variable metadata is added to DDI XML exports so that the exports are valid against DDI schema. GitHub. Available online at https://github.com/IQSS/dataverse/issues/6554, checked on 9/3/2020.

GESIS - Leibniz-Institute for the Social Sciences (n.d.): CESSDA - The Consortium of European Social Science Data Archives. Available online at https://www.gesis.org/en/institute/gesis-in-erics/cessda, checked on 9/1/2020.

GO FAIR (n.d.a): A1: (Meta)data are retrievable by their identifier using a standardised communication protocol. Available online at https://www.go-fair.org/fair-principles/metadata-retrievable-identifier-standardised-communication-protocol/, checked on 9/3/2020.

GO FAIR (n.d.b): F1: (Meta) data are assigned globally unique and persistent identifiers. Available online at https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/, checked on 9/3/2020.

GO FAIR (n.d.c): F2: Data are described with rich metadata. Available online at https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/, checked on 9/3/2020.

GO FAIR (n.d.d): FAIR Principles. Available online at https://www.go-fair.org/fair-principles/, checked on 9/3/2020.

GO FAIR (n.d.e): I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. Available online at https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/, checked on 9/3/2020.

GO FAIR (n.d.f): I3: (Meta)data include qualified references to other (meta)data. Available online at https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/, checked on 9/3/2020.

Google Docs (n.d.a): Dataverse Metadata v4.x. Available online at https://docs.google.com/spreadsheet/ccc?key=0AjeLxEN77UZodHFEWGpoa19ia3pldEFyVFR0aFVGa0E#gid=4, checked on 9/24/2020.

Google Docs (n.d.b): Dataverse Metadata v4.x. Available online at https://docs.google.com/spreadsheets/d/13HP-jI_cwLDHBetn9UKTREPJ_F4iHdAvhjmlvmYdSSw/edit#gid=0, checked on 9/24/2020.

Google Docs (n.d.c): Dataverse Metadata v4.x. Available online at https://docs.google.com/spreadsheet/ccc?key=0AjeLxEN77UZodHFEWGpoa19ia3pldEFyVFR0aFVGa0E#gid=1, checked on 9/24/2020.

Green, Ann E.; Humphrey, Chuck (2014): Building the DDI. In *IASSIST Quarterly* 37 (1-4), pp. 36–44. DOI: 10.29173/iq500.

Hausstein, Brigitte; Horton, Laurence (2020): CESSDA ERIC Persistent Identifier Policy 2019 (Version 2.0). Zenodo. Available online at http://doi.org/10.5281/zenodo.3611327.

Hoyle, Larry; Castillo, Fortunato; Clark, Benjamin; Kashyap, Neeraj Kumar; Perpich, Denise; Wackerow, Joachim; Wenzig, Knut (2011): Metadata for the Longitudinal Data Life Cycle. The Role and Benefit of Metadata Management and Reuse. In *DDI Working Papers Series* (3). DOI: 10.3886/DDILongitudinal03.

Ilijašić Veršić, Ivana; Ausserhofer, Julian (2019): Social sciences, humanities and their interoperability with the European Open Science Cloud: What is SSHOC? In *Mitteilungen der VÖB* 72 (2), pp. 383–391. DOI: 10.31263/voebm.v72i2.3216.

Ilijašić Veršić, Ivana; Dekker, Ron (2020): SSHOC - Position paper 2020 (Version v1.0). Zenodo. Available online at http://doi.org/10.5281/zenodo.3697121, checked on 9/3/2020.

Jaunsen, Andreas; Kleemola, Mari; Alaterä, Tuomas J.; Lehvaslaiho, Heikki; Hasan, Adil; Nordling, Josefine; Assinen, Pauli (2020): D4.1 An assessment of FAIR-uptake among regional digital repositories (Version 1.0).

Jensen, Uwe; Zenk-Möltgen, Wolfgang; Wasner, Catharina (2019): Metadatenstandards im Kontext sozialwissenschaftlicher Daten. In Uwe Jensen, Sebastian Netscher, Katrin Weller (Eds.): Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten. Opladen: Verlag Barbara Budrich, pp. 151–178.

Jünger, Stefan; Borschewski, Kerrin; Zenk-Möltgen, Wolfgang (2019): Documenting Georeferenced Social Science Survey Data: Limits of Metadata Standards and Possible Solutions. In *Journal of Map & Geography Libraries* 15 (1), pp. 68–95. DOI: 10.1080/15420353.2019.1659903.

King, Gary (2007): An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. In *Sociological Methods & Research* 36 (2), pp. 173–199. DOI: 10.1177/0049124107306660.

Library of Congress (2013): Codes for the Representation of Names of Languages. Alpha-3 codes arranged alphabetically by the English name of language. Available online at https://www.loc.gov/standards/iso639-2/php/English_list.php, checked on 9/24/2020.

Mendez, Eva; Hernandez, Tony; Whyte, Angus; Davidson, Joy (2020): D3.6 Proposal on integration of metadata catalogues to support cross-disciplinary FAIR uptake. Zenodo. Available online at https://doi.org/10.5281/zenodo.4134787, checked on 12/10/2020.

Metadata Encoding & Transmission Standard (2019): Home. Available online at https://www.loc.gov/standards/mets/, checked on 8/27/2020.

Mills, Hayley (2017): CLOSER Discovery. CLOSER. UCL Wiki. Available online at https://wiki.ucl.ac.uk/display/CLOS/CLOSER+Discovery, checked on 9/3/2020.

NADDI (n.d.): The Premier Data Documentation Conference. North American Data Documentation Initiative Conference. Available online at http://naddiconf.org/, checked on 9/24/2020.

Open Geospatial Consortium (2020): Observations and Measurements. Available online at https://www.ogc.org/standards/om?utm_source=phplist394&utm_medium=email&utm_content=HTML&utm_campaign=OGC+seeks+public+comment+on+SensorThings+API+standard+for+Internet+of+Things, checked on 8/27/2020.

PwC EU Services (2018a): Cost of not having FAIR research data. Cost-benefit analysis for FAIR research data. European Commission. Luxembourg.

PwC EU Services (2018b): Policy Recommendations. Cost-benefit analysis for FAIR research data. European Commission. Luxembourg.

Rasmussen, Karsten Boye (2014): Social Science Metadata and the Foundations of the DDI. In *IQ* 37 (1-4), pp. 28–35. DOI: 10.29173/iq499.

RDA (n.d.a): Metadata IG. Available online at https://www.rd-alliance.org/groups/metadata-ig.html, checked on 9/29/2020.

RDA (n.d.b): Research Metadata Schemas WG. Available online at https://www.rd-alliance.org/groups/research-metadata-schemas-wg, checked on 9/29/2020.

RDA (n.d.c): The Value of RDA for. Available online at https://www.rd-alliance.org/, checked on 9/29/2020.

RDA | Metadata Directory (2015): Encoded Archival Description (EAD). Available online at https://rd-alliance.github.io/metadata-directory/standards/EAD.html, checked on 8/27/2020.

RDA | Metadata Directory (2018a): MARC (Machine-Readable Cataloging). Available online at http://rd-alliance.github.io/metadata-directory/standards/marc.html, checked on 8/27/2020.

RDA | Metadata Directory (2018b): MODS (Metadata Object Description Schema). Available online at https://rd-alliance.github.io/metadata-directory/standards/mods.html, checked on 8/27/2020.

researchobject.org (2020): OAI-ORE (Object Reuse and Exchange). Available online at https://www.researchobject.org/initiative/oai-ore/, checked on 8/27/2020.

Riungu-Kalliosaari, Leah; Hooft, Rob; Kuijpers, Sylvia; Parland-von Essen, Jessica; Tana, Jonas (2020): D2.4 2nd Report on FAIR requirements for persistence and interoperability (Version v1.0 draft). Zenodo. Available online at https://doi.org/10.5281/zenodo.4001630, checked on 12/10/2020.

Simons, Gary; Bird, Steven (Eds.) (2006): OLAC Metadata. Available online at http://www.language-archives.org/OLAC/metadata-20060405.html, checked on 8/27/2020.

Smith, Dan (2019): Implementing DDI3.3 in Colectica Version 6. colectica, 2019. Available online at https://zenodo.org/record/3228361#.XwWeHedCSMo, checked on 7/31/2020.

Social Sciences & Humanities Open Cloud (2019): Developing the SSHOC Dataverse. Available online at https://www.sshopencloud.eu/news/developing-sshoc-dataverse, checked on 9/1/2020.

Social Sciences & Humanities Open Cloud (2020a): About SSHOC. Available online at https://www.sshopencloud.eu/about-sshoc, checked on 9/3/2020.

Social Sciences & Humanities Open Cloud (2020b): SSHOC & EOSC. Available online at https://sshopencloud.eu/sshoc-eosc, checked on 8/12/2020.

Social Sciences & Humanities Open Cloud (2020c): SSHOC Principal Goals. Available online at https://sshopencloud.eu/, checked on 9/3/2020.

Text Encoding Initiative (n.d.): TEI: Text Encoding Initiative. Available online at https://tei-c.org/, checked on 8/27/2020.

Thomas, Wendy; Gregory, Arofan; Gager, J.; Johnson, Jon (2014): Data Documentation Initiative (DDI) Technical Specification. Part II: User Guide - Version 3.2, 2/15/2014, pp. 1–25.

Thomas, Wendy; Gregory, Arofan; Gager, J.; Kuo, I-Lin; Wackerow, Achim; Nelson, Chris (2008): Data Documentation Initiative (DDI) Technical Specification. Part I: Overview - Version 3.0, 4/28/2008, pp. 1–162.

