A metadata schema for documenting material samples from multiple domains

Stephen M Richard[1], Dave Vieglais[2], Andrea Thomer[3], Hyunju Song[3], Neil Davies[4], John Deck[4], Quan Gan[3], Eric C Kansa[5], Sarah Kansa[5], John Kunze[6], Kerstin Lehnert[7], Danny Mandel[3], Chris Meyer[10], Rebecca Snyder[10], Ramona Walls[12,13], Yuxuan Zhou[3], and Hong Cui[3]

(1) U.S. Geoscience Information Network, Tucson, AZ, United States,
(2) University of Kansas, Lawrence, KS, United States,
(3) University of Arizona, Tucson, United States,
(4) University of California Berkeley, Berkeley, CA, United States,
(5) Open Context, The Alexandria Archive Institute, San Francisco, United States,
(6) Ronin Institute for Independent Scholarship, Westminster, United States,
(7) Columbia University, Lamont-Doherty Earth Observatory, Palisades, United States,
(10) Smithsonian National Museum of Natural History, Invertebrate Zoology, Washington, DC,
(12) CyVerse, BIO5 Institute, Tucson, AZ, United States,
(13) Critical Path Institute (C-Path), Tucson, United States United States,

Abstract

The Internet of Samples (iSamples) project brings together material sample metadata from the System for Earth Sample Registration (SESAR), Open Context, the Genomic Observatories Meta-Database (GEOME), and Smithsonian Institution Museum of Natural History (NMNH), representing geoscience, archaeology/anthropology, and biology disciplines. To create an index for sample discovery across these disparate domains, we reviewed the metadata schema and example metadata from each project partner to develop the sample description scheme described in this document. We determined that a single sample type classification vocabulary could not account for the spectrum of samples without becoming very large and unwieldy. By factoring the categorization into material type, material sample object type, and sampled feature type, it has been possible to classify the approximately 6,000,000 samples in the combined corpus. High-level vocabularies were developed based on random subsamples and unique values summaries from related fields in the source sample metadata, and tested with the project team using a card sorting exercise, and by developing code for automated mapping. Our goal was that each vocabulary should have on the order of 20 values and some hierarchy, values should be covering, but might overlap. These vocabularies are documented here, and registered with the ARDC Research Vocabularies Australia (RVA) vocabulary service for use by the community. The metadata schema is implemented as a JSON schema that is used to validate instance documents. To further test and evaluate the schema, mapping to the DataCite schema now used by IGSN (International Generic Sample Number), schema.org JSON-LD, the Biodiversity Information Standards (TDWG) Minimum Information about a Digital Specimen (MIDS), and Distributed System of Scientific Collections (DiSSCo) Open Digital Specimen (openDS) schema are included as appendices.

Introduction

The Internet of Samples (iSamples) project, funded by the U.S. National Science Foundation, is working to enable connections between diverse data derived from material samples across science domains to facilitate interdisciplinary collaborations. The project brings together sample collection and data repository managers from the System for Earth Sample Registration (SESAR)[1], Open Context (a publishing service maintained by the Alexandria Archive Institute, a metadata repository for archaeological artifacts and ecofacts)[2], the Genomic Observatories Meta-Database (GEOME)[3], and Smithsonian Institution Museum of Natural History (NMNH)[4], representing geoscience, archaeology/anthropology, and biology disciplines. The goal is a searchable global index of material samples described by rich metadata and linked to derived research products. iSamples aims to (i) enable connections between diverse and disparate information derived from material samples; (ii) support existing research programs and facilities that collect and manage diverse sample types; (iii) facilitate new interdisciplinary collaborations; and (iv) provide an efficient solution for FAIR samples, avoiding duplicate efforts in different domains.

To achieve these goals, iSamples must incorporate and help advance development and adoption of metadata vocabularies and content standards across science domains. As a starting point, we have developed a core sample description scheme applicable to material samples from any of the partner systems. Our approach to developing the schema was empirical, based on reviewing existing sample description schemes in use (Table 1) and on reviewing the content in the metadata records from project partner systems to evaluate what fields are populated.

Another important source for developing the schema is a metadata cross walk document prepared in conjunction with research for Damerow et al. [2021], and shared with the iSamples team by J E. Damerow (pers. Comm). Results from several workshops over the last several years have also be used; these include workshops to develop the basic and description metadata for IGSN, and to review the material sample metadata for the USGS National Geological and Geophysical Preservation program.

Table 1. Existing systems with some material sample description models.

System	Name	Notes
IGSN	International Generic sample Number	originally focused on Earth Science material samples, then expanded to register other kinds of samples. Original design had very simple 'registration' scheme and more in depth descriptive scheme that could be extended for different sample types[5,6]. IGSN is now incorporated into the DOI system under the authority of DataCite, and uses the DataCite metadata schema.

[1] https://www.geosamples.org/
[2] https://opencontext.org/
[3] https://geome-db.org/
[4] https://naturalhistory.si.edu/research
[5] https://igsn.github.io/metadata/, https://github.com/IGSN/metadata
[6] https://archive-intranet.ardc.edu.au/display/DOC/IGSN+Descriptive+Metadata

System	Name	Notes
GeoSciML[7]	Geoscience markup language	XML implementation based on a conceptual model for geologic data. Model is presented using UML. Material samples are modeled as a kind of sampling feature[8] based on the OGC Observation and Measurement model.
ODM2[9]	Observation data model	Entity relation model developed by the Critical Zone Community, similar to OGC Observation and Measurement model; treats material sample as 'specimen', a kind of sampling feature[10].
TDWG-MIDS	Biodiversity Information Standards Minimum Information about a Digital Specimen	Basic content model for a metadata record describing a material sample[11]
SESAR[12]	System for Earth SAmple Registration	US Node in IGSN network; extends IGSN metadata schema with additional content[13].
W3C SSN[14]	World Wide Web Consortium activity developing an ontology for sensor networks;	Sample and sampling activity are included in the model[15]
ESS-DIVE	Environmental System Science Data Infrastructure for a Virtual Ecosystem	ESS-DIVE Sample ID and Metadata Reporting Format (IGSN-ESS)[16]
DISSCo[17]	Distributed Systems of Scientific Collections	DiSSCo specimen & collection classification. Focused on digital representation of material samples, and linkage to related resources[18].

In the context of this work, we are interested in material samples. These are material entities collected and identified with the intention of being representative of some feature of interest in the world[19]. The term 'sample' as used in this document should be understood to mean such a 'material sample'.

[7] https://geosciml.org/

[8] see https://docs.ogc.org/is/16-008/16-008.html#357 and https://docs.ogc.org/is/16-008/16-008.html#443

[9] https://github.com/ODM2/ODM2

[10] https://github.com/ODM2/ODM2/blob/master/doc/ODM2Docs/ext_samplingfeatures.md#sampling-features-that-are-specimens;
https://odm2.github.io/ODM2/schemas/ODM2_Current/tables/ODM2SamplingFeatures_Specimens.html

[11] https://github.com/tdwg/mids. Work in progress; see working draft branch for current version.

