				Colour Legend	
Final version after processing the results of discussion in the working group				**Colour Legend**	
This will become a collection of all tables; after reviewing per item				**Row colours**	**mandatory/recommended/optional column**
Review the sheet Sunflower_Sample_Review				Confirmed elements	mandatory and mandatory if applicable
Please leave comments including your name and date				Newly added elements - for review	recommended
If done with review (even if no remarks/comments) add your name below under Reviewers of this version					
					optional
Note: The HealthDCAT-AP column will be mostly complete for the "metadata submission class (new class name for Submitter) since that class is all about metadata during submission of a dataset. For other classes we only provide an entry in this column if information of that specific concept can be harvested/used to provide metadata during submission.					conditional
Note on mandatory fields with lists (but not strings): also HL7 NULL flavors will be allowed					

Version history; Items added per version

Items	Date version release
Submitter, Sample	2024-09-04

Decision(s):	Date decision	Motivation/Information:
Submitter should not be part of the minimal dataset. Submitter is part of metadata. Within UML-diagram this will be explicitly visualised	2024-09-18	Submitter of a dataset within GDI should not be part of the minimal dataset per se, it is the metadata provided to make clear who submits a dataset within GDI.
The class "Submitter" was renamed to "Metadata Submission" to make it clearer it is a class concerning metadata.	2024-10-02	The name of the class Submitter was confusing and wasn't properly representing the class.
New class "Identification" was added. The class contains Identifier, Identifier Type and Name properties.	2024-10-02	Multiple elements about identifiers got feedback that only providing the ID is not sufficient. New class "Identification" was added, where other items can point to.

Authors of latest version:

Name	Date version release
Ana Konrad	2025-01-30
Hannah Neikes	2025-01-30
Jeroen Belien	2025-01-30
Joeri van der Velde	2025-01-30
Abhishek Nayak	2025-01-30

Aedin Culhane	2025-01-30							
Reviewers of latest version								
Name	Date							
Mikael Kronborg	2025-02-10							
Richard Hagan	2025-02-10							
Evita Lindholm	2025-02-10							
Hannele Laivuori	2025-02-11							
Milan Ojsteršek	2025-02-11							
Michela Riba	2025-02-11							
Edel Cahalin	2025-02-10							

