Overview full metadata model

Name metadata field	Description	Level of necessity	Tips
	Note: All metadata fields need to be filled out in English.	• *Mandatory = necessary for curation flow and publishing* • *Recommended = advised to fill out, but will not block publishing* • *Optional = your choice*	
Citation metadata			
Title	Enter the main title of your dataset.	Mandatory	If your dataset is linked to a publication, you may click *"Replication data for"* below the title field and add the title of your publication.
Subtitle	A secondary title that amplifies or states certain limitations on the main title.	Optional	
Alternative title	A title that is commonly used to refer to the dataset or an abbreviation of the title. This field can also be used to add a title in another language.	Optional	By clicking "+", you can add multiple titles.
Alternative URL	Another URL where the data of the dataset can be viewed or accessed, such as a project or personal webpage. An example can be GitHub/GitLab.	Optional	
Other identifier *(consists of 2 subfields)*	Another unique identifier for the dataset, such as a producer's database number or another repository's identifier (e.g. accession number in a sequence database).		By clicking "+", you can add multiple identifiers.

Agency	The name of the producer or other repository that generated the other identifier.	Optional	
Identifier	Another identifier that uniquely identifies the dataset.	Optional	
Author *(consists of 4 subfields)*	The person(s) or organization(s) responsible for creating the data of the dataset. Note that the author names are added to the bibliographic citation in order of entry.		The name and affiliation associated with the account creating the dataset draft will be automatically included as the first Author of the dataset. By clicking "+", you can add co-authors.
Name	The author's last name, first name or the name of an organization responsible for creating this dataset.	Mandatory	Use the search function to find (co-) authors with an ORCID. In case they are missing an ORCID, you can add them manually (last name, first name).
Affiliation	The organization with which the author is affiliated.	Mandatory by FV	Use the search function to find the organization with a ROR-id. In case the ROR-id is missing, you can add the affiliation manually. Currently, it is not possible to add multiple affiliations for the same author.
Identifier type	The type of identifier that uniquely identifies the author (e.g., ORCID, ISNI)	Optional	This field only appears if you do not use the automated ORCID method listed above.
Identifier	Uniquely identifies the author when paired with an Identifier Type	Optional	This field is only appears if you do not use the automated ORCID method listed above.
Point of contact *(consists of 3 subfields)*	The person(s) or organization(s) that can be contacted by data users when questions arise. Note that we advise to add (a) person(s) with a more permanent function (e.g. promotor).		By clicking "+", you can add multiple points of contact.

Name	The contact's last name, first name or the name of an organization.	Mandatory	The name of the dataset creator in FAIRVault will automatically be added, but this can be changed manually. Best practice is to add the name of a senior scientist (e.g. the promotor).
Affiliation	The organization with which the point of contact is affiliated.	Mandatory	The affiliation of the dataset creator in FAIRVault will automatically be added, but this can be changed manually. Best practice is to add the affiliation of a senior scientist (e.g. the promotor).
E-mail	The email address of the point of contact. This email address will NOT be displayed publicly. A message to the point of contact can only be sent via the "contact owner" button once the dataset is published.	Mandatory	The email address of the dataset creator in FAIRVault will automatically be added, but this can be changed manually. Best practice is to add the email address of a senior scientist (e.g. the promotor).
Description *(consists of 2 subfields)*	A short description that summarizes the dataset itself. You can mention what the data include, how they were obtained and what they can be used for. Avoid copying the abstract of your publication.		By clicking "+", you can add more descriptions.
Text	A summary describing the purpose, nature, and scope of the dataset.	Mandatory	Avoid using special characters and html tags.
Date	The date when the description was added to the dataset. If the dataset contains more than one description, e.g. the data producer supplied one description and the data repository supplied another, this	Recommended	Add the date following the ISO 8601 standard, i.e. YYYY-MM-DD.