[12] https://www.geosamples.org/

[13] https://zenodo.org/doi/10.5281/zenodo.3875530

[14] https://www.w3.org/TR/vocab-ssn/

[15] https://www.w3.org/TR/vocab-ssn/#SOSASample

[16] https://github.com/ess-dive-community/essdive-sample-id-metadata

[17] https://dissco.tech/2020/03/31/what-is-a-digital-specimen/

[18] https://docs.google.com/document/d/19OPyOm9VF2qfI3M6Rm

[19] https://dwc.tdwg.org/list/#dwc_MaterialSample, http://www.opengis.net/doc/as/om/3.0

Categorization of sample type is an attribute typically included metadata systems for describing samples. In most cases this categorization is done with a controlled vocabulary to provide a means to zero in on kinds of sample a user is looking for. We reviewed the sample type classifications used in the metadata corpus from the project partners and determined that a single sample type classification vocabulary could not account for the spectrum of samples without becoming too large and unwieldy. By factoring sample type categorization into material type, material sample object type, and sampled feature type, it has been possible to classify the approximately 6,000,000 samples in the combined corpus. Material Type specifies the kind of substance that constitutes the sample, for example 'Rock', 'Organic material', 'Liquid water', 'Anthropogenic material'. Material Sample Object Type specifies the kind of physical object identified as the sample, for example 'Fossil', 'Artifact', 'Organism part'. Sampled Feature Type specifies the thing in the world the sample is intended to represent, for example 'Site of past human activity', 'Atmosphere', 'Extraterrestrial environment'. These vocabularies are described in more detail below.

This paper first describes an information model for the content of a material sample description. Next the vocabularies for documenting sample type are described, followed by a discussion of the physical implementation of the information model. The final section briefly summarizes testing of the metadata schema for integration of sample descriptions from the project partners.

Sample description information model

The information model for the material sample description scheme defines the content items used to describe a sample. Only the most critical items are made mandatory. The physical implementation used for the iSamples catalog is described in a subsequent section,

Registration metadata

Registration metadata includes information about the sample description (metadata) record itself and is useful for management of the metadata in a distributed information system such as is envisioned for the iSamples architecture (Lehnert et al., 2021).

Metadata identifier

This is an identifier for the metadata record, distinct from the identifier for the physical object the metadata documents. Dereferencing[20] the *Sample identifier* (see following section) will typically return information from the metadata record describing the sample and is thus commonly confused with the *Metadata identifier*. The material sample and the metadata record are considered separate resources. A distinct identifier for the metadata record allows statements to be made (annotation) about the metadata record, for example to make corrections, point out errors, or add links to new resources.

Registering agent

This identifies the agent, a person or organization, responsible for the metadata record. Having contact information for the agent included is important to enable harvesters or other users to notify the metadata originator about problems or suggest updates and improvements to the sample metadata

[20] Dereferencing is the process of retrieving information about the entity (resource) that an identifier represents.

content. The agent should be specified with a name for human use to identify the agent, a resolvable URI to identify the agent in a linked data context, and a point of contact address or phone number. ORCIDs are recommended for persons and ROR identifiers for organizations. An institutional role e-mail is recommended as a point of contact address, e.g. 'sampleCurator@AcmeUni.edu', as these are less likely to go stale when staffing changes occur.

Basic discovery metadata

The following properties are considered mandatory for a basic sample description supporting discovery and resolution of sample identifiers.

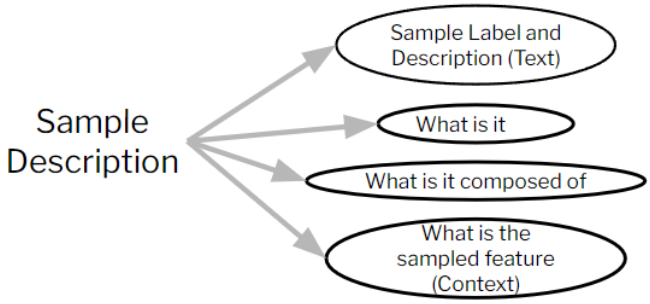

Figure 1. Questions addressed by basic discovery metadata.

Sample Identifier

This is a string that uniquely identifies the material sample. It should be a persistent, web-resolvable URI (e.g. an http URI). iSamples project participants are using ARK (Archival Resource Key) identifiers and IGSN[21] identifiers. The material sample is a physical entity and cannot be transmitted electronically. The iSamples metadata record is considered the default electronic representation of the physical entity. The material sample has an identifier, ideally attached to the physical entity; when this identifier is dereferenced electronically (on the web) the default electronic representation is provided as a proxy for the physical entity (Figure 2). The metadata record is the anchor for the 'Digital Specimen' network of linked data about the material sample. It has a separate identifier from the material sample, allowing statements to be made about the metadata distinct from statements about the material sample.

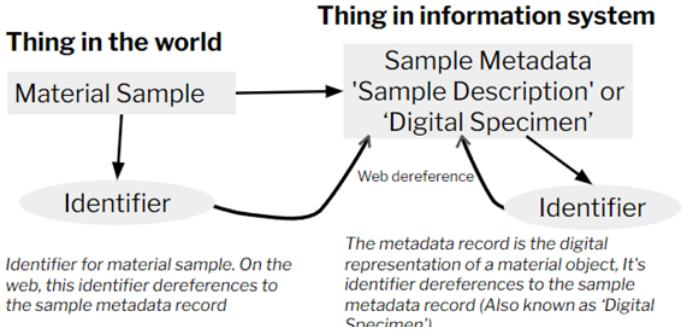

Figure 2. Relationship between the material sample and its representation in the digital information system.

[21] IGSN identifiers are now issued by Data Cite and conform to the DOI scheme. See https://support.datacite.org/docs/igsn-faq.

Label

A text string that identifies the sample for human users. In many cases this is an identifier string assigned by the original sample collector. Other identifiers or labels assigned to the sample can be recorded in the alternate identifier field (see below).

Alternate identifiers

This item contains other identifiers used to represent the sample. These might include a field identifier assigned by the sample collector, laboratory-assigned identifiers used in analysis workflows, or museum accession numbers. The identifier string value should be accompanied by a scheme name to identify the scope in which the identifier is used and assumed to be unique.

Description

A free text description of the sample. This text will be indexed by search applications and read by users to understand what the sample is. More information is better. The description should include how and why the sample was collected, any particulars of the collection context, and a physical description of the sample-size, mass, color, material composition, etc. Cryptic abbreviations and acronyms, as well as project- or domain-specific jargon should be avoided.

Sample type: Material type, Material sample object type and Sampled feature type

Samples can be categorized in various ways. After reviewing example metadata from project participants, it was determined that factoring sample type into three properties: material type, material sample object type and sampled feature type provides an effective scheme that can be implemented with ~20 classes in each property vocabulary. These properties answer the basic questions about the sample: what is it composed of?, what is it?, what does it represent? The vocabularies are described in more detail and listed in the vocabulary section, below.

Additional important information

These properties are not essential for basic discovery, but useful to enable access and reuse, and to fully implement the FAIR principles for material samples.

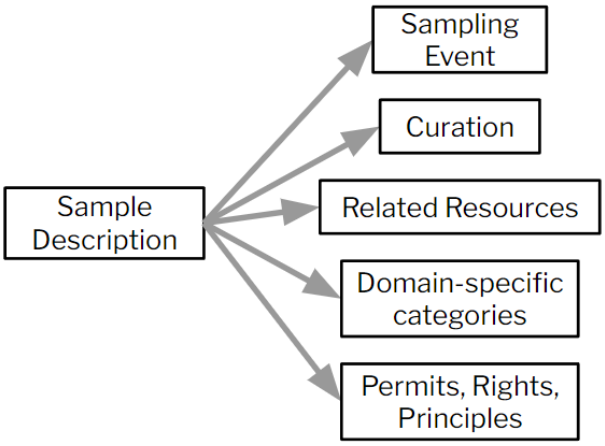

Figure 3. Other sample description elements

SamplingEvent

A sample is the product of a sampling event. Documentation of the sampling event comprises subitems that specify where and when the sampling took place, what was sampled, how the sampling was done, any conventions, protocols, or policies followed in sample collection, permissions obtained for sampling, and who did the sampling. These aspects are elucidated below.