Thompson, Mark; Burger, Kees; Kaliyaperumal, Rajaram; Roos, Marco; da Silva Santos, Luiz Olavo Bonino (2020): Making FAIR Easy with FAIR Tools: From Creolization to Convergence. In *Data Intelligence* 2, pp. 87–95. DOI: 10.1162/dint_a_00031.

Tóth-Czifra, Erzsébet (2020): A sneak peek into the SSH component of EOSC: interview with the SSHOC Marketplace team. DARIAH. Available online at https://dariahopen.hypotheses.org/834, checked on 9/3/2020.

UNESCO (2020): Vocabulary information. UNESCO Thesaurus. Available online at http://vocabularies.unesco.org/browser/thesaurus/en/, checked on 8/27/2020.

van Reisen, Mirjam; Stokmans, Mia; Basajja, Mariam; Ong'ayo, Antony Otieno; Kirkpatrick, Christine; Mons, Barend (2020): Towards the Tipping Point for FAIR Implementation. In *Data Intelligence* 2, pp. 264–275. DOI: 10.1162/dint_a_00049.

Vardigan, Mary; Heus, Pascal; Thomas, Wendy (2008): Data Documentation Intitiative: Toward a Standard for the Social Sciences. In *The International Journal of Digital Curation* 3 (1), pp. 107–113. DOI: 10.2218/ijdc.v3i1.45.

Wikipedia (2020): ISO 3166-1. Available online at http://en.wikipedia.org/wiki/ISO_3166-1, checked on 9/24/2020.

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, I. Jsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. In *Scientific Data* 3, pp. 1–9. DOI: 10.1038/sdata.2016.18.

Wittenberg, Marion; Tykhonov, Vyacheslav; Huis in 't Veld, Laura (2020): SSHOC Webinar: Discussion meeting about requirements of CESSDA Service Providers for a Dataverse repository. Zenodo. Available online at http://doi.org/10.5281/zenodo.3714988, checked on 9/1/2020.

Wittenburg, Peter; Lautenschlager, Michael; Thiemann, Hannes; Baldauf, Carsten; Trilsbeek, Paul (2020): FAIR Practices in Europe. In *Data Intelligence* 2, pp. 257–263. DOI: 10.1162/dint_a_00048.

Woollard, Matthew; Beedham, Hilary; L'Hours, Hervé (2016): CESSDA Data Access Policy, June 2016, pp. 1–9.

World Wide Web Consortium (2013): PROV-Overview. An Overview of the PROV Family of Documents; W3C Working Group Note 30 April 2013. Edited by Paul Groth, Luc Moreau. Available online at https://www.w3.org/TR/prov-overview/, checked on 8/27/2020.

World Wide Web Consortium (2014): The RDF Data Cube Vocabulary. W3C Recommendation 16 January 2014. With assistance of Jeni Tennison. Edited by Richard Cyganiak, Dave Reynolds. Available online at https://www.w3.org/TR/vocab-data-cube/, checked on 8/27/2020.

Zenk-Möltgen, Wolfgang (2012): Metadaten und die Data Documentation Initiative (DDI). In Reinhard Altenhöner, Claudia Oellers (Eds.): Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen. 1. Aufl. Berlin: Scivero Verl., pp. 111–126, checked on 9/3/2020.

Appendix

1) DDI Codebook (DDI-C) and DDI Lifecycle (DDI-L) – Further information on the General Description

The work of DDI started in 1995, when several international data archives and producers met to produce a metadata standard for the social science research community, "[…] because good documentation is paramount to effective data use."[111] The DDI's metadata standard is a free and international standard to describe data from the social, behavioral, economic (SBE) and health sciences.[112]

Since the beginnings of DDI, the DDI Alliance has produced different specifications – DDI-C, DDI-L, and DDI-CDI – with different versions of the metadata standard.

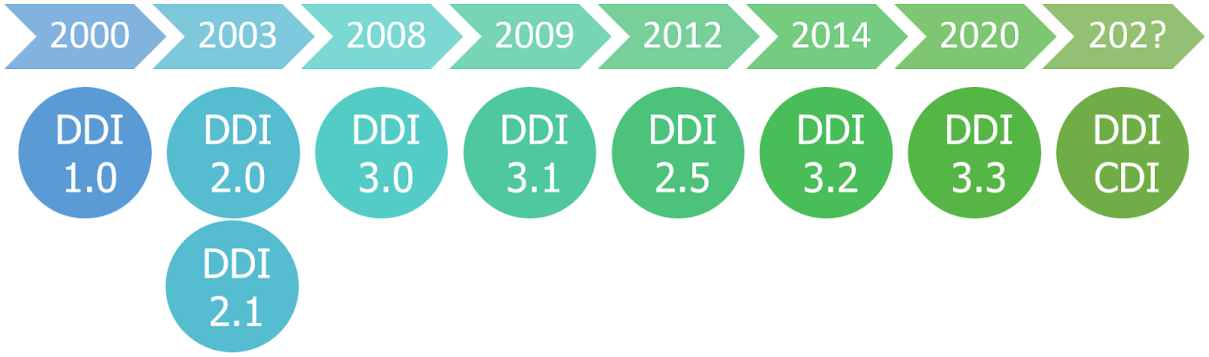

Figure 1: Timeline of DDI specification.[113]

The standard has been extended and improved with each new specification and version, from the beginnings of DDI-C to the newest DDI specification, DDI-CDI. Since DDI-CDI has only recently been made available for public review (in April 2020) and deserves a closer look for CESSDA purposes, a separate section in this report is dedicated to this specification. The present section focuses on DDI-C and DDI-L.

The development of the DDI standard allowed social science archives to move from paper catalogs to standardized and interoperable digital catalogs. The first specification published was DDI-C. DDI-C focuses on after the fact documentation and is less extensive than DDI-L and DDI-CDI. DDI-C covers information on document description, study description, variable description, and file description. DDI-L adds complexity and offers more than DDI-C. The DDI-L specification bases documentation on the data lifecycle.[114] It emphasizes the reuse of

[111] Vardigan et al. 2008, p. 108
[112] Vardigan et al. 2008; Förster et al. 2020a; Blank and Rasmussen 2016; Jünger et al. 2019; Rasmussen 2014; Jensen et al. 2019
[113] Source: Own representation based on Smith 2019; Jensen et al. 2019; DDI Alliance 2012; Zenk-Möltgen 2012; Thomas et al. 2014.
[114] See Figure 2.

metadata, starting from creating concepts to the collection, processing, archiving, distributing, discovering, analyzing, and repurposing research data and corresponding metadata.[115]

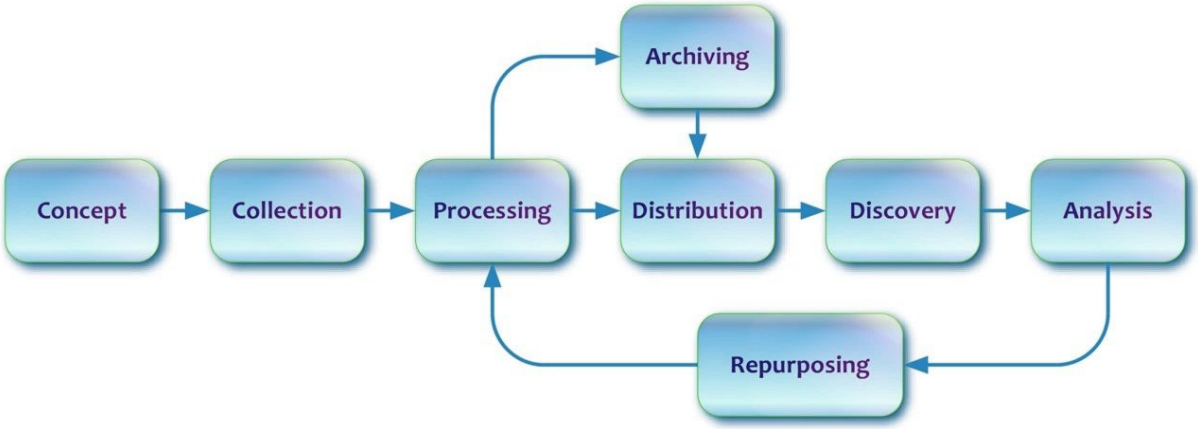

Figure 2: DDI Data Lifecycle.[116]

By focusing data documentation on the DDI Lifecycle, the scope of the DDI standard has increased. Whereas the DDI-C specification focused on data archiving, DDI-L made it possible to describe all activities concerning data throughout the data lifecycle, from the design of the study to the reuse of data. DDI-L "[…] places a major emphasis on being able to identify, version, and maintain the metadata throughout that process"[117] to enable the backtracking of versions of metadata. The documentation of the metadata from the different phases in the lifecycle is flexible and does not have to take place in chronological order. For DDI-L, the DDI Alliance developed modules that logically refer to the data lifecycle phases or technical aspects and summarize the corresponding metadata elements. With DDI-L, the DDI instances became larger since the focus of describing a single data file (primarily of cross-sectional studies) had moved to a more flexible description of data from studies (including longitudinal studies) with several data files. This also means that DDI-L enabled the recording of relationships between studies and the description of study groups.[118]

Another significant change introduced in DDI-L resulted from advances in XML (eXtensible Markup Language) technology. For DDI 3.0, the DDI Alliance moved from DDI Document Type Definition (DTD) as the canonical expression of the standard to the World Wide Web Consortium (W3C) XML Schema Definition (XSD). Another typical XML practice introduced with DDI-L was the use of XML namespaces, which allowed the expanded "vocabulary to be modularized, making it more manageable and maintainable over the long run." [119] Every

