

## ABC metadata model

Amsterdam UMC Biomedical Concise metadata model

Version 1.0, 31-03-2023

The ABC metadata model (Amsterdam UMC Biomedical Concise metadata model) is a **model** that provides guidance for **minimal** metadata implementation in the field of **biomedicine**.

Being a **model** implies that it isn't a ready-made solution, but it offers guidance to what metadata is needed for FAIR data management. It is deliberately kept **minimal**, thus facilitating data exchange between persons and systems, reducing workload for researchers and support alike. It could be considered to be a checklist at the utmost minimum of items that should be reported for FAIR data management. It aims to bridge the gap between generic metadata standards and detailed metadata generated by man and machine in the field of **biomedicine**, covering all disciplines in health and life sciences.

The ambition of the development team is that the ABC metadata model eventually will grow to become a standard by wide adaptation in FAIR RDM practice, collective collaboration in improvement of the model and community endorsement.

The model consist of **three levels of granularity**:

- 1) Data collection: items of a generic nature that completely and fully comply to Dublin Core, DataCite and DCAT metadata standards;
- 2) Data types: generic biomedical data types, not represented in generic metadata standards;
- 3) Data types specification: a more detailed but very concise elaboration per data type of the 2<sup>nd</sup> level of granularity.

For questions or more information please contact [rdm@amsterdamumc.nl](mailto:rdm@amsterdamumc.nl).

© 2022-2023 Joost Daams and Evelien van der Schaaf-de Wolf, Research Data Management, Research Support, Amsterdam UMC.

This work is licensed under a CC BY-NC-SA 4.0 license. <https://creativecommons.org/licenses/by-nc-sa/4.0/>

# ABC metadata model - Data collection level

Item number	Field label	Optiongroup	Definition/clarification	Datatype	Variable name	Required y/n	DataCite property	Dublin Core property	DCAT Class (W3C v2)	DCAT Property (W3C v2)
1.1	Unique Persistent Identifier		A unique persistent identifier is a long-lasting unique reference to a document, file, web page, or other object.	text	identifier	y	Identifier	Identifier	Resource	dct:identifier
1.2	Creator		Person(s) who bear(s) responsibility for the creation of the file/object	text	creator	y	Creator	Creator	Resource	dct:creator
1.3	Creator Unique Identifier		A unique identifier to the entity (person/organisation) primarily responsible for creation, preferably an ORCID. See <a href="https://orcid.org/register">https://orcid.org/register</a> to create an ORCID iD.	text	creator_id	y	nameIdentifierScheme			
1.4	Creator Affiliation		Affiliation of the creator. Example: Amsterdam UMC, University of Amsterdam, department(s), research institute(s), Amsterdam, Netherlands.	text	creator_affiliation	y	affiliation of affiliationIdentifier			
1.5	License		A legal document giving official permission to do something with the resource	text	licence	y	Rights	Rights	Resource	dct:license
1.6	Title		A name given to the resource	text	title	y	Title	Title	Resource	dct:title
1.7	Topic		Subject of the resource	text	topic	y	Subject	Subject	Resource	dcat:keyword
1.8	Year of Creation		The date the resource itself was put together; this could refer to a timeframe in ancient history, a date range, or a single date for a final component, e.g., the finalised file with all the data.	text	year_created	y	Date (DateType = Created)	Date	Dataset	dct:temporal [ a dct:PeriodOfTime ; dcat:startDate ; dcat:endDate ]
1.9	Year of Collection		The date or date range in which the resource content was collected.	text	year_collected	y	Date (DateType = Collected)	Date	Resource	dct:modified
1.10	Year of Update		The date of the last update to the resource, when the resource is being added to. May be a range	text	year_updated	y	Date (DateType = updated)	Date	Resource	dct:modified
1.11	Year of Publication/Issue		Issued = The date that the resource is published or distributed, e.g., to a data centre. (DataCite)	number	year_published	y	PublicationYear		Resource	dct:issued
1.12	Version Number		The version number	text	version	n	Version			
1.13	Stored Format		Technical format of the resource (DataCite)	text	format	n	Format	Format	Distribution	dct:format
1.14	Stored Size		Size (e.g., bytes, pages, inches, etc.) or duration (extent), e.g., hours, minutes, days, etc., of a resource (DataCite)	text	size	n	Size		Distribution	dcat:byteSize
1.15	Language	English, Dutch, Other	The primary language of the resource (DataCite)	dropdown	language	n	Language	Language	Resource	dct:language
1.16	Funder		Information about financial support (funding) for the resource being registered	text	funder	n	FunderName			
1.17	Publisher		The name of the entity that holds, archives, publishes, prints, distributes, releases, issues, or produces the resource. (DataCite)	text	publisher	y	Publisher	Publisher	Resource	dct:publisher
1.18	Contributor		The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource. (DataCite)	text	contributor	y	Contributor	Contributor		
1.19	Contact Details		E-mail address of the individual or organization holding more information about the data set or locating the person responsible.	text	contact	y			Resource	dcat:contactPoint