Suggested_Union_Domain/class/table	Suggested Union Item	Suggested_Union_Proposed_Definition	Suggested_Union_Proposed_Values	Suggested_Union_Proposed_cardinality	Suggested_union_mandatory/recommended/optional	If conditional, terms of condition stated here	Part of sunflower	Link to terminology/ontology that defines item	Link to terminology/ontology that defines value(s)/valueset(s)	Reasoning/explanation/evidence of/for suggestion	Concept: Exact match with	Concept: close match with	Value list: Exact match with	Value list: Close match with	Example data	Link to terminology/ontology that defines the class	Class: Exact match with	Class: close match with
Sample	Anatomical sample location	Anatomic site from which the sample was taken.	ICD-11 Anatomy and topography	1..1	conditional	In use case cancer, the exact sample location is relevant; therefore the more detailed value list of ICD-O3 can be used (which is included in ICD-11)	Core	https://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C13717	ICD-11 for Mortality and Morbidity Statistics			OncoTree		SNOMED Anatomical structure				
Sample	Pathological state	The pathological condition of the sample.	Tissue Normal, Germline Normal, Primary Tumor, Tumor Metastasis, Recurrent Tumor, (Xenograft), (unknown)	1..1	conditional	In cancer use case, the pathological state of the sample must be recorded; several instances of the sample class can be filled if matching samples of different pathological states (normal, primary tumor, ...) are present	Core	NCIT Tissue Sample	Tissue Normal, Germline Normal, Primary Tumor, Tumor Metastasis, Recurrent Tumor, (Xenograft), (unknown)									
Sample	Date	Defines the date of sampling	Date, ISO 8601 format, YYYY-MM-DD	0..1	recommended		Core	https://bioportal.bioontology.org/ontologies/NCIT?p=classes&conceptid=http%3A%2F%2Fncicb.nci.nih.gov%2Fxml%2Fowl%2FEVS%2FThesaurus.owl%23C25164&jump_to_nav=true	https://www.iso.org/iso-8601-date-and-time-format.html									
Sample	ID	Unique identifier for a collected specimen assigned by data provider.	String	0..1	recommended		Core		NA									
Sample	Organism	A living entity	List of siblings/children of https://bioportal.bioontology.org/ontologies/NCIT?p=classes&conceptid=http%3A%2F%2Fncicb.nci.nih.gov%2Fxml%2Fowl%2FEVS%2FThesaurus.owl%23C14250	1..1	mandatory	If Organism is for example "Virus" then Biospecimen_Type should not be something like "tissue"	Core	https://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C14250 or https://bioportal.bioontology.org/ontologies/NCIT?p=classes&conceptid=http%3A%2F%2Fncicb.nci.nih.gov%2Fxml%2Fowl%2FEVS%2FThesaurus.owl%23C14250	List of children/siblings of https://bioportal.bioontology.org/ontologies/NCIT?p=classes&conceptid=http%3A%2F%2Fncicb.nci.nih.gov%2Fxml%2Fowl%2FEVS%2FThesaurus.owl%23C14250									
Sample	Biospecimen_Type	The type of a material sample taken from a biological entity for testing, diagnostic, propagation, treatment or research purposes. This includes particular types of cellular molecules, cells, tissues, organs, body fluids, embryos, and body excretory substances.	SPREC codes	1..1	conditional	For Cancer use case, this field is mandatory, otherwise optional.	Core / Cancer	http://purl.obolibrary.org/obo/NCIT_C70713	https://cdn.cmoase.com/wee.obar.org/resource/wsmp/obar_20160pdf/standard_preanalytical_code_.pdf						B4S6TS (where B4 = blood tube with EDTA, S6 = centrifugation at 2000g for 10 minutes, TS = storage at -80°C).			
Sample	Extraction_Technique	The technique of extraction of the sample.	Values within MESH Specimen Handling	0..1	recommended		Core											
Sample	Storage_Conditions	Storage conditions of the sample.	https://github.com/fairgenomes/fairgenomes-semantic-model/blob/main/lookups/StorageConditions.txt	0..n	recommended		Core		https://github.com/fairgenomes/fairgenomes-semantic-model/blob/main/lookups/StorageConditions.txt									
Sample	Assayed_Biological_Macromolecule	Macromolecule derived from the sample.	Children of biological macromolecule	0..1	recommended		Core		Children of EFO biological macromolecule				https://bioportal.bioontology.org/ontologies/NCIT?p=classes&conceptid=http%3A%2F%2Fncicb.nci.nih.gov%2Fxml%2Fowl%2FEVS%2FThesaurus.owl%23					
Sample	Sampling_Intent	Describes the purpose for taking the sample.	String	0..n	recommended		Core	https://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&version=24.06&ns=ncit&code=C25726&key=945876985&b=1&n=null	NA						The sample was taken for germline testing.			
Sample_Relation					Sample relation class is not mandatory, but sample ID is mandatory inside this class.													
Sample_Relation	Related_Sample	Points to a related sample.	Sample ID	1..n	mandatory	If relation to other sample is provided, this item describes the type of relationship between the two samples (for example: matched healthy and tumor samples)	Core		NA									
