Version 1.0
Last updated: 15.01.2020

NFDI4Health Task Force COVID-19 Metadata Schema

Authors:

Aliaksandra Shutsko, Birte Lindstädt, ZB MED − Information Centre Life Sciences, Cologne

In close cooperation with:

Carsten Oliver Schmidt, University Medicine of Greifswald
Johannes Darms, ZB MED − Information Centre Life Sciences, Bonn
Martin Golebiewski, Heidelberg Institute for Theoretical Studies, Heidelberg
Moritz Lehne, Sophie Klopfenstein, Carina Nina Vorisek, Berlin Institute of Health, Berlin
Matthias Löbe, University of Leipzig, Leipzig

Introduction

The NFDI4Health Task Force COVID-19 Metadata Schema (Metadata Schema) contains a list of properties describing a resource being registered in the Study Hub of the NFDI4Health Task Force COVID-19 (Study Hub).

Table 1 provides an overview of eight main properties of the Metadata Schema. Each property is further subdivided into several items. Definitions, allowed values, formats, and examples of the items can be found in Table 2.

Table 1. Overview of the main Metadata Schema properties

No	Property	Description
1	Resource	Resource is any object to be published on the study hub. This can be a study, a sub-study, a study document, a data collection instrument or some other resource types.
2	Resource Title	Scientific title of the resource.
3	Resource Acronym	Acronym of the resource (if existing).
4	Resource Description	General description of the resource.
5	Resource ID	Related and alternate identifier(s) of the resource.
6	Resource Role	Function performed by the responsible party, including description of the institution(s) or person(s) contributing to the development of the resource.
7	Study Design	Description of a study or sub-study design. The properties should be filled in only if a study or sub-study is being registered as a resource. The properties refer to both interventional and non-interventional studies.
8	Interventional Study Design	Description of an interventional study or sub-study design. The properties should be filled in only if an interventional study or sub-study is being registered.

Table 2 contains the full list of Metadata Schema items.

There are two different levels of obligation (Ob) for the Metadata Schema items (see the third column of Table 2):

- Mandatory (M) items must be provided,
- Recommended (R) items are strongly recommended.

To make the resource findable and items interoperable, it is strongly recommended to submit the values for both Recommended and Mandatory items.

The fourth column, Occurrence (Occ), indicates cardinality/quantity constraints for the items as follows:

- 0-n = optional and repeatable,
- 0-1 = optional, but not repeatable,
- 1-n = required and repeatable,
- 1 = required, but not repeatable.

As indicated in the Publication Policy, the metadata should be preferably provided in English. However, the information in German can also be accepted (in this case resource description (see Table 2, Property 4) must be provided in both German and English). Additionally, resource title (see Table 2, Property 2) can be provided in the original language (i.e. German, French, Spanish) but together with its English translation.

Table 2. Extended description of the Metadata Schema properties

No	Property/Item	Ob	Occ	Definition	Allowed values and formats, examples, other notes
1	**Resource**				
1.1	resource_type	M	1	A short description of the resource.	Select one value from the list. *Controlled List Values:* Study, Substudy, Dataset, Study Protocol, Protocol Amendment, Data Dictionary, Informed Consent Form, Patient Information Sheet, Manual of Operations (SOPs), Statistical Analysis Plan, Data Management Plan, Case Report Form, Code Book, Questionnaire, Interview Scheme and Themes, Observation Guide, Discussion Guide, Participant Tasks, Other Data Collection Instrument, Other Study Document, Other
1.2	resource_type_general	M	1	A more general description of the resource.	It is used to enable DOI-assignment. The categories are predefined by DataCite. For category definitions and examples, see the latest version of the DataCite Metadata Schema at https://schema.datacite.org/. Select one value from the list. *Controlled List Values:* Audiovisual Collection Data Paper Dataset Event Image Interactive Resource Model Physical Object Service Software Sound