Where:

For many samples, the location where the sample was collected provides important information for data integration and understanding the sample context. Location can be specified at different levels of resolution, ranging from a general place name to a very specific location in a high-resolution local coordinate system like a survey grid at an excavation. Reporting the location with latitude and longitude coordinates in decimal degrees with 2 to 4 decimal places precision using the WGS84 coordinate reference system is required for ease of use and comparison to sample locations reported from different collections. To report locations using other reference systems, e.g. site-specific grids or linear reference like depth in a non-vertical borehole, these should be described in the sampling site description as free text. Be sure to include description of the coordinate reference system used.

If access to the sampling location or removal of material from the location required one or more permits, information about permits obtained for access and sampling should be included in the authorized_by property. Sampling locations might also be approximated to avoid abuse; the metadata scheme includes a field ('obfuscated') to indicate if the location is intentionally approximate.

Procedure

The sampling procedure or protocol should be documented as a property of the sampling event. The procedure might be specified with a text description, or if it is formally documented with an identifier (e.g. protocols.io[22]), a label and identifier should be provided.

[22] https://www.protocols.io/

Sampling event responsible parties

The parties responsible for the sample collection event should be acknowledged in this section. 'Party' can be a person or an organization. Each party has a role such as collector or funder that specifies their relationship to the sampling event. Role names from a controlled vocabulary should used, but this specification does not specify a particular vocabulary; ISO19115 Role Codes[23] or the Contributor Role Ontology[24] are suggested for use. A name must be provided for each party. Organization affiliation can be specified for persons. Each party should have a persistent, resolvable identifier to support data integration and interoperability. The use of ORCID[25] to identify persons and ROR[26] to identify organizations is recommended.

Curation and access

The curation and access description documents where the sample is currently located, how it has been preserved or otherwise modified, procedures to access the sample, and constraints on usage of the sample.

In analytical chemistry, sample preparation refers to the ways in which a sample is treated prior to its analyses.[27] For biological samples (and some other types) the 'Preparation' or preservation method is an important consideration. In Darwin core 'preparations'[28] is defined as 'A preparation or preservation method for a dwc:MaterialEntity [material entity]'. For iSample purposes, we consider curation to be the set of activities between when a sample was originally collected and assigned an identifier and its current state. These activities might include various preservation steps to stabilize the sample in its original state. Preparation is the subset of curation activities specifically focused on changes to the sample necessary to support some analytical activity.

Use of a controlled vocabulary to describe sample processing would help search precision for use cases when sample processing is a filter criteria. One possible vocabulary is the Sample Processing and Separation Techniques Ontology[29], and projects are under way to develop conventions in the Biomolecular Ocean Observing and Research community (Pesant et al., 2023). Description of curation activities in text will assist users in evaluating the fitness of a sample for reuse purposes.

Curation responsible parties

The parties responsible for curation of the sample should be acknowledged in this section. 'Party' can be a person or an organization. Each party has a role such as curator, collections manager or analyst. Role names from a controlled vocabulary should be used, but this specification does not specify a particular vocabulary. A name must be provided for each party; organization affiliation can be specified for persons. Each party should have a persistent, resolvable identifier to support data integration and

[23] https://wiki.esipfed.org/ISO_19115-3_Codelists#CI_RoleCode
[24] Contributor Role Ontology https://doi.org/10.5281/zenodo.2593738
[25] https://orcid.org/
[26] ROR https://ror.org/
[27] https://en.wikipedia.org/wiki/Sample_preparation
[28] https://dwc.tdwg.org/list/#dwc_preparations
[29] https://bioportal.bioontology.org/ontologies/SEP

interoperability. The use of ORCID[30] to identify persons and ROR[31] to identify organizations is recommended. At least one party should be identified as the point of contact if the sample is available for viewing or possible loan, with included contact information.

Storage location

The location where the sample is physically stored should be identified at the organization or facility level, if known or applicable. Details about specific shelf or drawer is useful for collection curators, but not necessary for sample search and evaluation.

Access and usage constraints

Any restrictions or policies that determine whether or how the sample may be viewed or borrowed should be explained in the metadata. If the access and usage policies are defined in online documents, URLs to access those documents should be provided.

Domain-Specific Categories: Keywords

Keywords are not required but are recommended. They are useful for providing other categorization or descriptive words to make discovering and evaluating a sample more accurate. The three iSamples sample type vocabularies are very high level and need to be supplemented by keywords that categorize the material sample in more domain-specific terms. The keyword implementation allows for a keyword term, keyword identifier, keyword scheme name and scheme identifier. The keyword term is the minimal requirement; if no identifier or keyword scheme is provided the term is treated as a simple 'tag'—a word or phrase that expected users might associate with the sample. If the keyword is from a controlled vocabulary with more precise semantics and can be used for cross-domain searches, the keyword identifier, scheme name, or scheme identifier should be provided. Use of keywords from controlled vocabularies can be used to associate categorical property values with the sample, e.g. geologic age, biological taxonomy classification, or archaeologic material culture.

Related Resources

To take advantage of the linked data capabilities of the World Wide Web, it is useful to provide links to related resources for understanding the sample, discovering data derived from the sample, or finding related research. The most important relationship to report is linkage between parent and child samples. Other related resources include publications or datasets using data from the sample, and collections in which the sample is a member. Links to related resources should include a resolvable identifier (e.g. an http URI) for the target resource, a label for the link, and a relationship type term from a controlled vocabulary to categorize the nature of the relationship. A sample relationship type vocabulary is in development.[32]

[30] https://orcid.org/
[31] ROR https://ror.org/
[32] https://docs.google.com/spreadsheets/d/1yk3ZEWzEc1cXBOwb-foCtq3isaBBa95Rv8eA6TIyieM

Permits, Rights, Principles

The sample metadata includes properties to specify legal, cultural, or policy considerations or constraints that might be critical to enable use of the sample and derived data. As mentioned in the Sampling event section (above), permits that provide permission to access sampling locations or remove material should be cited in the 'authorized_by' property on the sampling event. Restrictions on access to or use of the sample can be specified in the 'dc_rights' property on the sample. Conformance with other policies and procedures, e.g. those related to the CARE Principles[33] (Carroll et al., 2020), can be asserted using the 'complies_with' property on the sample. These metadata fields currently contain string values, allowing text descriptions, or links to documents or other resources.

Other Properties, Domain Specific Categories and Contexts

There are a variety of other properties that are important for samples in some domains, for example dimensions, mass, origin age, rock type, mineral type, biological taxon, tectonic setting, archaeological culture, or biome. If such properties have categorical values, they can be asserted in the sample descriptions as keywords (see Domain-Specific Categories: Keywords). In version one sample description implementation for iSamples, properties with numeric values should be summarized with text statements in the sample descriptions. This will make the information available to human users. The text will be indexed as free text for search, but the results will be somewhat unpredictable because the information is not structured. We envision future development of extension profiles to define structured and interoperable scheme for assigning such properties.

Vocabularies

A single classification vocabulary for sample types cannot account for the spectrum of possibilities in the project scope without becoming very large and unwieldy. By factoring the categorization into material sample object type, material type, and sampled feature type, it has been possible to classify the approximately 6,000,000 samples in the combined corpus. Material sample object type specifies the kind of physical object identified as the 'sample', for example 'Fossil', 'Artifact', 'Experiment product', 'Organism part'. Material type specifies the kind of substance that constitutes the sample, for example 'Rock', 'Organic material', 'Liquid water', 'Anthropogenic material'. Sampled feature type specifies the thing in the world the sample is intended to represent, for example 'Site of past human activity', 'Atmosphere', 'Extraterrestrial environment'. High-level vocabularies were developed based on subsamples and unique value summaries from related fields in the source sample metadata. The draft vocabularies were tested with the project team using a card-sorting exercise, and by developing code for automated mapping from source metadata to iSamples metadata. Our goal was that each vocabulary should have around 20 values, have some hierarchy, and should cover the range of possible values. Vocabulary concepts are possibly overlapping, such that some samples might be categorized in more than one class.