[115] Vardigan et al. 2008; Jünger et al. 2019; Förster et al. 2020a; Hoyle et al. 2011; Green and Humphrey 2014; Rasmussen 2014; Thomas et al. 2008; Jensen et al. 2019
[116] Source: http://www.ddialliance.org/training/why-use-ddi. © 2020 DDI Alliance (CC BY 4.0).
[117] Thomas et al. 2008, p. 9
[118] Vardigan et al. 2008; Thomas et al. 2008; Jünger et al. 2019; Förster et al. 2020a; Hoyle et al. 2011; Green and Humphrey 2014; Jensen et al. 2019
[119] Thomas et al. 2008, p. 8

module in DDI-L received a namespace to be used in the XML files.[120] As the timeline shows, in 2012, after two versions of DDI-L had been published, DDI-C with version DDI2.5 "[…] was updated in response to requests for new DDI elements and attributes and to facilitate the transformation of metadata from DDI-C to DDI-L and back."[121] While the DDI metadata standard is a standard of its own, the DDI Alliance has always emphasized DDI's alignment with other relevant standards (for instance, SDMX (Statistical Data and Metadata eXchange), Dublin Core, and ISO codes). It continues to do so and even adds new alignments to external standards with every new specification.[122]

The DDI Alliance currently consists of 30 full members[123] and 15 associated members from many countries worldwide. The number of DDI adopters[124] continually grows. Currently, there are 38 different adopters listed on the DDI Alliance's website. Among those adopters are CESSDA and the Dataverse Network (DVN).

The importance of the DDI standard is also reflected in the conferences on DDI. Every year two DDI-specific conferences are held, the North American Data Documentation Initiative Conference (NADDI)[125] and the European DDI User Conference (EDDI)[126]. NADDI and EDDI are international conferences for those using metadata standards (especially DDI) or interested in learning more about DDI. EDDI has taken place annually since 2009. The first NADDI took place in 2013 and was modeled on the already successful EDDI conference. NADDI is also held every year (except for 2020, when it had to be canceled due to the repercussions COVID-19 pandemic).

2) FAIR Principles (on Metadata) – Further information on the General Description

The FAIR principles for scientific data management and stewardship are based on four goals; namely, to make scientific digital data more **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable. For each goal, the authors defined more specific guidelines, resulting in a total of 15 principles. The FAIR Guiding principles do not refer to any specific standard or specification but rather act as best practices to data producers and publishers in dealing with data and their associated metadata.[127] They apply to data (or any digital object), metadata (information about that digital object), and infrastructure (FAIR Principles).

Since the first publication of the FAIR guidelines in 2016, the principles have become widely popular. Numerous organizations, projects, and initiatives concerned with data

[120] Thomas et al. 2008
[121] DDI Alliance 2012, p. 1
[122] Thomas et al. 2008; Vardigan et al. 2008; Jensen et al. 2019
[123] DDI Alliance 2020f
[124] DDI Alliance 2020h
[125] NADDI n.d.
[126] EDDI n.d.
[127] Wilkinson et al. 2016; Thompson et al. 2020

infrastructures, data sharing, and metadata have adopted the FAIR principles as guidelines.[128] The FAIR guiding principles have been embraced mainly by European countries (67%) and the American geographies (14%) (predominantly the United States of America)[129]. The Southern hemisphere is almost completely excluded. Despite efforts to integrate the principles across all research domains, their implementation is highly biased towards the bio- and natural sciences (95%). In contrast, the applications of the principles in the social sciences is extremely limited (5%). In order to reach a tipping point for the FAIR implementation, the inclusion of non-European geographies and scientific domains other than the bio- and natural sciences is necessary.[130]

The question of the benefits of FAIR and if they are worth the effort for their compliance has been raised several times. This is a valid question that needs to be evaluated. A cost-benefit-analysis of the FAIR principles, by PricewaterhouseCoopers EU Services (2018)[131], shows that compliance with the FAIR principles has many advantages for all research stakeholders, such as researchers, archives, funders, and the whole research community. While some investments are necessary to make (meta)data FAIR, those investments are nothing compared to the benefits that come with the application of FAIR and to the costs that arise when not complying with FAIR. PricewaterhouseCoopers EU Services considered three areas in which the FAIR principles impact: "research activities," "opportunities for further research," and "innovation." They identified and evaluated seven indicators for those three areas of impact.

The impact on research activities concerns the indicators "time spent," "cost of storage," "license costs." For the impact on opportunities for further research, the indicators "research retraction," "double funding," and "cross-fertilization" were considered. And "potential economic growth (as % of GDP)" refers to the impact on innovation. For the impact on research activities, PricewaterhouseCoopers found that FAIR has positive effects on the time, storage costs, and license costs needed to discover, access, and alter (meta)data. This is because FAIR promotes the findability (e.g., by requiring the use of global persistent identifiers and rich quality documentation) of data. Rich, structured metadata also help humans and machines (machine-readability) comprehend data and decide faster if (meta)data are relevant for one's research, and if their quality is sufficient for reuse. FAIR advocates single point access for (meta)data, which would reduce the time needed for searching for (meta)data. FAIR also has a positive impact on the openness of (meta)data and scientific publications, and therefore access fees, which consequently increases the citation ratio of (meta)data and publications. FAIR impacts the time needed to create, curate, clean, process, and maintain (meta)data. If data are FAIR, it is easier to integrate data from different sources. It is also less time-consuming to verify other studies' findings if

[128] Wittenburg et al. 2020
[129] FAIR implication is indicated by the number of articles citing the original FAIR principles paper Wilkinson et al. 2016.
[130] van Reisen et al. 2020
[131] PwC EU Services 2018a, 2018b

their (meta)data are fair. The costs for storage of (meta)data can be reduced following FAIR. If (meta)data and publications are accessible, redundant separate copies held by universities, research institutions, funders, and journals are no longer necessary, which means that an ample amount of storage costs could be saved. Furthermore, applying the FAIR principles helps to reduce the time spent on peer review. The use of international standards for metadata minimizes the time needed to find data in diverse access points. Also, less time is required for authentication, and the interoperability of (meta)data is improved. Another advantage of FAIR is that it reduces the costs for licenses when data are made available openly (and FAIR). FAIR improves the collaboration in research, which indirectly fosters cross-fertilization and innovation, since FAIR fosters (meta)data reuse and citation.

PricewaterhouseCoopers EU Services claim that the implementation of the FAIR principles increases research quality and decreases fraud and plagiarism, and thus minimizes the need for article retraction. Moreover, since FAIR improves the findability, accessibility, and interoperability of data, redundant research can be avoided more efficiently, and double research funding diminished. FAIR facilitates interdisciplinary research and cross-fertilization. Not having FAIR research (meta)data would hamper innovation due to a lack of access to high quality (meta)data. Applying FAIR principles also makes it easier to identify possible partners and experts for research and collaboration. Furthermore, as PricewaterhouseCoopers EU Services point out, building on previous research results enables innovation and progress. Finally, the FAIR principles foster efficiency and have a positive effect on potential economic growth.[132] "By unlocking the value of research and facilitating the progress of science, FAIR has a positive impact on innovation which translates into job creation and higher GDP."[133]

3) European Open Science Cloud (EOSC) and the Social Sciences and Humanities Open Cloud (SSHOC) project – Further information on the General Description

The vision of EOSC was conceived in 2015. EOSC is envisaged as a supporting landscape to foster open science and open innovation. Also, to remove human, policy, and technical barriers by bringing together existing scientific data infrastructures that are currently spread across countries and disciplines and are not at all or not sufficiently connected.

The cloud will provide a competitive European data and knowledge economy to ensure Europe a global lead in scientific data infrastructures. EOSC is intended to be Europe's virtual environment that enables seamless and efficient access to high-quality research data and data tools for all scientific disciplines. "The EOSC will offer 1.7 million European researchers

[132] PwC EU Services 2018b, 2018a
[133] PwC EU Services 2018a, p. 17

and 70 million professionals in science, technology, the humanities and social sciences a virtual environment with open and seamless services for storage, management, analysis and reuse of research data, across borders and scientific disciplines by federating existing scientific data infrastructures, currently dispersed across disciplines and the EU Member States."[134] EOSC will use the existing infrastructures, bring them together, and ensure their interoperability to allow users to work in multiple ecosystems.[135]

When discussing EOSC, there is one part most relevant for CESSDA and the Social Sciences, namely the SSHOC project. This project started on the first of January 2019 and is planned for 40 months. As a thematic cluster project for the Social Sciences and Humanities (SSH), the SSHOC project contributes to the realization of the EOSC.[136][137] With SSHOC, a "full-fledged Social Sciences and Humanities Open Cloud (SSHOC) where data, tools, and training are available and accessible for users of SSH data"[138] shall be provided. SSHOC sets out to develop and provide EOSC services for the SSH community by interconnecting existing and new infrastructures. The project adheres to the FAIR principles[139] and the principles of Open Science to maximize data reuse.[140] It provides resources, activities and a network for training and prepares an appropriate governance model for SSH-EOSC.[141] "Development, realisation and maintenance of user-friendly tools & services, covering all aspects of the full research data cycle will be built, taking into account human-centric approach and creating links between people, data, services and training."[142] [143]

[134] European Open Science Cloud 2020
[135] Dekker 2020; Ilijašić Veršić and Ausserhofer 2019; EOSC Secretariat 2019; Ilijašić Veršić and Dekker 2020; Wittenberg et al. 2020; European Open Science Cloud 2020
[136] Broeder et al. 2019
[137] Social Sciences & Humanities Open Cloud 2020a, 2020b; Ilijašić Veršić and Dekker 2020; Tóth-Czifra 2020; Wittenberg et al. 2020; Ilijašić Veršić and Ausserhofer 2019; Social Sciences & Humanities Open Cloud 2020c
[138] CESSDA 2020
[139] For more information, see section 4 in this report "FAIR principles (on Metadata)."
[140] Broeder et al. 2019
[141] Social Sciences & Humanities Open Cloud 2020a, 2020b; Ilijašić Veršić and Dekker 2020; Tóth-Czifra 2020; Wittenberg et al. 2020; Ilijašić Veršić and Ausserhofer 2019; Social Sciences & Humanities Open Cloud 2020c
[142] CESSDA 2020
[143] See also Figure 3 "ESFRI thematic cluster view."