					Text Workflow Other *Example:* If a resource (i.e., resource_type) is a "Study", use "Other" for this property. If a resource (i.e., resource_type) is an "Other Study Document", use "Text" for this property. If you are not sure which type to use, select "Other".
1.3	resource_language	M	1	The primary language of the resource, i.e. the language, in which a study was conducted, or the language, in which a study document has been originally composed.	Select one value from the list. *Controlled List Values:* EN (English) DE (German) ES (Spanish) FR (French)
1.4	resource_use_rights	M	1	License defining the rights to re-use the resource. In lines with the FAIR Guiding Principles, an open license is preferred.	License is applicable only if a study document is to be published. Select one value from the list. *Controlled List Values:* CC BY 4.0 CC BY-ND 4.0 N/A Other "N/A" can be selected if a license is not applicable (e.g., for metadata of a resource). "Other" can be selected in case if another license type is assigned or if it is unknown.
1.5	resource_web_page	R	0-1	If existing, web page directly relevant to the resource.	Format: URL starting with http://
2	**Resource Title**				
2.1	title	M	1-2	Scientific unabbreviated title of the resource.	The title must be provided in English. If original title of the resource is non-English, please provide both titles: an original title and its English translation.
2.2	title_language	M	1-2	Language of the title.	Four languages are currently accepted. *Controlled List Values:* EN (English), DE (German), ES (Spanish), FR (French)
3	**ResourceAcronym**				
3.1	acronym	R	0-1	If existing, the acronym of the resource.	Only use an acronym that is officially used for the resource. Otherwise, leave this property empty.
4	**ResourceDescription**				

4.1	description_text	M	1-2	A short plain text summary describing the resource in English. The provided information should be understandable by a scientific audience.	The description must be written in English. Additionally, text in German can be provided.
4.2	description_language	M	1-2	The language of the description text.	Two languages are accepted. *Controlled List Values:* EN (English) DE (German)
5	**Resource IDs**				
5.1	ID	R	0-n	Any identifiers associated with the resource. This can be an identifier or identifiers of the related resource(s), contributor(s), institution(s) or an alternate identifier of the given resource provided by another registering system.	The format is a free text and depends on the ID Type.
5.2	ID_type	R	0-n	The type of the ID.	Four main groups of ID types are available: (*) IDs usually applied to documents and publications; (**) IDs applied to persons or institutions; (***) IDs applied to clinical trials which were registered in some clinical trials register(s); (****) any other related identifier (e.g., grant number). Select appropriate values from the list. *Controlled List Values:* (*) arXiv, bibcode, DOI, EAN13, EISSN, ISBN, ISSN, ISTC, LISSN, LSID, PMID, PURL, URL, URN; (**) ORCID, ISNI, ROR, GRID; (***) DRKS, UTN, ISRCTN, EudraCT, EUDAMED, NCT (clinicaltrials.gov); (****) Other
5.3	ID_date	R	0-n	If known, date of ID assignment.	Please, enter the date in the following format: DD.MM.YYYY

5.4	relationType	R	0-n	Description of the relationship of the resource being registered and a related resource. If an alternate identifier is being registered, select "HasAlternateID" relation type.	Select an appropriate value from the list. *Controlled List Values:* IsCitedBy, Cites, IsSupplementTo, IsSupplementedBy, IsContinuedBy, Continues, IsDescribedBy, Describes, HasMetadata, IsMetadataFor, HasVersion, IsVersionOf, IsNewVersionOf, IsPreviousVersionOf, IsPartOf, HasPart, IsReferencedBy, References, IsDocumentedBy, Documents, IsCompiledBy, Compiles, IsVariantFormOf, IsOriginalFormOf, IsIdenticalTo, IsReviewedBy, Reviews, IsDerivedFrom, IsSourceOf, IsRequiredBy, Requires, IsObsoletedBy, Obsoletes, HasGrantNumber, HasAlternateID For definitions, examples and usage notes, see the latest version of the DataCite Metadata Schema at https://schema.datacite.org/.
6	**Roles**				
6.1	role_type	M	1-n	The type of the role, i.e. the type of the resource contributor. For each resource, there may be multiple contributors, but at least one contributor must be specified.	Select appropriate values from the list. *Controlled List Values:* Contributor, Contact Person, Data Collector, Data Curator, Data Manager, Distributor, Editor, Hosting Institution, Producer, Project Leader, Project Manager, Project Member, Registration Agency, Registration Authority, Related Person, Researcher, Research Group, Rights Holder, Supervisor, Work Package Leader, Sponsor, Sponsor-Investigator, Funder, Publisher, Other For definitions, see the latest version of the DataCite Metadata Schema at https://schema.datacite.org/. Please note that the following terms slightly differ from DataCite definitions: "Sponsor" is an organization or person taking responsibility for and initiating a study.