[33] https://www.gida-global.org/care

The vocabularies have been implemented in Resource Description Framework (RDF)[34] using the Simple Knowledge Organization System (SKOS)[35] vocabulary. Each concept has a Uniform Resource Identifier (URI)[36], preferred label, and a definition; notes and examples are included for some concepts. Hierarchical relations are represented using skos:broader. Where mapping to concepts in other vocabularies has been found, relationships are represented using one of the SKOS mapping relations: broadMatch, narrowMatch, closeMatch, exactMatch. The vocabularies are defined and maintained as Terse RDF Triple Language (Turtle)[37] files in the iSamples Github, and published with the Australian Research Data Commons vocabulary service[38] or the ESIPFed Community Ontology Repository[39]. URIs are defined for each category and resolved using the W3ID re-direction service for Web applications[40]. We have opted to generate the URI tokens based on the preferred labels for the concepts, favoring user-friendliness over other considerations. The URIs include a version identifier token, with the value '1.0' for the current release. Formal versioning policies for individual categories or the vocabulary as a unit have not been defined. The following tables list the category terms for each of the three base sample type vocabularies. Accompanying figures display the hierarchical relationships of categories.

Type	Definition
Anthropogenic material	Material produced by human activity. Material that would not be found in nature without human intervention. Thus clay would be a 'Mineral' material, but fired clay in a brick or ceramic would be an 'Anthropogenic material'. Native copper would be a Mineral, but smelted copper, extracted from ore that might include native copper (among other sulfide and oxide minerals) would be 'Anthropogenic metal material'.
Anthropogenic metal material	Metal that has been produced or used by humans. Samples of naturally occurring metallic material (e.g. native copper, gold nuggets) should be considered mineral material. Metallic material is material that when polished or fractured, shows a lustrous appearance, and conducts electricity and heat relatively well. Metals are typically malleable (they can be hammered into thin sheets) or ductile (can be drawn into wires). The boundaries between metals, nonmetals, and metalloids fluctuate slightly due to a lack of universally accepted definitions of the categories involved. (https://en.wikipedia.org/wiki/Metal, c.f. http://purl.obolibrary.org/obo/ENVO_01001069)
Any Ice	a solid material that is in a solid state under the temperature and pressure conditions of the preserved sample, that but is normally a liquid or gas at Standard Temperature and Pressure (STP)[41] that is in a solid state under the observed temperature and pressure conditions.

[34] https://www.w3.org/RDF/
[35] https://www.w3.org/TR/skos-reference/
[36] https://www.ietf.org/rfc/rfc3986.txt
[37] https://www.w3.org/TR/turtle/
[38] https://vocabs.ardc.edu.au/search/#!/?activeTab=vocabularies&pp=15&q=isamples
[39] http://cor.esipfed.org/ont#/so/isample
[40] https://github.com/perma-id/w3id.org#permanent-identifiers-for-the-web
[41] The US National Institute of Standards and Technology (NIST) standard temperature and pressure (STP) is 20°C (68°F) and 1 atm (14.696 psi, 101.325 kPa). This standard is also known as normal temperature and pressure (NTP).

Type	Definition
Biogenic non-organic material	Material produced by an organism but not composed of 'very large molecules of biological origin.' E.g. bone, tooth, shell, coral skeleton,
Dispersed media	A material contains discrete elements of one medium that are dispersed in a different, continuous fluid medium. The dispersed component can be a gas, a liquid or a solid (based on https://en.wikipedia.org/wiki/Dispersed_media). Does not include mixtures of granular material like soil, sediment, particulate, or solids that would be considered a rock. E.g. aerosol ENVO_00010505, foam ENVO_00005738, emulsion ENVO_00010506, colloidal suspension ENVO_01001560, scum(?)ENVO:00003930, clathrate?
Fluid material	a substance that continually deforms (flows) under an applied shear stress, or external force. Fluids are a phase of matter and include liquids, gases and plasmas. They are substances with zero shear modulus, or, in simpler terms, substances that cannot resist any shear force applied to them. (https://en.wikipedia.org/wiki/Fluid)
Frozen water	Water that is in a solid state.
Gaseous material	Material composed of one or more chemical entities that has neither independent shape nor volume but tends to expand indefinitely (http://purl.obolibrary.org/obo/ENVO_01000797). Infer that the sample is curated in some kind of container.
Liquid water	A material primarily composed of dihydrogen oxide in its liquid form; infer that the sample is curated in some kind of container.
Material	Top Concept in iSamples Material Category scheme
Mineral	Material consists of a single mineral or mineraloid phase. . 'A mineral is an element or chemical compound that is normally crystalline and that has been formed as a result of geological processes.' (Nickel, Ernest H. , (1995), The definition of a mineral, The Canadian Mineralogist. 33 (3): 689–90). Include mineraloids. ... A material primarily composed of some substance that is naturally occurring, solid and stable at room temperature, representable by a chemical formula, usually abiogenic, and that has an ordered atomic structure. (http://purl.obolibrary.org/obo/ENVO_01000256). Comment: t The identity of a mineral species is defined by a crystal structure and a chemical composition that might include various specific elemental substitutions in that structure. This class includes mineraloids,Mineraloid: A naturally occurring mineral-like substances that does not demonstrate crystallinity. Mineraloids possess chemical compositions that vary beyond the generally accepted ranges for specific minerals. Examples: obsidian, Opal. (https://en.wikipedia.org/wiki/Mineraloid)
Mixed soil, sediment, or rock	Material is mixed aggregation of fragments of undifferentiated soil, sediment, or rock origin. e.g. cuttings from some boreholes (rock fragments and caved soil or sediment).
Natural Solid Material	A naturally occurring solid material that is not anthropogenic, biogenic, or ice.Undifferentiated, soil, sediment, rock, or natural particulates. Typically (necessarily?) a granular aggregate that might include any of the previous constituents. Use for Earth Material aggregates of uncertain origin

Type	Definition
Non-aqueous liquid material	Liquid that is composed dominantly of material other than water. Includes liquids that do not fit in any other category. E.g. alcohol, petroleum.Liquid composed dominantly of material other than water. Includes liquids that do not fit in any other category. E.g. alcohol, petroleum.
Organic material	Environmental mMaterial derived from living organisms and composed primarily of one or more very large molecules of biological origin. Examples: body (animal or plant), body part, tissue, biological fluids, biological waste, algal material, biofilm, necromass, plankton. source: http://purl.obolibrary.org/obo/ENVO_01000155
Other anthropogenic material	Non-metallic material produced by human activity. Organic products of agricultural activity are both anthropogenic and organic. Include lab preparations like XRF pellet and rock powders. Examples: ceramics, concrete, slag, (anthropogenic) glass, mine tailing, plaster, waste.
Particulate	Material consists of microscopic particulate material derived by precipitation, filtering, or settling from suspension in a fluid, e.g. filtrate from water, deposition from atmosphere, astro material particles. Might include mineral, organic, or biological material. ENVO definition (ENVO_01000060) has "composed of microscopic portions of solid or liquid material suspended in another environmental material.", refine here to define as the solid particles, distinct from a material in which they are suspended. A material that includes solid or liquid particles suspended in another material would be a dispersed media in this scheme, not defined in ENVO. Human manufactured particulates (e.g. rock powder) should be categorized as 'anthropogenic 'Anthropogenic material' as well as 'Particulate'. {@en}
Rock	Consolidated aggregate of particles (grains) of rock, mineral (including native elements), mineraloid, or solid organic material. Includes mineral aggregates such as granite, shale, marble; natural glass such as obsidian; organic material formed by geologic processes such as coal; extraterrestrial material in meteorites; and crushed rock fragments like drill cuttings from rock. (based on http://resource.geosciml.org/classifier/cgi/lithology/rock, same as http://purl.obolibrary.org/obo/ENVO_00001995)
Rock or sediment	Material is rock or sediment. E.g. for samplesFor example core from boreholes from subsurface cores that are not well described, from drill holes that likely penetrate sediment near the surface and might be sampling rock at greater depth, with descriptions that do not clearly distinguish non-consolidated sediment from rock. Also for samples like dredge hauls and remote vehicle scoops that mix rock and sediment from a water body bottom.
Sediment	Solid granular material transported by wind, water, or gravity, not modified by interaction with biosphere or atmosphere (to differentiate from soil). Particles are derived by erosion of pre-existing rock, from shell or other body parts from organisms, or precipitated chemically in the surficial environment (http://resource.geosciml.org/classifier/cgi/lithology/sediment). Sediment is not consolidated, i.e. Particulate particulate constituents of a compoundthe material do not adhere to each other strongly enough that the aggregate can be considered a solid material in its own right.