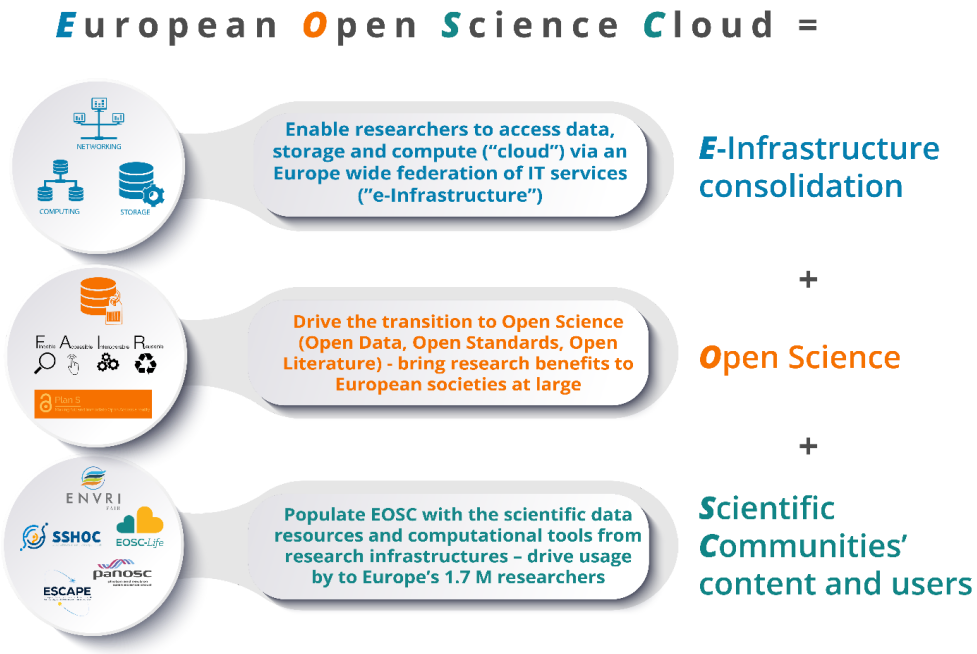

Figure 3: ESFRI thematic cluster view on EOSC.[144]

The result of the SSHOC project will be a data ecosystem for the SSH, integrated into the EOSC, that allows for seamless access to high-quality data. This means that the currently separated e-infrastructure facilities of the SSH need to be integrated into a cloud-based infrastructure.[145]

[144] Source: Blomberg and Petzold 2020.
[145] Ilijašić Veršić and Dekker 2020

4) Mapping between Dataverse and CMM

Table 3: Mapping between Dataverse[146] and CMM[147].

No.	Element	Status (Mandatory / Recommended / Optional)	Occ	Dataverse: DB Element	Dataverse: Field label
COMPLETE LIST OF ELEMENTS					
Information on Study:					
1	Study	M	1		
Information on Study: Bibliographic Information					
1.1	Bibliographic Information	M	1		
1.1.1	Study ID	M (for DDI3.2)	1-2		
1.1.2	Study Number	M (for DDI3.2)	1-n	otherID	Other ID
1.1.2.1	Type of Study Number (for DDI3.2)	M (for DDI3.2)	1		
1.1.3	Study Title	M (CDC)	1-n	title	Title
1.1.3.1	Language of Study Title	M (for DDI3.2)	1	No corresponding element in Dataverse to CMM element	
1.1.3.2	Translation Status of Study Title	R	0-1		
1.1.4	Subtitle	O	0-n	subtitle	Subtitle
1.1.4.1	Language of Subtitle	M	1		
1.1.4.2	Translation Status of Subtitle	R	0-1	alternativeTitle	Alternative Title
1.1.5	Alternative Title	O	0-n		

[146] Dataverse 2020
[147] Borschewski et al. 2019

1.1.5.1	Language of Alternative Title	M	1		
1.1.5.2	Translation Status of Alternative Title	R	0-1		
1.1.6	Funding Information	O	0-n		
1.1.6.1	Funding Agency Reference	O	0-n		
1.1.6.1.1	Type of Funding Agency Reference (for DDI3.2)	M (for DDI3.2)	1		
1.1.6.2	Grant Number	R	0-n	grantNumberValue	Grant Number
1.1.6.3	Grant Title	R	0-n	grantNumber	Grant Information
1.1.7	Principal Investigator Reference	R	1-n	dataCollector	Data Collector
1.1.7.1	Principal Investigator (Person Reference)	O	0-2		
1.1.7.1.1	Type of Principal Investigator (Person Reference) (for DDI3.2)	M (for DDI3.2)	1		
1.1.7.2	Principal Investigator (Institution Reference)	O	0-2		
1.1.7.2.1	Type of Principal Investigator (Institution Reference) (for DDI3.2)	M (for DDI3.2)	1		
1.1.8	Publisher	M (CDC)	1-n	distributorName	Name
1.1.8.1	Language of Publisher	M (for DDI3.2)	1	No corresponding element in Dataverse to CMM element	
1.1.8.2	Translation Status of Publisher	R	0-1		
1.1.9	Publication Year (controlled [YYYY])	R	1	distributionDate	Distribution Date
1.1.10	Study Version	R	0-1		
1.1.10.1	Version Number of the Study (freetext)	O	0-1		
1.1.10.2	Version Date of the Study (controlled)	O	0-1		
1.1.10.3	Reason for Versioning the Study	O	0-1		
1.1.10.3.1	Language of Versioning Reason	M	1		
1.1.10.3.2	Translation Status of Versioning Reason	R	0-1		
1.1.10.4	Study PID	R	0-n	otherIdValue	Identifier
1.1.10.4.1	Study PID Type	M	1	otherIdAgency	Agency

1.1.10.4.2	Study PID URL	R	0-1		
1.1.10.5	Study SPURL	M	1-n	distributorURL	URL
1.1.11	Contributor Reference	O	0-n	contributor	Contributor
1.1.11.1	Contributor (Person Reference)	O	0-2		
1.1.11.1.1	Type of Contributor (Person Rereference) (for DDI3.2)	M (for DDI3.2)	1		
1.1.11.2	Contributor (Institution Reference)	O	0-2	contributorName	Name
1.1.11.2.1	Type of Contributor (Institution Reference) (for DDI3.2)	M (for DDI3.2)	1		
1.1.11.3	Role of Contributor (freetext)	R	0-n	contributorType	Type
1.1.11.4	Role of Contributor (controlled)	M	0-n		
1.1.11.4.1	Role of Contributor Code ID (Descriptive Term)	R	0-n		
1.1.11.4.1.1	Language of Role of Contributor Code ID (Descriptive Term)				
1.1.11.4.2	Role of Contributor Vocabulary	M	1		
1.1.11.4.3	Role of Contributor Vocabulary Agency	O	0-1		
1.1.11.4.4	Role of Contributor Vocabulary URN	M	1		
1.1.11.5	Reference Study to Document	O	0-n		
Information on Study: Content Information					
1.2	Content Information	R	0-1		
1.2.1	Abstract	M (CDC)	1-n	description	Description
1.2.1.1	Language of Abstract	M (CDC)	1		
1.2.1.2	Translation Status of Abstract	R	0-1		
1.2.2	Study Topic (controlled)	M	1-n	topicClassification	Topic Classification
1.2.2.1	Study Topic Code ID (Descriptive Term)	R	0-n	topicClassValue	Term
1.2.2.1.1	Language of Study Topic (Descriptive Term)	M	1		
1.2.2.1.2	Translation Status of Study Topic (Descriptive Term)	R	0-1		

1.2.2.2	Topic Classification Vocabulary	M	1	topicClassVocab	Vocabulary
1.2.2.3	Topic Classification Agency	O	0-1		
1.2.2.4	Topic Classification URN	M	1	topicClassVocabURI	Vocabulary URL
1.2.3	Keyword (controlled)	O	0-n	keyword	Keyword
1.2.3.1	Keyword Code ID (Descriptive Term)	M	0-n	keywordValue	Term
1.2.3.1.1	Language of Keyword (Descriptive Term)	M	1		
1.2.3.1.2	Translation Status of Keyword (Descriptive Term)	R	0-1		
1.2.3.2	Keyword Vocabulary	M	1	keywordVocabulary	Vocabulary
1.2.3.3	Keyword Vocabulary Agency	O	0-1		
1.2.3.4	Keyword Vocabulary URN	O	0-1	keywordVocabularyURI	Vocabulary URL
Information on Study: Methodical Information					
1.3	Methodical Information	R	0-1		
1.3.1	Time Method (controlled)	M	1-n	timeMethod	Time Method
1.3.1.1	Time Method CodeID (Descriptive Term)	R	0-n		
1.3.1.1.1	Language of Time Method (Descriptive Term)	M	1		
1.3.1.2	Time Method Vocabulary	M	1		
1.3.1.3	Time Method Vocabulary Agency	O	0-1		
1.3.1.4	Time Method Vocabulary URN	M	1		
1.3.2	Study Area as a Country (controlled)	R	1-n	geographicUnit	Geographic Unit
1.3.2.1	Study Area Vocabulary	R	0-1		
1.3.2.2	Study Area Vocabulary Agency	O	0-1		
1.3.2.3	Study Area Vocabulary URN	O	0-1		
1.3.3	Study Area as a Country (freetext)	O	0-n	country	Country / Nation
1.3.3.1	Language of Study Area (freetext)	M	1		
1.3.3.2	Translation Status of Study Area (freetext)	R	0-1		