					"Sponsor-Investigator" is the person both initiating and conducting the study. "Funder" is an organization or person providing financial support for a study. "Project Leader" is equivalent to the Principal Investigator of a study, i.e. the person both initiating and conducting the study
6.2	role_name	M	1-n	Name associated with the role. This may be a personal or an institutional name.	For personal names, format should be: family name, given name.
6.3	role_address	R	0-n	If applicable, address associated with the role.	The address includes the institution/affiliation name and the elements of a postal address, i.e. street, house number, city, postal code, state/province (if applicable), and country. *Example:* ZB MED – Information Centre for Life Sciences Gleueler Str. 60 50931 Cologne, Germany
6.4	role_email	R	0-n	If applicable, email address associated with the role.	-
6.5	role_phone	R	0-n	If applicable, phone number associated with the role.	Use the International ITU-T E.164-number structure for geographic areas (https://www.itu.int/rec/T-REC-E.164-201011-I). *Example:* +49 (0)153 1112233
6.6	role_web_page	R	0-n	If applicable, web page associated with the role.	Format: URL starting with http://
6.7	role_specific_type	R	0-n	The classification of the role types "Sponsor" and "Funder" into a predefined list of specific role types.	For the role type "Sponsor" *Controlled List Values:* Primary, Secondary For role type "Funder" *Controlled List Values:* Public, Private
7	**Study Design**				
7.1	study_primary_design	M	1	The primary design of a study. Select between interventional and non-interventional study design.	Select one value from the list. *Controlled List Values:* Interventional Non-interventional Definitions: *Non-interventional study* refers to a study that does not aim to alter study outcomes of interest. *Interventional study* refers to a study that aims to alter study outcomes of interest.
7.2	study_type	R	0-1	In case of an interventional study, the strategy for assigning interventions to participants.	Select one value from the list. *Controlled List Values:*

				In case of a non-interventional (observational) study, the primary strategy for participant identification and follow-up.	*Interventional study types (models):* Single Group Parallel Crossover Factorial Sequential Other *Observational study types (models):* Cohort Case-Control Case-Only Case-Crossover Ecologic or Community Studies Family-Based Other For definitions, see the ClinicalTrials.gov Protocol Registration Data Element Definitions at https://prsinfo.clinicaltrials.gov/definitions.html (Section "7. Study Design" -> "Interventional Study Model" and "Observational Study Model")
7.3	study_hypothesis	R	0-1	A statement of the hypotheses underlying the study.	Format: free text.
7.4	study_analysis_unit	R	0-1	Primary unit of analysis.	Select one value from the list. *Controlled List Values:* Individual, Organization, Family, Household, Event/Process, Geographic Unit, Time Unit, Text Unit, Group, Object, Pathogens, Twins, Other For definitions, see the DDI Controlled Vocabulary: https://ddialliance.org/Specification/DDI-CV/AnalysisUnit_1.0.0.html
7.5	study_status	R	0-1	Overall recruitment status for the study.	Select one value from the list. *Controlled List Values:* Not yet recruiting, Recruiting, Enrolling by invitation, Active, not recruiting, Completed, Suspended, Terminated, Withdrawn For definitions, see the ClinicalTrials.gov Protocol Registration Data Element Definitions at