	date is used to distinguish between the descriptions.		
Subject	The area of study relevant to the dataset.	Mandatory	Use the drop-down list to choose the subject(s) that best fit(s) your dataset. You can choose multiple. In case none of the subjects fit, choose "other".
Keyword (consists of 4 subfields)	Key terms describing important aspects of the dataset. These keywords will improve the findability of your dataset.		Each keyword needs to be entered separately by clicking "+".
Term	Key terms describing important aspects of the dataset.	Mandatory	Preferably choose keywords from a controlled vocabulary. For example, for the term "Plasmodium vivax", you could provide a link to the specific term in the term URI field (https://www.ncbi.nlm.nih.gov/mesh/68010966), the name "MeSH" in the Controlled Vocabulary Name field, and a link to the MeSH homepage in the Controlled Vocabulary URL field (https://www.nlm.nih.gov/mesh/).
Term URI	A URI (Uniform Resource Identifier) that points to the web presence of the keyword term.	Optional	For example, for the term "Plasmodium vivax", you could provide a link to the specific term in the term URI field (https://www.ncbi.nlm.nih.gov/mesh/68010966).
Controlled Vocabulary Name	The controlled vocabulary used for the keyword term, such as LCSH, MeSH or others.	Recommended	
Controlled Vocabulary URL	The URL where one can access information about the term's controlled vocabulary.	Optional	Enter the URL of the keyword vocabulary website (in general; e.g. https://www.nlm.nih.gov/mesh/).
Topic Classification (consists of 3 subfields)	Indicates a broad, important topic or subject that the dataset covers.		By clicking "+", you can add topic classifications.
Term	Indicates a broad, important topic or subject that the dataset covers.	Optional	

Controlled Vocabulary Name	The controlled vocabulary used for the topic classification term, such as LCSH, MeSH or others.	Optional	
Controlled Vocabulary URL	The URL where one can access information about the term's controlled vocabulary.	Optional	
Related publication *(consists of 5 subfields)*	Publications that use the data of this dataset.		By clicking "+", you can add multiple related publications.
Relation type	The nature of the relationship between this dataset and the related publication.	Mandatory *(if applicable)*	Most frequently, for a dataset supporting a research article/publication, use the "Is Supplement To" relation type.
Citation	The full bibliographic citation for the related publication.	Mandatory *(if applicable)*	In case your publication is submitted to the journal but not yet accepted, add "submitted for review".
Identifier Type	The type of identifier that uniquely identifies a related publication (e.g. DOI = Digital Object Identifier).	Mandatory *(if applicable)*	Choose from the drop-down list.
Identifier	The identifier for a related publication.	Mandatory *(if applicable)*	
URL	Link to the website where the publication is available (e.g. journal article page).	Recommended *(if applicable)*	
Notes	Additional important information about the dataset.	Optional	
Language	The language that the dataset's files are written in.	Recommended	Choose from the drop-down list.
Producer *(consists of 5 subfields)*	Person(s) or organization(s) managing the finances or other administrative processes involved in the creation of the dataset.		By clicking "+", you can add multiple producers.

Name	The producer's last name, first name or the name of an organization.	Optional	
Affiliation	The organization with which the producer is affiliated.	Optional	
Abbreviated Name	The producer's abbreviated name.	Optional	
URL	The URL of the producer's website.	Optional	
Logo URL	The URL of the producer's logo.	Optional	
Production Date	The date when the data were produced (not distributed, published, or archived).	Optional	Add the date following the ISO 8601 standard, i.e. YYYY-MM-DD.
Production Location	The location where the data and any related materials were produced or collected.	Optional	By clicking "+", you can add multiple locations.
Contributor *(consists of 2 subfields)*	The person(s) or organization(s) responsible for collecting, managing, or otherwise contributing to the development of the dataset.		By clicking "+", you can add multiple contributors. Note that contributors do not appear in the bibliographic citation.
Type	Indicates the type of contribution made to the dataset.	Recommended *(if applicable)*	Choose from the drop-down list.
Name	The contributor's last name, first name or the name of an organization.	Recommended *(if applicable)*	
Funding Information *(consists of 2 subfields)*	Information about the dataset's financial support.		By clicking "+", you can add multiple funders.
Agency	The agency that provided financial support for the dataset.	Mandatory *(if applicable)*	Use the search function to find the funding agency with a ROR-id. In case the ROR-id is missing, you can add the funding agency manually.