1.3.4	Universe (freetext)	R	0-n	universe	Universe
1.3.4.1	Language of Universe (freetext)	M	1		
1.3.4.2	Translation Status of Universe (freetext)	R	0-1		
1.3.5	Unit of Analysis (controlled)	M	1-n	unitOfAnalysis	Unit of Analysis
1.3.5.1	Unit of Analysis CodeID (Descriptive Term)	R	0-n		
1.3.5.1.1	Language of Unit of Analysis (Descriptive Term)				
1.3.5.2	Unit of Analysis Vocabulary	M	1		
1.3.5.3	Unit of Analysis Vocabulary Agency	O	0-1		
1.3.5.4	Unit of Analysis Vocabulary URN	M	1		
1.3.6	Sampling Information	R	0-n		
1.3.6.1	Type of Data Source (controlled)	M	1-n	dataSources	Data Sources
1.3.6.1.1	Type of Data Source (Descriptive Term)	R	0-n		
1.3.6.1.1.1	Language of Type of Data Source (Descriptive Term)	M	1		
1.3.6.1.2	Type of Data Source Vocabulary	M	1	originOfSources	Origin of Sources
1.3.6.1.3	Type of Data Source Vocabulary Agency	O	0-1		
1.3.6.1.4	Type of Data Source Vocabulary URN	M	1		
1.3.6.2	Type of Sampling Procedure (controlled)	M	1-n	samplingProcedure	Sampling Procedure
1.3.6.2.1	Type of Sampling Procedure (Descriptive Term)	R	0-n		
1.3.6.2.1.1	Language of Type of Sampling Procedure (Descriptive Term)	M	1		
1.3.6.2.2	Type of Sampling Procedure Vocabulary	M	1		
1.3.6.2.3	Type of Sampling Procedure Vocabulary Agency	O	0-1		
1.3.6.2.4	Type of Sampling Procedure Vocabulary URN	M	1		
1.3.6.3	Sampling Procedure (freetext)	O	0-n	targetSampleSize	Target Sample Size

1.3.6.3.1	Language of Sampling Procedure (freetext)	M	1		
1.3.6.3.2	Translation Status of Sampling Procedure (freetext)	R	0-1		
1.3.6.4	Mode of Data Collection (controlled)	M	1-n	collectionMode	Collection Mode
1.3.6.4.1	Mode of Data Collection (Descriptive Term)	R	0-n		
1.3.6.4.1	Language of Mode of Data Collection (Descriptive Term)	M	1		
1.3.6.4.2	Mode of Data Collection Vocabulary	M	1		
1.3.6.4.3	Mode of Data Collection Vocabulary Agency	O	0-1		
1.3.6.4.4	Mode of Data Collection Vocabulary URN	M	1		
1.3.6.5	Mode of Data Collection (freetext)	O	0-n		
1.3.6.5.1	Language of Mode of Data Collection (freetext)	M	1		
1.3.6.5.2	Translation Status of Mode of Data Collection (freetext)	R	0-1		
1.3.7	Survey Period	R	0-n	dateOfCollection	Date of Collection
1.3.7.1	Data Collection Period Startdate (controlled) The complete calender date with the structure YYYY-MM-DD e.g. 2018-03-24	R	0-n	dateOfCollectionStart	Start
1.3.7.2	Data Collection Period Enddate (controlled) The complete calender date with the structure YYYY-MM-DD e.g. 2018-03-24	R	0-n	dateOfCollectionEnd	End
1.3.7.3	Data Collection Single Date	O	0-n		
1.3.7.4	Data Collection Period (freetext)	O	0-n		
1.3.7.4.1	Language of Data Collection Period (freetext)	M	1		
1.3.7.4.2	Translation Status of Data Collection Period (freetext)	R	0-1		
Information on Study: Legal Information					
1.4	Access Information	M	1-n	termsofaccess	Terms of Access
1.4.1	Data Access (freetext)	M	1-n	No corresponding element in Dataverse to CMM element	
1.4.1.1	Language of Data Access (freetext)	M	1		
1.4.1.2	Translation Status of Data Access (freetext)	R	0-1		

1.4.2	Data Use Restrictions	O	0-n	restrictions	Restrictions
1.4.2.1	Language of Data Use Restrictions	M	1		
1.4.2.2	Translation Status of Data Use Restrictions	R	0-1		
Information on Persons:					
2	Person	R	0-n		
2.1	Person ID	M (for DDI3.2)	1-2	authorIdentifier	Identifier
2.2	Person Name	M	1	authorName	Name
2.2.1	Person: First Given Name	M	1	authorName	Name
2.2.2	Person: Middle Name	O	0-1		
2.2.3	Person: Last / Family Name	M	1	authorName	Name
2.3	Associated organization	O	0-n		
2.3.1	Associated Organization (Individual) Reference	M (for DDI3.2)	1-2	authorAffiliation	Affiliation
2.3.1.1	Type of Associated Organization (Individual) (for DDI3.2)	M (for DDI3.2)	1		
2.3.2	Role of the Target Object (Organization) in this Relation	M (for DDI3.2)	1		
2.3.3	Associated Organization (Organization)	M (for DDI3.2)	1-2		
2.3.3.1	Type of Associated Organization (Organization) (for DDI3.2)	M (for DDI3.2)	1		
2.4	Person PID	R	0-n		
2.4.1	Person PID Type	M	1		
2.4.2	Person PID URL	R	0-1		
Information on Organizations:					
3	Organization / Consortium	R	0-n	distributor	Distributor
3.1	Organization ID	M (for DDI3.2)	1-2		
3.2	Full Name of Organization	M	1-n	distributorName	Name
3.2.1	Language of Full Name of Organization	M	1		
3.2.2	Translation Status of Full Name of Organization	R	0-1		
3.3	Organization Name Abbreviation/Acronym	O	0-1	distributorAbbreviation	Abbreviation

3.4	Country (using ISO 3166 2-letter code)	O	0-n	country	Country / Nation
3.4.1	Country Vocabulary	M	1		
3.4.2	Country Vocabulary Agency	R	0-1		
3.4.3	Country Vocabulary URN	R	0-1		
Information on Dataset:					
4	Dataset	R	0-n		
Information on Dataset: Identification Information					
4.1	Identification Information	R	0-1		
4.1.1	Dataset ID	M (for DDI3.2)	1-2		
4.1.2	Dataset PID	R	0-n	[system generated]	Dataset Persistent ID
4.1.2.1	Dataset PID Type	M	1		
4.1.2.2	Dataset PID URL	R	0-1		
4.1.3	Reference of Study to Dataset	R	0-n		
4.1.3.1	Type of Reference between Study and Dataset (for DDI3.2)	M (for DDI3.2)	1		
4.1.4	Dataset Publication Date	R	0-1	productionDate	Publication Date
4.1.5	Dataset Version	R	0-1		
4.1.6	Dataset Version Description	O	0-1		
4.1.6.1	Language of Dataset Version Description	M	1		
4.1.6.2	Translation Status of Dataset Version Description	R	0-1		
4.1.7	Dataset Language	R	0-n		
Information on Dataset: Content Information			**0-1**		
4.2	Content Information	R	0-1		
4.2.1	Dataset File Description	O	0-n		
4.2.1.1	Language of Dataset File Description	M	1	language	Language

4.2.1.2	Translation Status of Dataset File Description	R	0-1		
4.2.2	File Name	M	1-n	relatedDatasets	Related Datasets
4.2.2.1	Language of File Name	M	1		
4.2.2.2	Translation Status of File Name	R	0-1		
4.2.3	Number of Units	O	0-1		
4.2.4	Number of Variables	O	0-1		
Information on Dataset: Variables					
4.3	Variable	R	0-n		
4.3.1	Basic Information about Variable	M	1		
4.3.1.1	Variable ID	M (for DDI3.2)	1-2		
4.3.1.2	Variable Name	M	1-n		
4.3.1.2.1	Language of Variable Name	R	0-1		
4.3.1.2.2	Translation Status of Variable Name	R	0-1		
4.3.1.3	Variable Label	R	0-n		
4.3.1.3.1	Language of Variable Label	M	1		
4.3.1.3.2	Translation Status of Variable label	R	0-1		
4.3.1.4	Derivation Instructions	O	0-n		
4.3.1.5	Variable Description	O	0-n		
4.3.1.5.1	Language of the Variable Description	M	1		
4.3.1.5.2	Translation Status of the Variable Description	R	0-1		
4.3.1.6	Variable Note	O	0-n		
4.3.1.6.1	Language of Variable Note	M	1		
4.3.1.6.2	Translation Status of Variable Note	R	0-1		
4.3.1.6.3	Type of Note (controlled)	M	1		