Sample_Relation	Relation_Type	Describes the relationship type between the two connected samples.	List to be defined/extended: New sample, Derived, Aliquot, Control (experiment), Disease versus Control, Matched, Longitudinal, Paired, Familial, Spatial	1..1	mandatory	If relation to other sample is provided, this item describes the type of relationship between the two samples (for example: matched healthy and tumor samples)	Core		Potential value-set, either if it fulfils demands/needs use it, or make a subselection: https://ncit.nci.nih.gov/ncitbrowser/action=create_src_vs_free&xml_url=http://evs.nci.nih.gov/valueset/sas02rpu/C0/12B	In Cancer use case: If normal and tumor samples are present, corresponding to the same specimen (e.g. normal tissue adjacent to tumor tissue, should be matched by the sample date, but also by a direct match					Matched Tumor-Normal: Taking a sample from a tumor and a corresponding sample from the normal surrounding tissue. This helps in comparing the genetic and molecular differences between cancerous and non-cancerous tissues. Longitudinal Samples: These are samples taken from the same individual at different time points. For example, blood samples collected from a patient before, during, and after a treatment to monitor changes over time. Paired Samples: These involve taking two different types of samples from the same individual at the same time. For instance, collecting both blood and urine samples from a patient to study different biomarkers. Familial Samples: Samples collected from related individuals, such as siblings or parent-child pairs. This can be useful in studying hereditary conditions or genetic predispositions. Pre- and Post-Treatment Samples: Samples taken before and after a specific treatment or intervention. For example, tissue biopsies taken before starting chemotherapy and after completing the treatment to assess its effectiveness. Disease vs. Control Samples: Samples from individuals with a specific disease compared to samples from healthy individuals. For instance, comparing blood samples from diabetic patients with those from non-diabetic individuals to identify disease-specific biomarkers. Spatially Related Samples: Samples taken from different locations within the same tissue or organ. For example, multiple biopsies from different regions of a tumor to study intratumoral heterogeneity.			
Sample_Relation	Relation_Description	Free text describing the Relation Type of two samples.	String	0..1	recommended		Core		NA	Suggested by Alfonso: Regarding the samples, given that in many cases more than one sample is collected it will be key to indicate the relation with other samples of the same donor (e.g. different tumors) or from different donors (a family for RDs)								
TNM	Tumor_size	Indicates the size of the primary tumor.	Double	0..1	optional		Cancer	https://bioportal.bioontology.org/ontologies/NCIT?p=classes&conceptid=http%3A%2F%2Fncicb.nci.nih.gov%2Fxml%2Fowl%2FEVS%2FThesaurus.owl%23C25164&jump_to_nav=true	NA									
TNM	Tumor_size_unit	Indicate the unit in which tumor size is stated.	UCUM	0..1	optional		Cancer	https://build.fhir.org/ig/HL7/fhir-mCODE-ig/ValueSet-mcode-tumor-size-units-vs.html	https://ucum.org/									
TNM	Lymph_Node_Status	Indicates lymph node involvement according to the international TNM classification for solid tumors	Children of 1279504008 (American Joint Committee on Cancer pcN category allowable value (qualifier value)	0..1	optional		Cancer	1279504008 (American Joint Committee on Cancer pcN category allowable value (qualifier value)	258828003 (symptomatous nodal involvement (tumor staging))	Values of this field use definitions from AJCC, but via SNOMED to make it machine-readable.								
TNM	Metastasis	Indicate presence or absence of metastasis according to the international TNM classification for solid tumors	Children of American Joint Committee on Cancer clinical M category allowable value (qualifier value) or American Joint Committee on Cancer pathological M category allowable value (qualifier value) and HL7 NULL flavor in case not determined	0..1	optional		Cancer	Children of American Joint Committee on Cancer clinical M category allowable value (qualifier value) or American Joint Committee on Cancer pathological M category allowable value (qualifier value) and HL7 NULL flavor in case not determined	Values of this field use definitions from AJCC, but via SNOMED to make it machine-readable.								
TNM	Version	Version of the TNM classification	String	0..1	optional		Cancer											
TNM	Stage	The extent of a cancer in the body. Staging is usually based on the size of the tumor, whether lymph nodes contain cancer, and whether the cancer has spread from the original site to other parts of the body.	Children of American Joint Committee on Cancer clinical stage group allowable value (qualifier value) or American Joint Committee on Cancer pathological stage group allowable value (qualifier value) or American Joint Committee on Cancer ss stage group allowable value (qualifier value) or UICC values (to be added to SNOMED in 2025)	0..1	optional		Cancer	Children of American Joint Committee on Cancer clinical stage group allowable value (qualifier value) or American Joint Committee on Cancer pathological stage group allowable value (qualifier value) or American Joint Committee on Cancer ss stage group allowable value (qualifier value) or UICC values (to be added to SNOMED in 2025)	Values of this field use definitions from AJCC, but via SNOMED to make it machine-readable. https://staging.seer.cancer.gov/tnm/home/1.9/								
TNM	Pathology_Clinical	Indication whether TNM classification is based on clinical or pathological evaluation. If pathological is available, this always goes before clinical.	Clinical or pathological	1..1	mandatory		Cancer											
TNM	Date of evaluation	Date of the clinical/pathological evaluation.	Date, ISO 8601 format, YYYY-MM-DD	0..1	optional		Cancer	https://www.iso.org/iso-8601-date-and-time-format.html										