# ABC metadata model - Data types level

Section	Item number	Field label	Optiongroup	Definition/clarification	HasDependentField	Dependencies	Datatype	Variable name	Required y/n
Data types	2.1	Data Type	i Biological samples, ii Genetic / omics data, iii Imaging data, iv Qualitative data and quantitative questionnaires, v Physiological data, vi Synthetic data		y		checkbox	datatype	n
	2.2	Subject Data			n				
	2.2.1	Subject Type	Human, Animal		n		checkbox	subject_type	n
	2.2.2	Number of Subjects			n		number	subject_amount	n
	2.2.3	Subject Health Condition			n		text	subject_health	n
	2.2.4	Laboratory Data	Yes, No		n		dropdown	lab_data	n
Data types specification	3.1				n				
	[header	Biological Samples			n				
	3.1.1	Biological Sample Type	DNA, RNA, Serum, Plasma, Urine, Faeces		n	datatype=i Biological samples	checkbox	sample_type	n
	3.1.2	Storage Temperature			n	datatype=i Biological samples	number	storage_temp	n
	3.2				n				
	[header	Genetic / Omics Data			n				
	3.2.1	Genetic / Omics Data Type	Genomics, Metabolomics, Proteomics, Transcriptomics, RNASeq		n	datatype=ii Genetic / omics data	checkbox	omics_datatype	n
	3.2.2	Primary Data	Yes, No		y	datatype=ii Genetic / omics data	dropdown	primary_data	n
	3.2.3	Primary Data Location			n	primary_data=Yes	text	primary_data_location	n
	3.2.4	Secondary Data	Yes, No		y	datatype=ii Genetic / omics data	dropdown	secondary_data	n
	3.2.5	Secondary Data Location			n	secondary_data=Yes	text	secondary_data_location	n
	3.3				n				
	[header	Imaging Data			n				
	3.3.1	Imaging Type	Diagnostic imaging, Interventional radiology		n	datatype=iii Imaging data	dropdown	imaging_type	n
	3.3.2	Imaging Modality	X-Ray, ECHO, CT, CTA, MRI, fMRI, PET, PET-CT, PET-MRI, SPECT, CBCT, Fluoroscopy		n	datatype=iii Imaging data	checkbox	imaging_modality	n
	3.3.3	Imaging Description			n	datatype=iii Imaging data	text	imaging_description	n
	3.3.4	Imaging Data Location			n	datatype=iii Imaging data	text	imaging_storage_system	n
	3.4	Qualitative Data and Quantitative			n				
	[header	Questionnaires			n				
	3.4.1	Questionnaire Name			n	datatype=iv Qualitative data and quantitative questionnaires	text	questionnaire_name	n
	3.4.2	Questionnaire Version Number			n	datatype=iv Qualitative data and quantitative questionnaires	text	questionnaire_version	n
	3.4.3	Protocol Name Video Interview			n	datatype=iv Qualitative data and quantitative questionnaires	text	video_interview_name	n
	3.5				n				
	[header	Physiological Data			n				
				Physical data: e.g., length, weight, heart rate, blood pressure, saturation. Biochemical data: e.g., drugs, bacterial cultures, protein levels. Machine readable biosignals: e.g., ECG, EEG.					
	3.5.1	Physiological Data Type	Physical data, Biochemical data, Machine readable biosignals		n	datatype=v Physiological data	checkbox	physiological_data_type	n
	3.5.2	Data Collection Method	Self-measured, Measured by health care provider, Measured by someone else		n	datatype=v Physiological data	checkbox	physiological_data_collection	n
	3.6				n				
	[header	Synthetic Data			n	datatype=vi Synthetic data			