4.3.1.6.3.1	Type of Note Vocabulary	M (for DDI3.2)	1		
4.3.1.6.3.2	Type of Note Vocabulary Agency	R	0-1		
4.3.1.6.3.3	Type of Note Vocabulary URN	M (for DDI3.2)	1		
4.3.1.7	Variable Type	R	0-1		
4.3.1.8	Measurement Level	O	0-1		
4.3.1.9	Weighting Variable	R	0-1	weighting	Weighting
4.3.2	Reference Information	R	0-1		
4.3.2.1	Reference Variable to Concept	R	0-n		
4.3.2.1.1	Type of Reference between Variable and Concept (for DDI3.2)	M (for DDI3.2)	1		
4.3.2.2	Reference Variable to Question	R	0-n		
4.3.2.2.1	Type of Reference Variable to Question (for DDI3.2)	M (for DDI3.2)	1		
4.3.2.3	Reference Variable to Codelist	O	0-2		
4.3.2.3.1	Type of Reference between Variable and Codelist (for DDI3.2)	M (for DDI3.2)	1		
4.3.3	Statistical Information	R	0-1		
4.3.3.1	Minimum (Summary Statistics)	R	0-1		
4.3.3.1.1	Minimum Summary Statistics Vocabulary	R	0-1		
4.3.3.1.2	Minimum Summary Statistics Vocabulary Agency	O	0-1		
4.3.3.1.3	Minimum Summary Statistics Vocabulary URN	R	0-1		
4.3.3.1.4	Minimum Summary Statistics Value	M	1		
4.3.3.2	Maximum (Summary Statistics)	R	0-1		
4.3.3.2.1	Maximum Summary Statistics Vocabulary	R	0-1		
4.3.3.2.2	Maximum Summary Statistics Vocabulary Agency	O	0-1		
4.3.3.2.3	Maximum Summary Statistics Vocabulary URN	R	0-1		
4.3.3.2.4	Maximum Summary Statistics Value	M	1		
4.3.3.3	Arithmetic Mean (Summary Statistics)	R	0-1		

4.3.3.3.1	Arithmetic Mean Summary Statistics Vocabulary	R	0-1		
4.3.3.3.2	Arithmetic Mean Summary Statistics Vocabulary Agency	O	0-1		
4.3.3.3.3	Arithmetic Mean Summary Statistics Vocabulary URN	R	0-1		
4.3.3.3.4	Arithmetic Mean Summary Statistics Value	M	1		
4.3.3.4	Standard Deviation (Summary Statistics)	R	0-1		
4.3.3.4.1	Standard Deviation Summary Statistics Vocabulary	R	0-1		
4.3.3.4.2	Standard Deviation Summary Statistics Vocabulary Agency	O	0-1		
4.3.3.4.3	Standard Deviation Summary Statistics Vocabulary URN	R	0-1		
4.3.3.4.4	Standard Deviation Summary Statistics Value	M	1		
4.3.4	Codelist	R	0-n		
4.3.4.1	Codelist ID	M (for DDI3.2)	1-2		
4.3.4.2	Variable Value Code	M	1		
4.3.4.3	Category ID	M (for DDI3.2)	1-2		
4.3.4.4	Category Name	M	1		
4.3.4.4.1	Language of Category Name	M	1		
4.3.4.4.2	Translation Status of Category Name	R	0-1		
4.3.4.5	Reference Code to Category Name	M (for DDI3.2)	1-2		
4.3.4.5.1	Type of Reference between Code and Category Name (for DDI3.2)	M (for DDI3.2)	1		
Information on Dataset: Technical Information					
4.4	Technical Information	R	0-1		
4.4.1	Dataset Software	M	1	softwareName	Name
4.4.2	Dataset File Checksum Value	M	1		
4.4.2.1	Dataset Checksum Algorithm	M	1		
Information on Instrument:					

5	Instrument	R	0-n		
5.1	Instrument ID	M (for DDI3.2)	1-2		
5.2	Reference Instrument to Study	M (for DDI3.2)	1-n		
5.2.1	Type of Reference Instrument to Study (for DDI3.2)	M (for DDI3.2)	1		
Information on Instrument: Content Information					
5.3	Content Information	M	1		
5.3.1	Type of Instrument (controlled)	M	1	researchInstrument	Type of Research Instrument
5.3.1.1	Type of Instrument (Descriptive Term)	R	0-n		
5.3.1.1.1	Language of Type of Instrument (Descriptive Term)	M	1		
5.3.1.2	Type of Instrument Vocabulary	M	1		
5.3.1.3	Type of Instrument Vocabulary Agency	O	0-1		
5.3.1.4	Type of Instrument Vocabulary URN	M	1		
5.3.2	Instrument Name	M	1-n		
5.3.2.1	Language of Instrument Name	M	1		
5.3.2.2	Translation Status of Instrument Name	R	0-1		
5.3.3	Instrument Language (controlled)	M (for DDI3.2)	1		
5.3.4	Instrument Source	O	0-1		
5.3.5	Instrument Source Reference	O	0-n		
5.3.5.1	Type of Instrument Source Reference (for DDI3.2)	M (for DDI3.2)	1		
Information on Instrument: Technical Information					
5.4	Technical Information	O	0-1		
5.4.1	Reference of Instrument to Question	M (for DDI3.2)	1-2		
5.4.1.1	Type of Reference of Instrument to Question	M (for DDI3.2)	1		
5.4.2	Instrument PID	O	0-n		
5.4.2.1	Instrument PID Type	M	1		

5.4.2.2	Instrument PID URL	R	0-1	
5.4.3	Instrument Description	O	0-1	
5.4.4	Instrument File Checksum Value	O	0-1	
5.4.4.1	Instrument Checksum Algorithm	M	1	
Information on Questions and Responses:			**0-n**	
6	Questions and Responses	R	0-1	
Information on Questions and Responses: Questions				
6.1	Questions	R	1-n	
6.1.1	Question Item	R	0-n	
6.1.1.1	Question Item ID	M (for DDI3.2)	1-2	No corresponding element in Dataverse to CMM element
6.1.1.2	Question Item Version	R	0-1	
6.1.1.3	User ID for Question Item	M	1-n	
6.1.1.3.1	Type of User ID for Question Item	M (for DDI3.2)	1	
6.1.1.4	Question Item Name	O	0-n	
6.1.1.4.1	Context for the Question Item Name (to separate Question Number and Label for DDI3.2)	M (for DDI3.2)	1	
6.1.1.4.2	Language of Question Item Name	M	1	
6.1.1.4.3	Translation Status of Question Item Name	R	0-1	
6.1.1.5	Question Item Text	M	1-n	No corresponding element in Dataverse to CMM element
6.1.1.5.1	Language of Question Text	M	1	No corresponding element in Dataverse to CMM element
6.1.1.5.2	Translation Status of Question Text	R	0-1	
6.1.1.6	Question Item Statement	O	0-n	
6.1.1.6.1	Language of Question Item Statement	M	1	
6.1.1.6.2	Translation Status of Question Item Statement	R	0-1	
6.1.1.7	Question Item Interviewer Instruction	R	0-n	
6.1.1.7.1	Language of Question Item Interviewer Instruction	M	1	
6.1.1.7.2	Translation Status of Question Item Interviewer Instruction	R	0-1	

6.1.1.8	Question Item Number	R	0-n	
6.1.1.8.1	Context for the Question Item Number (to separate Question Number and Label for DDI3.2)	M (for DDI3.2)	1	
6.1.1.8.2	Language of Question Item Number	M	1	
6.1.1.8.3	Translation Status of Question Item Number	R	0-1	
6.1.1.9	Reference of Question Item to Response Domain	R	0-2	
6.1.1.9.1	Type of Response Domain (Domain Reference)	M (for DDI3.2)	1	
6.1.1.10	Reference of Question Item to Concept	R	0-2	
6.1.1.10.1	Type of Concept (for DDI3.2)	M (for DDI3.2)	1	
6.1.1.11	Showcards	R	0-n	
6.1.1.11.1	Showcard File Name	M	1	
6.1.1.11.1.1	Language of the Showcard File Name	M	1	
6.1.1.11.1.2	Translation Status of the Showcard File Name	R	0-1	
6.1.1.11.2	Showcard URL	R	0-1	
6.1.2	Questions - Question Grid	R	0-n	
6.1.2.1	Question Grid Version	O	0-1	
6.1.2.2	Question Grid ID	M	1-n	No corresponding element in Dataverse to CMM element
6.1.2.2.1	Type of Question Grid ID (for DDI3.2)	M (for DDI3.2)	1	
6.1.2.3	Question Grid Name	O	0-n	
6.1.2.3.1	Context of Question Grid Name (to separate label and number for DDI3.2)	M (for DDI3.2)	1	
6.1.2.3.2	Language of Question Grid Name	M	1	
6.1.2.3.3	Translation Status of Question Grid Name	R	0-1	
6.1.2.4	Question Grid Text	M	1-n	No corresponding element in Dataverse to CMM element
6.1.2.4.1	Language of Question Grid Text	M	1	No corresponding element in Dataverse to CMM element
6.1.2.4.2	Translation Status of Question Grid Text	R	0-1	