Identifier	The grant identifier or contract identifier of the agency that provided financial support for the dataset.	Mandatory *(if applicable)*	
Distributor *(consists of 5 subfields)*	Person(s) or organization(s) designated to generate copies of the dataset, including any editions or revisions.		By clicking "+", you can add multiple distributors.
Name	The distributor's last name, first name or the name of an organization.	Optional	
Affiliation	The organization with which the point of contact is affiliated.	Optional	
Abbreviated Name	The distributor's abbreviated name.	Optional	
URL	The URL of the distributor's webpage.	Optional	
Logo URL	The URL of the distributor's logo image, used to show the image on the dataset's page.	Optional	
Distribution Date	The date when the dataset was made available for distribution/presentation.	Optional	Add the date following the ISO 8601 standard, i.e. YYYY-MM-DD.
Depositor	Person(s) or organization(s) that deposited the dataset in the repository.	Recommended	The name of the dataset creator in FAIRVault will automatically be added. We recommend not to manually change this field.
Deposit Date	The date when the dataset was deposited into the repository.	Recommended	The deposit date field is prefilled in the metadata form when clicking to add a new dataset. The date on which you click to add a new dataset is taken as the prefilled deposit date. Add the date following the ISO 8601 standard, i.e. YYYY-MM-DD.
Time period *(consists of 2 subfields)*	The time period that the data refer to. Also known as span. This is the time		By clicking "+", you can add multiple time periods.

	period covered by the data, not the dates of coding, collecting data, or making documents machine-readable.		
Start Date	The start date of the time period that the data refer to.	Optional	Add the date following the ISO 8601 standard, i.e. YYYY-MM-DD.
End date	The end date of the time period that the data refer to.	Optional	Add the date following the ISO 8601 standard, i.e. YYYY-MM-DD.
Date of Collection *(consists of 2 subfields)*	The dates when the data were collected or generated.		By clicking "+", you can add multiple collection periods.
Start Date	The date when the data collection started.	Optional	Add the date following the ISO 8601 standard, i.e. YYYY-MM-DD.
End date	The date when the data collection ended.	Optional	Add the date following the ISO 8601 standard, i.e. YYYY-MM-DD.
Data Type	The type of data included in the files.	Recommended	By clicking "+", you can add multiple data types. Choose your data type from the drop down menu (compiled data, simulation data, experimental data, aggregated data, survey data, observational data, recorded data, measurement and test data, clinical trial data, genomic data, laboratory notebook, encoded data or geospatial data).
Series *(consists of 2 subfields)*	Information about the dataset series to which the dataset belong.		By clicking "+", you can add multiple series.
Name	The name of the dataset series.	Optional	
Information	This can include a history of the series and/or a summary of features that apply to the series.	Optional	
Software *(consists of 2 subfields)*	Information about the software used to generate the dataset.		By clicking "+", you can add multiple software.
Name	The name of software used to generate the dataset.	Optional	

Version	The version of the software used to generate the dataset.	Optional	
Related Material	Information, such as a persistent ID or citation, about the material related to the Dataset, such as appendices or sampling information available outside of the dataset.	Optional	By clicking "+", you can add multiple related materials.
Related Dataset	Information, such as a persistent ID or citation, about a related dataset, such as previous research on the dataset's subject.	Recommended *(if applicable)*	By clicking "+", you can add multiple related datasets.
Other Reference	Information, such as a persistent ID or citation, about another type of resource that provides background or supporting material to the dataset.	Optional	By clicking "+", you can add multiple references.
Data Source	Information, such as a persistent ID or citation, about sources of the dataset (e.g. a book, article, serial, or machine-readable data file).	Optional	By clicking "+", you can add multiple data sources.
Origin of Historical Sources	For historical sources, the origin and any rules followed in establishing them as sources.	Optional	
Characteristics of Sources	Characteristics not already noted elsewhere.	Optional	
Documentation and Access to Sources	Methods or procedures for accessing data sources and/or any special permissions needed for access.	Optional	
FAIRVault Decision Tree			
Ethical Concerns?	Are there ethical concerns regarding disclosure of the data?	Mandatory	Choose from the drop-down list.