Type	Definition
	(http://resource.geosciml.org/classifier/cgi/consolidationdegree/consolidated). Similar to http://purl.obolibrary.org/obo/ENVO_00002007
Soil	Mixed granular mineral and organic matter modified by interaction between earth material, biosphere, and atmosphere, consisting mostly of varying proportions of sand, silt, and clay, organic material such as humus, gases, liquids, and a broad range of resident micro- and macroorganisms. (https://en.wikipedia.org/wiki/Soil,) Soil consists of horizons near the Earth's surface that, in contrast to the underlying parent material, have been altered by the interactions of climate, relief, and living organisms over time. (http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/edu/?cid=nrcs142p2_054280, http://purl.obolibrary.org/obo/ENVO_00001998)

Table 1. Material type vocabulary. Categories are listed in alphabetic order. See Figure 4 for hierarchical structure of the vocabulary.

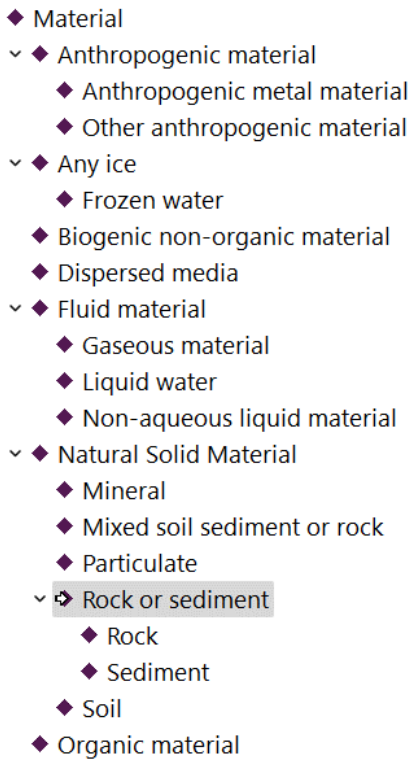

Figure 4. Hierarchical relationships between material types.

Table 2. Material sample object type vocabulary
Categories are listed in alphabetic order. See Figure 5 for hierarchical structure of the vocabulary.

Type	**Definition**
Aggregation	An aggregate material sample that is not biogenic or composed of anthropogenic material fragments. Examples: loose soil or sediment (e.g. in a bag), rock chips, particulate filtrate or precipitate; rock powders.
Analytical preparation	Material sample that is a product of processing required for some observation procedure, e.g. thin section, XRF bead, SEM stub, rock powder. If identified separately, this should have a 'parent' link to the original sample.
Anthropogenic aggregation	An aggregate material sample consisting of fragments of material produced by human activity, not described individually, and generally not all originating from the same object. Includes pottery in an excavation unit that gets an aggregate description, production waste, production raw-materials, or other residues (broken bits of plaster from a destroyed wall), synthetic powders.
Any aggregation material sample	Sample consists of a bunch of material fragments, not related to the same object (e.g. not a bunch of broken pot sherds that might be reassembled), but taken together representative of the sampled feature. Examples: loose soil, sediment, crushed rock, particulate, bunches of unrelated pot sherd, human production waste, filtrates and residues. The sample requires some kind of container to keep it together. Cores of loosely consolidated material are considered 'Solid material sample' when preserved such that the internal parts have spatial relationships (e.g. upper part, lower part, sedimentary structures).
Biological material sample	Material sample representative of one or more living organisms from a particular biome context, megascopic or microscopic
Artifact	An object made (manufactured, shaped, modified) by a human being, or precursor hominid. Include a set of pieces belonging originally to a single object and treated as a single sample. A sample is classified as an artifact if it is intended to represent the manufactured item. For example, core, thin section, peel, glass slide smear are products of human manufacture, but are intended to represent materials or biological entities, thus not considered 'Artifact' as a material sample object type.
Biome aggregation sample	Material sample that is an aggregation of whole or fragmentary parts of multiple organisms, microscopic or megascopic, representative of some sampled feature.
Bundle biome aggregation	Material sample that is an aggregation of whole organisms representative of some biome.
Experiment product	Material sample that is the product of an experimental procedure (e.g. synthetic material)

Type	Definition
Fossil	Material sample that is the remains or trace of one or more organisms preserved in rock; includes whole body, body parts (usually bone or shell), and trace fossils. An organism or organism part becomes a fossil when it has undergone some fossilization process that generally entails physical and chemical changes akin to diagenesis in a sedimentary rock. Trace fossils are manifestations of biologic activity preserved in a rock body (typically sedimentary), without included preserved body parts. There are many processes that lead to fossilization, including permineralization, casts and molds, authigenic mineralization, replacement, recrystallization, adpression, carbonization, and bioimmuration (https://en.wikipedia.org/wiki/Fossil).
Fluid in container	Material sample is a liquid, gas, or mixed dominantly fluid phase material that is necessarily inside some container. Fluids might include minor solid particles. The container is typically human made, but might be a natural fluid container, e.g. fluid inclusion in a mineral grain. Fluids might be colloids, foams, gels, or suspensions. The sample is the fluid substance; fluid samples collected to analyze the contained biome should be considered 'Biome Aggregation'
Non biologic solid object	Material sample is an individual solid object, the substance of which is not formed directly by or part of a living organism
Organism part	Material sample is part of an organism, e.g. a tissue sample, plant leaf, flower, bird feather. Includes internal parts not composed of organic material (e.g. teeth, bone), and hard body parts that are not shed (hoof, horn, tusk, claw). Hair is tricky, include here for now. Does not necessarily imply existence of parent sample. Not fossilized; generally includes organism parts native to deposits of Holocene to Recent age.
Organism product	Material sample is a thing produced by some organism, generally not composed of organic material or including biological tissue, e.g. Shell, antler, egg shell, coral skeleton (organic tissue not included), fecal matter, cocoon, web. Consider internal parts not composed of organic material (e.g. teeth, bone) and hard body parts that are not shed (hoof, horn, tusk) to be organism parts.
Other solid object	A non-biologic solid object that is not one of the other types. The sample is not intended as representative of its constituent material. Examples -- a bird nest, termite tower, ventrifact, geode.
Material sample	A material entity that represents an entity of interest in whole or in part (http://rs.tdwg.org/dwc/terms/MaterialSample). Top concept in material sample object type hierarchy. Represents any material sample object type.
Research product	Material sample is a product of some research workflow, e.g. a thin section, an XRF pellet, a grain mount, SEM stub, synthetic rock or mineral ... In general there should be a link to a parent material sample from which this was derived. Might be aggregation (e.g. a synthetic material powder) or a solid object.
Slurry biome aggregation	Material sample that consists of mixed organic and inorganic material, including whole organisms and organism fragments.

Type	Definition
Solid material sample	Individual solid object, not formed directly by or part of a living organism, that is representative of some material. Many sediment cores consist of non-consolidated or weakly consolidated material but are considered solid objects if the core is preserved intact to observe the sedimentary structures and particle relationships within the sediment. If this material were 'disaggregated' into a mass of granular material to put in a bag, it would become an aggregate (msot:genericaggregation).
Whole organism material sample	Material sample consists of the bodies of one or more entire organisms of the same species, from any kingdom.