6.1.2.5	Question Grid Statement	O	0-n
6.1.2.5.1	Language of Question Grid Statement	M	1
6.1.2.5.2	Translation Status of Question Grid Statement	R	0-1
6.1.2.6	Question Grid Interviewer Instruction	O	0-n
6.1.2.6.1	Language of Question Grid Interviewer Instruction	M	1
6.1.2.6.2	Translation Status of Question Grid Interviewer Instruction	R	0-1
6.1.2.7	Question Grid Number	O	0-n
6.1.2.7.1	Context of Question Grid Number (to separate label and number for DDI3.2)	M (for DDI3.2)	1
6.1.2.7.2	Language of Question Grid Number	M	1
6.1.2.7.3	Translation Status of Question Grid Number	R	0-1
6.1.2.8	Reference of Question Grid to a Concept	R	0-n
6.1.2.8.1	Type of Object for Concept (for DDI3.2)	M (for DDI3.2)	1
6.1.2.9	Grid Dimension	R	0-n
6.1.2.9.1	Subquestion Number	R	0-n
6.1.2.9.2	Subquestion Text	M	1-n
6.1.2.9.2.1	Language of Subquestion Text	M	1
6.1.2.9.2.2	Translation Status of Subquestion Text	R	0-1
6.1.2.9.3	Response Domain for Grid Dimension	R	0-n
6.1.2.9.3.1	Reference of Question Grid to Response Domain	R	0-2
6.1.2.9.3.1.1	Type of Response Domain (Domain Reference) for Question Grid	M (for DDI3.2)	1
6.1.2.9.3.2	Out Parameter	O	0-1
6.1.2.9.3.3	Grid Attachment	M	1
6.1.2.10	Showcards Question Grid	R	0-n
6.1.2.10.1	Showcard File Name Question Grid	M	1

6.1.2.10.1.1	Language of the Showcard Question Grid	M	1
6.1.2.10.1.2	Translation Status of the Showcard Question Grid	R	0-1
6.1.2.10.2	Showcard Question Grid URL	O	0-1
Information on Questions and Responses: Responses			
6.2	Response Domain Type	R	0-n
6.2.1	Code Domain	O	0-1
6.2.1.1	Codelist	M (for DDI3.2)	1
6.2.1.2	Codelist UserID	M (for DDI3.2)	1-2
6.2.1.3	User ID for Codelist	M	1-n
6.2.1.3.1	Type of User ID for Codelist	M (for DDI3.2)	1
6.2.1.4	Codelist Name	O	0-1
6.2.1.4.1	Language of the Codelist Name	M	1
6.2.1.4.2	Translation Status of the Codelist Name	R	0-1
6.2.1.5	Codelist Label	O	0-1
6.2.1.5.1	Language of the Codelist Label	M	1
6.2.1.5.2	Translation Status of the Codelist Label	R	0-1
6.2.1.6	Codelist Description	O	0-1
6.2.1.6.1	Language of the Codelist Description	M	1
6.2.1.6.2	Translation Status of the Codelist Description	R	0-1
6.2.1.7	Codelist Response Cardinality	O	0-1
6.2.1.7.1	Codelist Response Cardinality Minimum	O	0-1
6.2.1.7.2.	Codelist Response Cardinality Maximum	O	0-1
6.2.1.8	Answer Code	M	1-n
6.2.1.9	Answer Category	M	1-n
6.2.1.9.1	Language of the Answer Category	M	1
6.2.1.9.2	Translation Status of the Answer Category	R	0-1

6.2.1.9.3	Answer Missing Value	O	0-1
6.2.1.10	Reference of Answer Code to Answer Category	R	0-n
6.2.1.10.1	Type of Object for Category (for DDI3.2)	M (for DDI3.2)	1
6.2.2	Numeric Domain	O	0-1
6.2.2.1	Numeric Domain Name	O	0-1
6.2.2.1.1	Language of the Numeric Domain Name	M	1
6.2.2.1.2	Translation Status of the Numeric Domain Name	R	0-1
6.2.2.2	Numeric Domain Label	O	0-1
6.2.2.2.1	Language of the Numeric Domain Label	M	1
6.2.2.2.2	Translation Status of the Numeric Domain Label	R	0-1
6.2.2.3	Numeric Domain Description	O	0-1
6.2.2.3.1	Language of the Numeric Domain Description	M	1
6.2.2.3.2	Translation Status of the Numeric Domain Description	R	0-1
6.2.2.4	Numeric Domain Range: Low	O	0-1
6.2.2.5	Numeric Domain Range: High	O	0-1
6.2.3	Text Domain	O	0-1
6.2.3.1	Text Domain Name	O	0-1
6.2.3.1.1	Language of the Text Domain Name	M	1
6.2.3.1.2	Translation Status of the Text Domain Name	R	0-1
6.2.3.2	Text Domain Label	O	0-1
6.2.3.2.1	Language of the Text Domain Label	M	1
6.2.3.2.2	Translation Status of the Text Domain Label	R	0-1
6.2.3.3	Text Domain Description	O	0-1
6.2.3.3.1	Language of the Text Domain Description	M	1
6.2.3.3.2	Translation Status of the Text Domain Description	R	0-1
6.2.3.4	Text Domain Length: Min	O	0-1
6.2.3.5	Text Domain Length: Max	O	0-1

Information on Concepts:					
7	Concept	O	0-n		
7.1	Concept ID	M (for DDI3.2)	1-2		
7.2	Concept Name	R	0-1		
7.2.1	Language of the Concept Name	M	1		
7.2.2	Translation Status of the Concept Name	R	0-1		
7.3	Concept Label	O	0-1		
7.3.1	Language of the Concept Label	M	1		
7.3.2	Translation Status of the Concept Label	R	0-1		
7.4	Concept Description	O	0-1		
7.4.1	Language of the Concept Description	M	1		
7.4.2	Translation Status of the Concept Description	R	0-1		
7.5	Concept Reference	O	0-1		
7.5.1	Type of Reference for Concept (for DDI3.2)	M (for DDI3.2)	1		
Information on further Documents:		**O**	**0-n**		
8	Document	O	0-n	relatedMaterial	Related Material
8.1	Document ID	R	0-1		
8.1.1	Type of Document ID (for DDI3.2)	M (for DDI3.2)	1		
8.2	Document Title	M	1		
8.2.1	Language of Document Title	M	1		
8.2.2	Translation Status of Document Title	R	0-1		
8.3	Document URL	O	0-1		
8.4	Document Format	M	1		
Information on Publication (publications where data have been used):		**R**	**0-n**		
9	Publication	R	0-n	publication	Related Publication
9.1	Reference of Publication to Study	M (for DDI3.2)	1-n		

9.1.1	Type of Reference Publication to Study (for DDI3.2)	M (for DDI3.2)	1		
9.2	Type of Material of Publication (for DDI3.2)	M (for DDI3.2)	1		
9.3	Author Name (reference)	R	0-n	author	Author
9.3.1	Type of Author Name (reference) (for DDI3.2)	M (for DDI3.2)	1		
9.4	Editor Name (reference)	R	0-n		
9.4.1	Type of Editor Name (reference) (for DDI3.2)	M (for DDI3.2)	1		
9.4.2	Contributor Role (freetext)	R	0-n		
9.5	Publication Title	M	1-n		
9.5.1	Language of Publication Title	M	1		
9.5.2	Translation Status of Publication Title	R	0-1		
9.6	Publication Year [YYYY] (controlled)	M	1	distributionDate	Distribution Date
9.7	Publisher Name (reference)	R	0-2		
9.7.1	Type of Publisher Name (reference) (for DDI3.2)	M (for DDI3.2)	1		
9.8	Journal/Serial Name	R	0-1		
9.9	Volume Number	R	0-1		
9.10	Issue Number	R	0-1		
9.11	Compilation Name	R	0-1		
9.12	Pages	R	0-1		
9.13	Publication PID	R	0-n	publicationIDNumber	ID Number
9.13.1	Publication PID Type	M	1	publicationIDType	ID Type
9.13.2	Publication PID URL	R	0-1	publicationURL	URL
9.14	Publication URL	R	0-1		
Information on Group of Studies:					
10	Study Group	O	0-n	series	Series
10.1	Study Group ID	M (for DDI3.2)	1-2		

10.2	User ID for Study Group	O	0-n		
10.2.1	Type of UserID for Study Group ID (for DDI3.2)	M (for DDI3.2)	1		
10.3	Study Group Name	M	1-n	seriesName	Name
10.3.1	Language of Study Group Name	M	1		
10.3.2	Translation Status of Study Group Name	R	0-1		
10.4	Study Group Description	R	0-n	seriesInformation	Information
10.4.1	Language of Study Group Description	M	1		
10.4.2	Translation Status of Study Group Description	R	0-1		
10.4.3	Study Group Description URI	R	0-1		
10.5	Reference Study Group to Study Unit	M (for DDI3.2)	1-n		
10.5.1	Type of Reference Study Group to Study Unit (reference) (for DDI3.2)	M (for DDI3.2)	1		
10.6	Reference to Document	O	0-n		
Information on Document Description ("metadata about metadata"):					
11	Study Documentation Description	M	1		
11.1	Study Documentation Copyright	M	0-n		
11.1.1	Language of Study Documentation Copyright	M	1		
11.1.2	Translation Status of Study Documentation Copyright	R	0-1		
11.2	Study Documentation Publication Date (controlled)	O	0-1		
11.3	Study Documentation Publisher (reference)	M	1-2		
11.3.1	Type of Study Documentation Publisher (reference) (for DDI3.2)	M (for DDI3.2)	1		

5) Listing of further metadata initiatives, projects, institutions, etc. for the SSH

Table 4: Listing of further metadata initiatives, projects, institutions, etc.