					https://prsinfo.clinicaltrials.gov/definitions.html (Section "11. Contacts, Locations, and Investigator Information" -> "Individual Site Status")
7.6	study_sampling	R	0-1	Applied sampling method.	Select one value from the list. *Controlled List Values:* TotalUniverseCompleteEnumeration, Probability, Probability.SimpleRandom, Probability.SystematicRandom, Probability.Stratified, Probability.Stratified.Proportional, Probability.Stratified.Disproportional, Probability.Cluster, Probability.Cluster.SimpleRandom, Probability.Cluster.StratifiedRandom, Probability.Multistage, Nonprobability, Nonprobability.Availability, Nonprobability.Purposive, Nonprobability.Quota, Nonprobability.RespondentAssisted, MixedProbabilityNonprobability, Other For definitions, see the DDI Controlled Vocabulary: https://ddialliance.org/Specification/DDI-CV/SamplingProcedure_1.1.html
7.7	study_country	R	0-1	Country or countries in which a study takes place.	For country names, see the ISO 3166-1 list (https://www.iso.org/iso-3166-country-codes.html). If multiple countries are entered, they should be semicolon-separated.
7.8	study_region	R	0-1	If applicable, region(s) and/or cities within a country, in which the study takes place.	Format: free text.
7.9	study_conditions	R	0-1	Primary health condition(s) or problem(s) studied.	Format: free text.
7.10	study_population	R	0-1	Detailed description of the study population. Any information important for the given study can be specified here.	Format: free text.
7.11	study_target_sample_size	R	0-1	Intended number of observational units of a study (e.g., intended number of study participants).	Format: an integer value.
7.12	study_obtained_sample_size	R	0-1	Obtained number of observational units of a study (e.g., obtained number of study participants).	Format: an integer value. The values are only available after end of recruitment. Enter "-1", if obtained sample size is unknown.

7.13	study_age_min	R	0-1	Minimum age of participants for inclusion provided in years. If minimum age is measured not in years, specify the relevant unit of time (e.g., hours, days, weeks, etc.) in the "study_population" field.	Format: a float value. Enter "-1" if no limit is applicable.
7.14	study_age_max	R	0-1	Maximum age of participants for inclusion provided in years. If maximum age is measured not in years, specify the relevant unit of time (e.g., hours, days, weeks, etc.) in the "study_population" field.	Format: a float value. Enter "-1" if no limit is applicable.
7.15	study_eligibility	M	1	Eligibility criteria for the study, if possible organized by inclusion and exclusion criteria.	Format: free text
7.16	study_start_date	R	0-1	Start date of data collection for the study. In case of a planned study, it is the intended start date; in case of an ongoing study — the actual start date.	Please, enter the date in the following format: DD.MM.YYYY
7.17	study_end_date	R	0-1	End date of data collection for the study. In case of a planned or ongoing study, it is the intended end date; in case of a completed study — the actual end date.	Please, enter the date in the following format: DD.MM.YYYY
7.18	study_datasource	R	0-n	Sources from which the data are generated or extracted.	Multiple values can be selected from the list. *Multi-select List Values:* Questionnaires, Physical measures, Blood, Urine, Saliva, Cord blood, Tissues, Buccal cells, Hair, Nail, Cognitive measures, Administrative databases, Other For definitions, see the Maelstrom Inventory at https://www.maelstrom-research.org/maelstrom-catalogue (select "Individual Studies" -> "Individual study properties" -> "Data sources" and "Data sources - Biosamples")

7.19	study_design_comment	R	0-1	Any information on the specific study design aspects that cannot be captured by other fields.	Format: free text.
8	**Interventional Study Design**				
8.1	study_interventions	M	1	Specification of the intervention(s) associated with each arm or group; at least one intervention must be specified.	Format: free text. If possible, please indicate the type and the name of intervention(s). For definition of intervention types, see the ClinicalTrials.gov Protocol Registration Data Element Definitions at https://prsinfo.clinicaltrials.gov/definitions.html (Section "8. Arms, Groups, and Interventions" -> "Interventions" -> "Intervention Type")
8.2	study_primary_outcomes	R	0-1	Description of each primary outcome measure.	Format: free text.
8.3	study_secondary_outcomes	R	0-1	Description of each secondary outcome measure.	Format: free text.
8.4	study_phase	R	0-1	If applicable, numerical phase of a study.	Select one value from the list. *Controlled List Values:* N/A, Preclinical, Early-phase-1, Phase-1, Phase-1-phase-2, Phase-2, Phase-2-phase-3, Phase-3, Phase-4, Other For definitions, see the ClinicalTrials.gov Protocol Registration Data Element Definitions at https://prsinfo.clinicaltrials.gov/definitions.html (Section "7. Study Design" -> "Study Phase")

Contacts

For questions, please contact contact@nfdi4health.de.

Acknowledgement

This work was done as part of the NFDI4Health Task Force COVID-19 (www.nfdi4health.de). We gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 451265285.