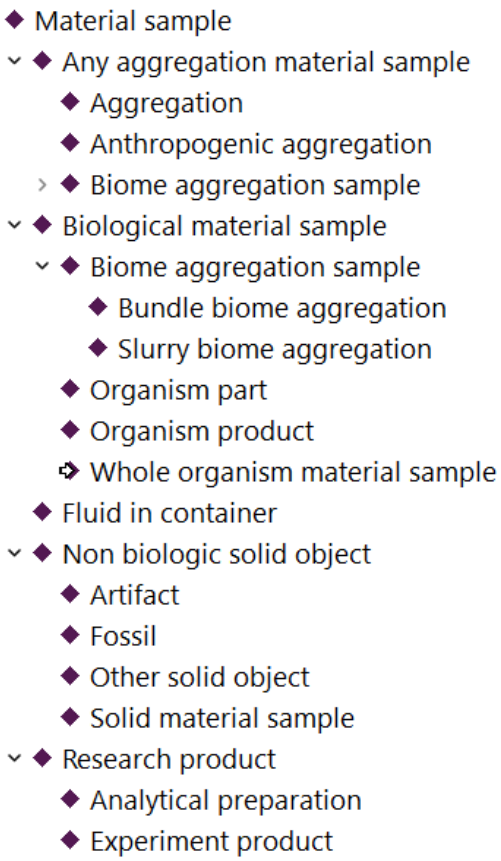

Figure 5. Hierarchical relationships of material sample object type terms.

Table 3. Sampled feature type vocabulary.
Categories are listed in alphabetic order. See Figure 6 for hierarchical structure of the vocabulary.

Type	Definition
Active human occupation site	Sampled feature is a site at which there are ongoing human activities

Type	Definition
Anthropogenic environment	Sampled feature is produced by or related to human activity past or present.
Any sampled feature	Any thing that can be sampled. Top concept in sampled feature type vocabulary.
Atmosphere	Sampled feature is the Earth's atmosphere
Biological entity	Sampled feature is an organism. Use for samples that represent some species of organism as the proximate sampled feature, not the environment that the organism inhabits.
Earth Surface	Sampled feature is the interface between solid earth and hydrosphere or atmosphere. Includes samples representing things collected on the surface, in the uppermost part of the material below the surface, or air or water directly at the contact with the Earth surface.
Earth environment	Sampled feature is the natural Earth environment
Earth interior	Sampled feature is solid material from within the Earth
Earth surface	Sampled feature is the interface between solid earth and hydrosphere or atmosphere. Includes samples representing things collected on the surface, in the uppermost part of the material below the surface, or air or water directly at the contact with the Earth surface.
Experiment setting	Sampled feature is an experimental set up that produced the sample; the sample is the product of the experiment.
Extraterrestrial environment	Sampled feature is the environment outside of solid earth, hydrosphere, or atmosphere.
Glacier environment	Sampled feature is a glacier, ice sheet, ice shelf, iceberg, or rock or water directly under or on top of such ice.
Laboratory or curatorial environment	Sampled feature is a laboratory or other research site, collected with intention of characterizing the environment in which data are collected or other research conducted, that might affect results or safety; e.g. lab blank measurements.
Lake river, or stream bottom	Sampled feature is the interface between the solid Earth interface and a lake or flowing water body.
Marine environment	Sampled feature is the marine hydrosphere.
Marine water body bottom	Sampled feature is the interface between the solid Earth and a marine or brackish water body. Includes benthic boundary layer: the bottom layer of water and the uppermost layer of sediment directly influenced by the overlying water.
Site of past human activities	Sampled feature is a place where humans have been and left evidence of their activity. Includes prehistoric and paleo hominid sites.
Subaerial surface environment	Sampled feature is the interface between solid Earth and atmosphere. Sample is collected on the surface, or immediately below surface (zone of bioturbation). Include soil and recently deposited subaerial sediment at the surface.

Type	Definition
Subsurface fluid reservoir	Sampled feature is fluid that resides in fractures, intergranular porosity or other open space in the solid earth.
Terrestrial water body	Sampled feature is terrestrial hydrosphere-- lake, other standing water, or a flowing water body (river, stream..). Include saline water in terrestrial evaporite environments.
Water body	Sampled feature is the Earth's hydrosphere.

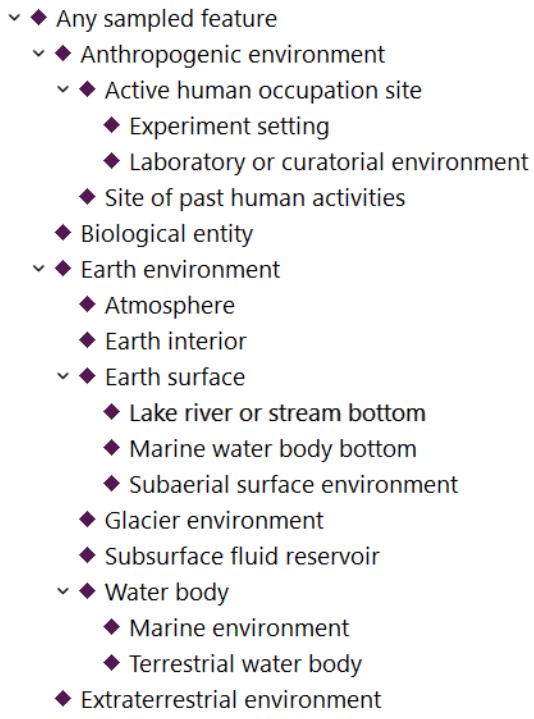

Figure 6. Hierarchical relationships of sampled feature types.

Vocabulary extensions

The vocabularies described in the previous section define categories at a high level, intended to cover material samples collected in any domain of interest. As such, they do not provide the granularity of categorization users in particular science communities will expect. To satisfy these use cases we expect user communities to develop extension vocabularies that provide the necessary granularity. The high-level vocabularies defined here provide a logical framework and extension points for these domain-specific vocabularies to be developed with more granular sub classes rooted in iSample vocabulary concepts. This approach will promote the development of extension vocabularies with a consistent logical basis, facilitating interoperability.

For example, in the geoscience community, classification of a sample as simply ‘mineral’ is too broad for most searches. A more specific mineral material classification is needed. As an example, we have

developed an extension vocabulary using the Nikel-Strunz mineral classification[42], which divides the domain of known minerals (with approximately 5800 species) into 10 classes. This mineral group vocabulary has the iSample material 'mineral' as top concept, with the Nikel-Strunz classes as the child concepts in the vocabulary.

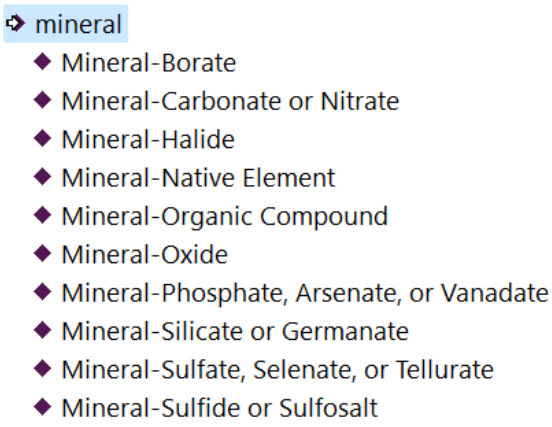

Figure 7. Extension vocabulary for mineral groups. 'mineral' is a category in the iSamples material type vocabulary. It is the top concept in the extension vocabulary for more granular categorization of minerals.

