FAIR Genomes semantic metadata schema

The FAIR Genomes semantic metadata schema to power reuse of NGS data in research and healthcare. Version 0.3-Minor, 2021-04-23. This model consists of 9 modules that contain 109 metadata elements and 84910 lookups in total (excluding null flavors).

Name	Ontology	Nr. of elements
Study	NCIT:C63536	9
Personal	NCIT:C90492	12
Leaflet and consent form	NCIT:C16468	8
Individual consent	NCIT:C16735	12
Clinical	NCIT:C25398	20
Material	NCIT:C43376	16
Sample preparation	OBI:0001902	9
Sequencing	EDAM:topic_3168	12
Analysis	EDAM:operation_2945	11

Table 1: FAIR Genomes v0.3-Minor overview of all modules.

Name	Ontology	Values
Identifier	OMIABIS:0000006	UniqueID
Name	OMIABIS:0000037	String
Description	OMIABIS:0000036	Text
Inclusion criteria	OBI:0500027	InclusionCriteria lookup (11 choices)
Principal investigator	OMIABIS:0000100	String
Contact information	OMIABIS:0000035	String
Study design	OBI:0500000	Text
Start date	NCIT:C69208	Date
Completion date	NCIT:C142702	Date

Table 2: Module: Study. A detailed examination, analysis, or critical inspection of a subject designed to discover facts about it. Ontology: NCIT:C63536.

Name	Ontology	Values
Personal identifier	NCIT:C164337	UniqueID
Phenotypic sex	PATO:0001894	PhenotypicSex lookup (4 choices)
Genotypic sex	PATO:0020000	GenotypicSex lookup (11 choices)
Country of residence	NCIT:C171105	Countries lookup (249 choices)
Ancestry	HANCESTRO:0004	Ancestry lookup (305 choices)
Country of birth	GENEPIO:0001094	Countries lookup (249 choices)
Year of birth	NCIT:C83164	Integer
Inclusion status	NCIT:C166244	InclusionStatus lookup (4 choices)
Age at death	NCIT:C135383	Integer
Primary affiliated institute	SIO:000688	Institutes lookup (218 choices)
Resources in other institutes	SIO:000688	Institutes lookup (218 choices)
Participates in study	RO:0000056	Reference to Study

Table 3: Module: Personal. Data, facts or figures about an individual; the set of relevant items would depend on the use case. Ontology: NCIT:C90492.

Name	Ontology	Values
Leaflet title	DC:title	String
Leaflet date	DC:date	Date
Leaflet version	DC:hasVersion	String
Consent form identifier	DC:identifier	UniqueID
Consent form accepted date	DC:dateAccepted	Date
Consent form valid until	DC:valid	Date
Consent form creator	SIO:000688	Institutes lookup (218 choices)
Consent form version	DC:hasVersion	String

Table 4: Module: Leaflet and consent form. A document explaining all the relevant information to assist an individual in understanding the expectations and risks in making a decision about a procedure. This document is presented to and signed by the individual or guardian. Ontology: NCIT:C16468.

Name	Ontology	Values
Individual consent identifier	ICO:0000044	UniqueID
Person consenting	IAO:0000136	Reference to Personal
Consent form used	IAO:0000136	Reference to Leaflet and consent form
Collected by	NCIT:C45262	Institutes lookup (218 choices)
Signing date	ICO:0000036	Date
Valid from	DC:valid	Date
Valid until	DC:valid	Date
Represented by	NCIT:C51828	RepresentedBy lookup (2 choices)
Data use permissions	DUO:0000001	DataUsePermissions lookup (5 choices)
Data use modifiers	DUO:0000017	DataUseModifiers lookup (23 choices)
Modifiers specification	SIO:000090	Text
Allow recontacting	NCIT:C25737	Recontacting lookup (3 choices)

Table 5: Module: Individual consent. Consent by a patient to a surgical or medical procedure or participation in a clinical study after achieving an understanding of the relevant medical facts and the risks involved. Ontology: NCIT:C16735.

Name	Ontology	Values
Clinical identifier	NCIT:C87853	UniqueID
Belongs to person	IAO:0000136	Reference to Personal
Phenotype	NCIT:C16977	Phenotypes lookup (15802 choices)
Unobserved phenotype	HL7:C0442737	Phenotypes lookup (15802 choices)
Phenotypic data available	NCIT:C15783	DCMITypes lookup (6 choices)
Clinical diagnosis	NCIT:C15607	Diseases lookup (9700 choices)
Molecular diagnosis gene	NCIT:C20826	Genes lookup (19202 choices)
Molecular diagnosis other	NCIT:C20826	Text
Age at diagnosis	SNOMEDCT:423493009	Integer
Age at last screening	NCIT:C81258	Integer
Medication	NCIT:C459	Drugs lookup (5632 choices)
Drug regimen	NCIT:C142516	Text
Family members affected	HP:0032320	FamilyMembers lookup (41 choices)
Family members sequenced	NCIT:C71384	FamilyMembers lookup (41 choices)
Consanguinity	GSSO:007578	String
Medical history	NCIT:C18772	MedicalHistory lookup (1154 choices)
Age of onset	Orphanet:C023	Integer
First contact	LOINC:MTHU048806	Date
Functioning	NCIT:C21007	Text
Material used in diagnosis	SIO:000641	String

Table 6: Module: Clinical. Relating to the examination and treatment of patients dependent on direct observation. Ontology: NCIT:C25398.