Name of Project/ Standard / Initiative	Fast Facts	Link
ARTS + HUMANITIES		
Encoded Archival Description (EAD)	"A standard for encoding archival finding aids using XML in archival and manuscript repositories, implementing the recommendations of the International Council on Archives."[148]	Standard website: https://www.loc.gov/ead/
TEI- Text Encoding Initiative	"The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics. Since 1994, the TEI Guidelines have been widely used by libraries, museums, publishers, and individual scholars to present texts for online research, teaching, and preservation."[149]	Standard website: https://tei-c.org/
MARC (Machine-Readable Cataloging)	"MARC is a standard and serialization format for representing bibliographic metadata, originally designed as a way of exchanging bibliographic records between library catalogs. Various different versions have been defined, mostly with national or regional scope, of which MARC 21 is probably the most widely used. There also exists an XML serialization of MARC 21, known as MARCXML."[150]	Standard website: http://www.loc.gov/marc/
SOCIAL + BEHAVIORAL SCIENCES		
QuDEx (Qualitative Data Exchange Format)	"The QuDEx standard/schema is a software-neutral format for qualitative data that preserves annotations of, and relationships between, data and other related objects. It can be viewed as the optimal baseline data exchange model for the archiving and interchange of data and metadata."[151]	Standard website: https://www.data-archive.ac.uk/managing-data/standards-and-procedures/metadata-standards/qudex/
SDMX (Statistical Data and	"A set of common technical and statistical standards and guidelines to be used for the efficient exchange and sharing of statistical data and metadata."[152]	Standard website: https://sdmx.org/

[148] RDA | Metadata Directory 2015
[149] Text Encoding Initiative n.d.
[150] RDA | Metadata Directory 2018a
[151] Digital Curation Centre 2020g
[152] Digital Curation Centre 2020h

Metadata Exchange)		
ARTS/ HUMANITIES + SOCIAL/BEHAVIORAL SCIENCES		
DDI (Data Documentation Initiative)	"The Data Documentation Initiative (DDI) is an international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences."[153] [154]	Standard website: http://www.ddialliance.org/
MIDAS-Heritage	"A British cultural heritage standard for recording information on buildings, archaeological sites, shipwrecks, parks and gardens, battlefields, areas of interest and artefacts."[155]	Standard website: http://www.heritage-standards.org.uk
OAI-ORE (Open Archives Initiative Object Reuse and Exchange)	"The goal of these standards is to expose the rich content in aggregations of Web resources to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. The standards support the changing nature of scholarship and scholarly communication, and the need for cyberinfrastructure to support that scholarship, with the intent to develop standards that generalize across all web-based information including the increasing popular social networks of 'Web 2.0.' "[156]	Standard website: http://www.openarchives.org/ore/
SSHOC	"The project aims to provide a full-fledged Social Sciences and Humanities Open Cloud (SSHOC) where data, tools, and training are available and accessible for users of SSH data. […] The intention is to create a European open cloud ecosystem for social sciences and humanities, consisting of an infrastructural and human component."[157] [158]	Standard website: https://sshopencloud.eu/
GENERAL RESEARCH DATA		
CERIF (Common European Research Information format)	"The Common European Research Information Format is the standard that the EU recommends to its member states for recording information about research activity. Since version 1.6 it has included specific support for recording metadata for datasets."[159]	Standard website: https://www.eurocris.org/cerif/main-features-cerif

[153] DDI Alliance 2020e
[154] See also sections 2 and 3 on DDI above in this report.
[155] Digital Curation Centre 2020e
[156] researchobject.org 2020
[157] CORDIS 2020
[158] See also section 5 on EOSC and SSHOC above in this report.
[159] Digital Curation Centre 2020a

Data Package	"The Data Package specification is a generic wrapper format for exchanging data. Although it supports arbitrary metadata, the format defines required, recommended, and optional fields for both the package as a whole and the resources contained within it."[160]	Standard website: https://specs.frictionlessdata.io/data-package/
DataCite Metadata Schema	"A set of mandatory metadata that must be registered with the DataCite Metadata Store when minting a DOI persistent identifier for a dataset. The domain-agnostic properties were chosen for their ability to aid in accurate and consistent identification of data for citation and retrieval purposes."[161]	Standard website: http://schema.datacite.org/
DCAT (Data Catalog Vocabulary)	"DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web. By using DCAT to describe datasets in data catalogs, publishers increase discoverability and enable applications easily to consume metadata from multiple catalogs. It further enables decentralized publishing of catalogs and facilitates federated dataset search across sites."[162]	Standard website: https://www.w3.org/TR/vocab-dcat/
Dublin Core	"A basic, domain-agnostic standard which can be easily understood and implemented, and as such is one of the best known and most widely used metadata standards. Sponsored by the Dublin Core Metadata Initiative, Dublin Core was published as ISO Standard 15836 in February 2009."[163]	Standard website: http://dublincore.org/
OLAC (Open Language Archives Community)	"For describing language resources in linguistics research."[164]	Standard website: http://www.language-archives.org/OLAC/metadata.html
MODS (Metadata Object Description Schema)	"The Metadata Object Description Schema (MODS) is a bibliographic metadata standard implemented in XML. It reimplements a subset of the elements of MARC (Machine Readable Cataloging) using language-based tags instead of numeric ones, and groups them somewhat differently. It is intended both as a simplified version of MARC 21 and as a richer alternative to Dublin Core for applications such as metadata syndication/harvesting and the documentation of digital information packages."[165]	Standard website: http://www.loc.gov/standards/mods/
Observations and Measurements	"This encoding is an essential dependency for the OGC Sensor Observation Service (SOS) Interface Standard. More specifically, this standard defines XML schemas for observations, and for features involved in sampling when making observations. These provide document models for	Standard website: http://www.opengeospatial.org/standards/om

[160] Digital Curation Centre 2020b
[161] Digital Curation Centre 2020c
[162] European Data Portal n.d.
[163] Digital Curation Centre 2020d
[164] Simons and Bird 2006
[165] RDA | Metadata Directory 2018b

	the exchange of information describing observation acts and their results, both within and between different scientific and technical communities."[166]	
PREMIS (Long term preservation)	"The PREMIS (Preservation Metadata Implementation Strategies) Data Dictionary defines a set of metadata that most repositories of digital objects would need to record and use in order to preserve those objects over the long term. It has its roots in the Open Archival Information System Reference Model but has been strongly influenced by the practical experience of such repositories. While the Data Dictionary can be used with other standards to influence the creation of local application profiles, an XML Schema is provided to allow the metadata to be serialized independently."[167]	Standard website: http://www.loc.gov/standards/premis/
PROV	"Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness. The PROV Family of Documents defines a model, corresponding serializations and other supporting definitions to enable the inter-operable interchange of provenance information in heterogeneous environments such as the Web."[168]	Standard website: https://www.w3.org/2001/sw/wiki/PROV
RDF Data Cube Vocabulary	"There are many situations where it would be useful to be able to publish multi-dimensional data, such as statistics, on the web in such a way that it can be linked to related data sets and concepts. The Data Cube vocabulary provides a means to do this using the W3C RDF (Resource Description Framework) standard. The model underpinning the Data Cube vocabulary is compatible with the cube model that underlies SDMX (Statistical Data and Metadata eXchange), an ISO standard for exchanging and sharing statistical data and metadata among organizations. The Data Cube vocabulary is a core foundation which supports extension vocabularies to enable publication of other aspects of statistical data flows or other multi-dimensional data sets."[169]	Standard website: https://www.w3.org/TR/vocab-data-cube/
Others		
Dataverse Project	"Dataverse is an open source web application to share, preserve, cite, explore, and analyze research data. It facilitates making data available to others, and allows you to replicate others' work more easily. Researchers, journals, data authors, publishers, data distributors, and affiliated institutions all receive academic credit and web visibility. […] A Dataverse repository is the software installation, which then hosts multiple virtual archives called Dataverses. Each Dataverse contains datasets, and each dataset contains descriptive metadata and data files	Standard website: https://dataverse.org/

[166] Open Geospatial Consortium 2020
[167] Digital Curation Centre 2020f
[168] World Wide Web Consortium 2013
[169] World Wide Web Consortium 2014

	(including documentation and code that accompany the data). As an organizing method, Dataverses may also contain other Dataverses."[170] [171]	
CLOSER	"CLOSER Discovery is a user-friendly online resource which allows users to search and locate variables across multiple UK longitudinal studies for research purposes. The aim of CLOSER Discovery is to provide a portal to hundreds of thousands of variables, questions and data collection instruments from across the CLOSER studies. […] CLOSER Discovery is built using metadata."[172]	Standard website: https://discovery.closer.ac.uk/
EOSC (European Open Science Cloud)	"The European Open Science Cloud (EOSC) initiative has been proposed in 2016 by the European Commission as part of the European Cloud Initiative to build a competitive data and knowledge economy in Europe."[173] [174]	Standard website: https://www.eosc-portal.eu/
CaRCC People Network Data-Facing Track	"CaRCC – the Campus Research Computing Consortium – is an organization of dedicated professionals developing, advocating for, and advancing campus research computing and data* and associated professions."[175]	Standard website: https://carcc.org/people-network/data-facing-track/
METS (Metadata Encoding and Transmission standard)	"The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium."[176]	Standard website: http://www.loc.gov/standards/mets/
UNESCO Thesaurus	"The UNESCO Thesaurus is a controlled and structured list of terms used in subject analysis and retrieval of documents and publications in the fields of education, culture, natural sciences, social and human sciences, communication and information. Continuously enriched and updated, its multidisciplinary terminology reflects the evolution of UNESCO's programmes and activities."[177]	Standard website: http://vocabularies.unesco.org/browser/thesaurus/en/

[170] Dataverse n.d.
[171] See also section 6 on Dataverse above in this report.
[172] Mills 2017
[173] European Open Science Cloud 2020
[174] See also section 5 on EOSC and SSHOC above in this report.
[175] Campus Research Computing Consortium 2020
[176] Metadata Encoding & Transmission Standard 2019
[177] UNESCO 2020