			Choose *'some ethical concerns'* if disclosing the data could raise ethical concerns — for example, if it risks harming individuals or groups, involves sensitive personal or cultural information, includes data from vulnerable populations, lacks proper informed consent for sharing, or could be misused in ways that cause discrimination, stigmatization, or other negative consequences. Ethical concerns also apply if disclosure would violate legal, institutional, or community agreements.
3rd Party Confidentiality?	Is the dataset subject to confidentiality obligations towards third parties?	Mandatory	Choose from the drop-down list. Choose *'some confidentiality concerns'* if the dataset contains information that is subject to confidentiality agreements with third parties — for example, data provided by external organizations, companies, or collaborators under a contract containing a confidentiality clause or a non-disclosure agreement (NDA).
Valorisation Potential?	Is there a valorisation potential related to the data?	Mandatory	Choose from the drop-down list. Choose *'some valorisation potential'* if the data has potential commercial value — for example, if it could be used for developing products, services, patents, or innovations, or if it holds relevance for industry partners, startups, or other external stakeholders.
Dual/Military Use?	Can the data be misused or is there any dual or military use potential?	Mandatory	Choose from the drop-down list. Choose *'potential for military application'* if the dataset (research results or techniques) may lend themselves to military purposes. Choose *'potential for dual use'* if the dataset (research results or materials) could be applied for both civil and military purposes, even if not intended for military use.

			Choose *'potential for misuse'* if the dataset (research results or techniques) could be used for terrorist or criminal activities or in applications that violate human rights or compromise the safety of people, animals or the environment. Choose *''concerns related to research security'* if the dataset risks undesired knowledge or technology transfer, foreign interference, or misuse of research results that could compromise academic freedom, research integrity, or lead to unethical applications, including military or criminal use.
Data Reuse? *(consists of 3 subfields)*			
Secondary Data?	Does the dataset contain secondary data?	Mandatory	Choose from the drop-down list. Choose *'yes, the data are copyrighted'* if the dataset includes secondary data — meaning data originally collected or generated by someone else.
Copyright?	Are the secondary data protected by third party copyright?	Mandatory *(in case you answered 'yes' to 'secondary data?')*	Choose from the drop-down list. Choose *'yes, the data are copyrighted'* if the dataset contains material protected by third-party copyright.
License?	Is there a clear license allowing you to openly share the material (e.g. CC-BY)?	Mandatory *(in case you answered 'yes' to 'secondary data?')*	Choose from the drop-down list.
Personal data *(consists of 4 subfields)*	Following Art. 4(1) of GDPR, personal data are any information which are related to an identified or identifiable natural person. These can include direct (e.g. name, identification number) and/or indirect identifiers (e.g. gender, date of birth).		
Does the dataset contain personal data?	Following Art. 4(1) of GDPR, personal data are any information which are related to	Mandatory	Choose from the drop-down list.

	an identified or identifiable natural person. These can include direct (e.g. name, identification number) and/or indirect identifiers (e.g. gender, date of birth).		
Anonymization/ Pseudonymization?	To what extent has your dataset been anonymized/pseudonymized?	Mandatory *(in case you answered 'yes' to 'does the dataset containpersonal data?')*	Choose from the drop-down list. • Anonymous data is (following GDPR) "information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable". • Pseudonymisation is defined in GDPR as "'the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person".
Consent?	Have your participants given explicit consent to have their data shared in its current form?	Mandatory *(in case you answered 'yes' to 'does the dataset contain personal data?')*	Choose from the drop-down list.
Special Categories?	Does the dataset contain special categories of personal data/vulnerable people?	Mandatory *(in case you answered 'yes' to 'does the dataset contain personal data?')*	Choose from the drop-down list. Special categories of personal data = Race, political views, religious or philosophical beliefs, trade union membership, genetic data, biometric data, health data, data on a person's sexual behavior or sexual orientation and data relating to criminal convictions or offences.

Access conditions

Access conditions *(consists of 3 subfields)*			
Sensitivity Level	Data Tags (originally developed at Harvard University) are a tool to quickly assess the sensitivity level of a dataset in a standardized way that is easy to understand for both humans and machines. See guide for more information.	Mandatory	The data curator of the FAIRVault will fill this out based on the answers from the FAIRVault Decision Tree metadata block. Choose from the drop-down list. - Blue: Non-confidential and non-sensitive data - Green: Non-confidential and non-sensitive data with some access control - Orange: Confidential and sensitive data - Red: Very confidential and sensitive data
Embargo until	End date of data embargo period. This is the date after which the data will become accessible (either openly or via access requests).	Mandatory by FV *(if applicable)*	Add the date following the ISO 8601 standard, i.e. YYYY-MM-DD.
Reason for restriction	Reason why access to (some) files in the dataset need to be restricted.	Mandatory by FV *(if applicable)*	Choose from the drop-down list. Multiple answers can be selected.