Other extension vocabularies might provide more granular subclasses under multiple concepts in the parent vocabulary. For instance in the materials extension for archaeological samples, various material types in the base iSamples vocabulary have more granular subclasses.

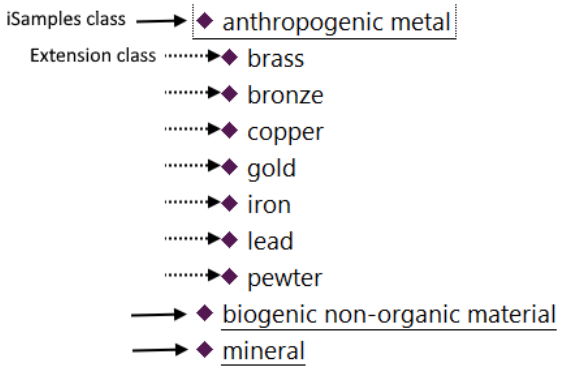

Figure 8. Material type extension vocabulary for archaeological materials. In this example, there are multiple top concepts in the extension vocabulary (the 'iSamples classes) that are categories in the base iSamples material type vocabulary. The extension vocabulary ('Extension classes') has more granular categories to subclass those from the base vocabulary.

[42] https://en.wikipedia.org/wiki/Nickel–Strunz_classification

Implementation

The conceptual model for sample description was constructed first as a UML model. The conceptual model is broadly based on the Sensor, Observation, Sample, and Actuator (SOSA) ontology (Janowicz et al., 2018) and the W3C Prov-O ontology (Lebo et al., 2013)[43]. The broad framework of the model is shown in the diagram in Figure 9. The iSamples model adds some classes. SamplingEvent is related to a SamplingSite that accounts for the spatial context that was sampled. A material sample (sosa:Sample) has related curation activities that document processing and preservation of the sample after it has been collected. The sample is linked to an open world of other resources through the SampleRelation class, which implements the Digital Extended Specimen information graph (Hardisty et al., 2022). Responsibility is a class linked to any prov:Activity subclass to document prov:Agents related to the

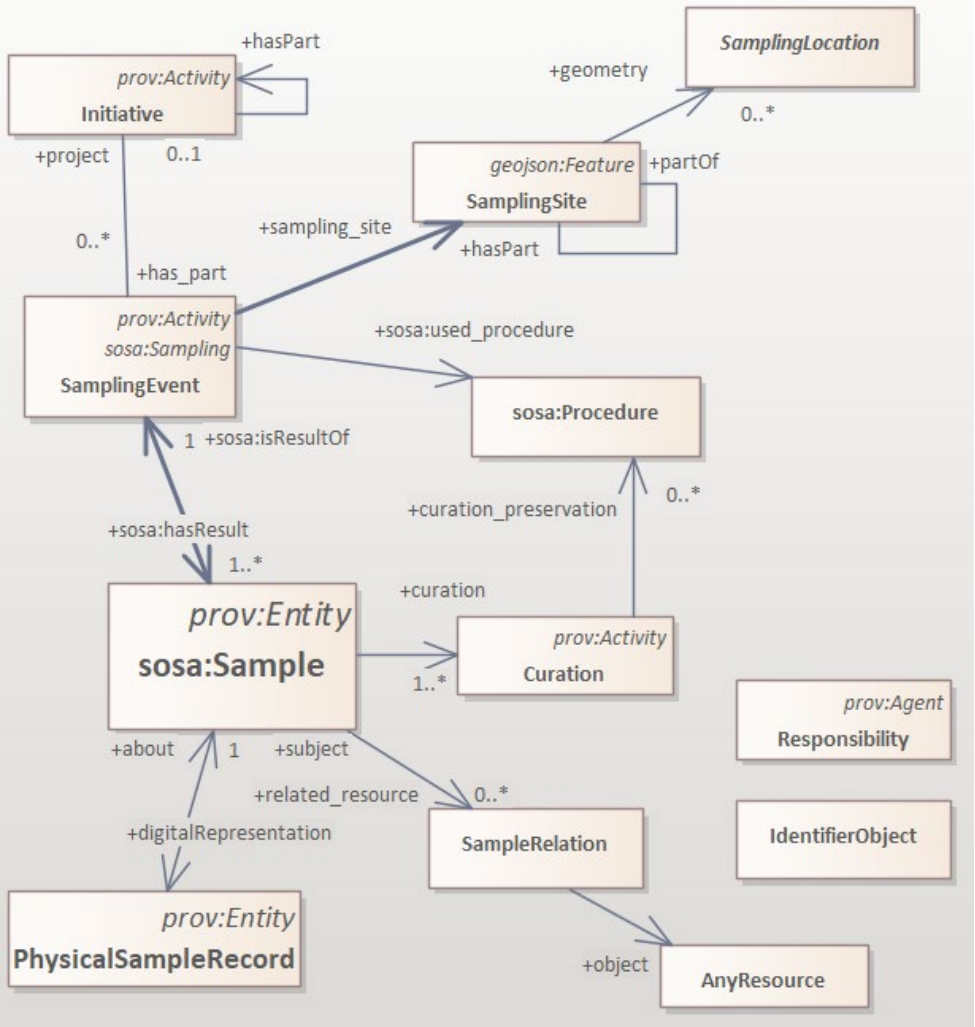

Figure 9. Conceptual framework for material sample description. Class names in italics in the upper right of boxes are parent classes. Namespace abbreviations: prov- https://www.w3.org/TR/prov-o/, geojson – GeoJSON, sosa: https://www.w3.org/ns/sosa/.

[43] https://www.w3.org/TR/prov-o/

activity through some role. IdentifierObject binds an identifier string with an authority agent that assures uniqueness of identifiers within some scope; any element that might be reused or referenced from an external source has an identifier. The content of these elements is outlined in the Sample description information model section, above. The digital representation of the sample is a MaterialSampleRecord (a Digital Object), which is expected to include elements documenting all of the classes associated with the material sample.

The iSamples cyberinfrastructure uses the JSON format for serializing sample descriptions for information interchange between iSamples instances. The conceptual model is implemented using LinkML tools[44]. The entity and property schema is described using a simple key-value structure encoded with YAML[45] syntax. This schema definition file includes documentation for each entity and property (see Appendix 1. YAML schema definition file used by LinkML tools). The LinkML tools generate the JSON schema that can be used to validate metadata instances, as well as generating HTML web pages with documentation for the schema[46]. For version one, implementation focus has been on simplicity and compatibility with legacy sample descriptions. The use of complex property values has been minimized, favoring free text content. Semantic interoperability is supported using URIs for responsible parties, iSamples controlled vocabulary terms, and optionally for keywords, as well as identifiers for SamplingEvents, SamplingSites, and Curation events. Properties that have text data type values can be populated with URIs, but the current user interface and search indexing will not take advantage of the linkages.

Relationships to other schemes and metadata formats

Schema.org

A proposed serialization of the iSamples metadata content using the Schema.org vocabulary is included in Appendix 2. Schema.org does not have properties that map to the iSamples sample curation property, or to the 'authorized_by' property related to permitting for a sampling event. Inclusion of this information can be done using the iSamples property URI as the key in the schema.org JSON. Schema.org is not designed for describing material samples, so usage of some properties does not follow the 'expected' domain or range for an entity-property pair. These interpretations of property semantics do not result in schema.org validation errors[47] because of the loose 'domainIncludes' and 'rangeIncludes' predicates used to define the schema.org RDF schema[48], but various warnings are raised by the validator, flagging the unexpected usages.