Name	Ontology	Values
Material identifier	NCIT:C93400	UniqueID
Collected from person	SIO:000244	Reference to Personal
Belongs to diagnosis	SIO:000068	Reference to Clinical
Sampling timestamp	EFO:0000689	DateTime
Registration timestamp	NCIT:C25646	DateTime
Sampling protocol	EFO:0005518	Text
Sampling protocol deviation	NCIT:C50996	String
Reason for sampling protocol deviation	NCIT:C93529	String
Biospecimen type	NCIT:C70713	MaterialTypes lookup (13 choices)
Anatomical source	UBERON:0001062	AnatomicalSources lookup (13827 choices)
Pathological state	NCIT:C25687	PathologicalState lookup (4 choices)
Storage conditions	NCIT:C96145	StorageConditions lookup (26 choices)
Expiration date	NCIT:C164516	Date
Percentage tumor cells	NCIT:C127771	Decimal
Physical location	GAZ:00000448	String
Derived from	NCIT:C28355	String

Table 7: Module: Material. Natural substances derived from living organisms such as cells, tissues, proteins, and DNA. Ontology: NCIT:C43376.

Name	Ontology	Values
Sampleprep identifier	NCIT:C132299	UniqueID
Belongs to material	NCIT:C25683	Reference to Material
Input amount	AFRL:0000010	Integer
Library preparation kit	GENEPIO:0000085	NGSKits lookup (615 choices)
PCR free	NCIT:C17003	Boolean
Target enrichment kit	GENEPIO:0000081	NGSKits lookup (615 choices)
UMIs present	EFO:0010199	Boolean
Intended insert size	FG:0000001	Integer
Intended read length	NCIT:C153362	Integer

Table 8: Module: Sample preparation. A sample preparation for assay that preparation of nucleic acids for a sequencing assay. Ontology: OBI:0001902.

Name	Ontology	Values
Sequencing identifier	NCIT:C171337	UniqueID
Belongs to sample	NCIT:C25683	Reference to Sample preparation
Sequencing date	GENEPIO:0000069	Date
Sequencing platform	GENEPIO:0000071	SequencingPlatform lookup (7 choices)
Sequencing instrument model	GENEPIO:0001921	SequencingInstrumentModels lookup (39 choices)
Sequencing type	NCIT:C25284	SequencingTypes lookup (35 choices)
Average read depth	NCIT:C155320	Integer
Observed read length	NCIT:C153362	Integer
Observed insert size	FG:0000002	Integer
Percentage Q30	GENEPIO:0000089	Decimal
Percentage TR20	FG:0000003	Decimal
Other quality metrics	EDAM:data_3914	Text

Table 9: Module: Sequencing. The determination of complete (typically nucleotide) sequences, including those of genomes (full genome sequencing, de novo sequencing and resequencing), amplicons and transcriptomes. Ontology: EDAM:topic_3168.

Name	Ontology	Values
Analysis identifier	AFR:0001979	UniqueID
Belongs to sequencing	NCIT:C25683	Reference to Sequencing
Physical data location	GAZ:00000448	String
Abstract data location	NCIT:C142494	String
Data formats stored	EDAM:format_1915	DataFormats lookup (582 choices)
Algorithms used	NCIT:C16275	Text
Reference genome used	EDAM:data_2340	GenomeAccessions lookup (29 choices)
Bioinformatic protocol used	EDAM:data_2531	Text
Bioinformatic protocol deviation	NCIT:C50996	String
Reason for bioinformatic protocol deviation	NCIT:C93529	String
WGS guideline followed	NCIT:C17564	String

Table 10: Module: Analysis. Apply analytical methods to existing data of a specific type. Ontology: EDAM:operation_2945.

Value	Ontology
NoInformation	HL7:NI
Invalid	HL7:INV
Derived	HL7:DER
Other	HL7:OTH
Negative infinity	HL7:NINF
Positive infinity	HL7:PINF
Un-encoded	HL7:UNC
Masked	HL7:MSK
Not applicable	HL7:NA
Unknown	HL7:UNK
Asked but unknown	HL7:ASKU
Temporarily unavailable	HL7:NAV
Not asked	HL7:NASK
Not available	HL7:NAVU
Sufficient quantity	HL7:QS
Trace	HL7:TRC

Table 11: Overview of null flavors. Each lookup in FAIR Genomes is supplemented with so-called 'null flavors' from HL7. These can be used to indicate precisely why a particular value could not be entered into the system, providing substantially more insight than simply leaving a field empty.