IGSN and DataCite

The IGSN sample registration system has been migrated to use digital object identifiers (DOIs) managed by DataCite. Thus, all IGSN sample metadata will be serialized using the DataCite metadata schema. A DataCite workgroup has developed recommendations for serializing material sample descriptions using

[44] https://linkml.io/linkml/
[45] https://yaml.org/spec/1.2.2/
[46] https://isamplesorg.github.io/metadata/
[47] See https://validator.schema.org/
[48] https://schema.org/version/latest/schemaorg-current-https.jsonld

the DataCite metadata schema[49]. A mapping from the original IGSN registration and description metadata scheme to the iSamples core scheme is provided in Appendix 3. A mapping from the DataCite v4.3 XML metadata schema to iSamples JSON is included in Appendix 4. Note that there are some minor discrepancies between the DataCite XML[50] and JSON[51] serialization. The XML schema appears to be more widely used, but the JSON schema offers some useful open-world flexibility.

TDWG Minimum Information about a Digital Specimen (MIDS)

Minimum Information about a Digital Specimen (MIDS)[52] is a specification defining information elements expected in a digital representation of a 'physical specimen', interpreted here to be equivalent to 'material sample'. The MIDS specification defines four levels of content for sample description, level 0 to level 3. The scope is focused on curated natural science collections (typically in museums), with the goal of making descriptions of physical objects available on the World Wide Web. Higher content levels include more detail in the sample description. This digital specimen object corresponds to the iSamples sample description record. A mapping from the MIDS digital specimen, levels 0-3 (including some yet-to-be adopted properties) is included in Appendix 5.

Digital extended specimen

The Digital extended specimen (DES) is a web-accessible representation of the digital assets related to a material sample (physical specimen) (Hardisty et al., 2022). These assets can include observations and measurements from the specimen, its sample context (sampled feature), or analyzed derivatives from an original material sample. This concept has evolved into the Open Digital Specimen (OpenDS) as the basis for the Distributed System of Scientific Collections (DiSSCo)[53] activity in the pan-European Research Infrastructure. The DiSSCo vision is a seamless and standardized digital representation of material samples in Natural Science Collections. While the DiSSCo vision parallels the goals of the iSamples project, it is founded in the Global Biodiversity Community, presenting opportunities for harmonization and interoperability of these efforts to enable integration of samples across these domains.

The OpenDS metadata schema[54] incorporates elements from Darwin Core, with numerous extensions specific to material samples. Appendix 6 presents a mapping from iSamples JSON metadata to the OpenDS metadata schema.

Testing

The JSON schema for iSamples material sample description has been tested via the process of mapping content from partner systems into the iSamples Central aggregator. The transformations are

[49] https://support.datacite.org/docs/igsn-id-metadata-recommendations
[50] https://github.com/datacite/schema/blob/master/source/meta/kernel-4.3/metadata.xsd
[51] https://github.com/datacite/schema/blob/master/source/json/kernel-4.3/datacite_4.3_schema.json
[52] https://github.com/tdwg/mids/blob/working-draft/current-draft%20/MIDS-definition-v0.17-13Jul2023.md
[53] https://www.dissco.eu/
[54] https://github.com/DiSSCo/openDS/blob/master/data-model/digitalobjects/0.1.0/digital-specimens/schema/digital-specimen.json

implemented in python[55]. Some minor issues have emerged, leading to updates in the core vocabulary or the metadata schema. The version 1 release tags are intended to provide a stable snapshot that other parties can use to implement a sample description and registration that can be integrated with the iSamples infrastructure.

References

Aryani, A., Fenner, M., Manghi, P., Mannocci, A. and Stocker, M, 2020, Open Science Graphs Must Interoperate!: In ADBIS, TPDL and EDA 2020 Common Workshops and Doctoral Consortium, Lyon, France, 2020, https://doi.org/10.1007/978-3-030-55814-7_16.

Damerow, JE, Varadharajan, C, Boye, K, Brodie, EL, Burrus, M, Chadwick, KD, Crystal-Ornelas, R, Elbashandy, H, Alves, RJE, Ely, KS, Goldman, AE, Haberman, T, Hendrix, V, Kakalia, Z, Kemner, KM, Kersting, AB, Merino, N, O'Brien, F, Perzan, Z, Robles, E, Sorensen, P, Stegen, JC, Walls, RL, Weisenhorn, P, Zavarin, M and Agarwal, D. 2021. Sample Identifiers and Metadata to Support Data Management and Reuse in Multidisciplinary Ecosystem Sciences. Data Science Journal, 20: 11, pp. 1–19. DOI: https://doi.org/10.5334/dsj-2021-011

Hardisty , Alex R., Ellwood, Elizabeth R., Nelson, Gil, Zimkus, Breda, Buschbom, Jutta, Addink, Wouter, Rabeler, Richard K., Bates, John, Bentley, Andrew, Fortes, José A. B., Hansen, Sara, Macklin, James A., Mast, Austin R., Miller, Joseph T., Monfils, Anna K., Paul, Deborah L., Wallis, Elycia, and Webster, Michael, 2022, Digital Extended Specimens: Enabling an Extensible Network of Biodiversity Data Records as Integrated Digital Objects on the Internet: BioScience 72: 978–987, https://doi.org//10.1093/biosci/biac060.

Haston, et al., 2023-07-13, The Minimum Information about a Digital Specimen (MIDS) specification for Natural Science Collections v0.17: https://github.com/tdwg/mids/blob/working-draft/current-draft%20/MIDS-definition-v0.17-13Jul2023.md.

Janowicz, Krzysztof, Haller, Armin, Cox, Simon J.D., Phuoc, Danh Le, and Lefrançois, Maxime, 2019, SOSA: A lightweight ontology for sensors, observations, samples, and actuators: eWeb Semantics: Science, Services and Agents on the World Wide Web v. 56, p 1–10, https://doi.org/10.1016/j.websem.2018.06.003.

Lebo, Timothy, Sahoo, Satya, and McGuinness, Deborah , 30 April 2013, PROV-O: The PROV Ontology: W3C Recommendation, http://www.w3.org/TR/2013/REC-prov-o-20130430/.

Lehnert, Kerstin, Sarah Ramdeen, Hong Cui, Neil Davies, John Deck, Eric Kansa, Sarah Kansa, John Kunze, Christopher Meyer, Thomas Orrell, Stephen Richard, Rebecca Snyder, Ramona Walls, & David Vieglais. (2021). iSamples (Internet of Samples): Cyberinfrastructure to support transdisciplinary use of material samples. EarthCube 2021 Annual Meeting. Zenodo. https://doi.org/10.5281/zenodo.4923861

Nickel, Ernest H., 1995, The definition of a mineral: The Canadian Mineralogist. 33 (3): 689–90.

Pesant, Stéphane, Goodwin, K. D., Magalhães, Catarina, Lacoursière-Roussel, Anaïs, eds., 2023, Marine Microbiomes: Towards Standard Methods and Best Practices: Frontiers in Microbiology, Frontiers in Marine Science, DOI 10.3389/978-2-8325-4131-9.

Schleidt, Katharina, and Rinne, Ilkka, eds., 2023, OGC Abstract Specification Topic 20: Observations, measurements and samples, version 3.0.0: Open Geospatial Consortium, document: 20-082r4, http://www.opengis.net/doc/as/om/3.0.

[55] https://github.com/isamplesorg/isamples_inabox/tree/develop/isamples_metadata. Note this is an active development environment and things can change at any time. See the …transformer.py files.

Appendix 1. YAML schema definition file used by LinkML tools

```
id: https://w3id.org/isample/schema/1.0
name: materialSample
title: Schema for documenting material samples
description: |
  SMR 2022-10-07. Schema for iSamples sample registry integration. Updated from 0.2 by synchronizing the vocabulary enumerations, change 'id' to '@id' and 'schema' to '$schema'. Schema name is iSamplesSchemaCore1.0.json. Target JSON schema version is https://json-schema.org/draft/2019-09/schema. SMR 2023-03-17. Move authorized_by into SamplingEvent, change keywords to list of Keyword objects with string, uri, scheme and scheme URI. Reorder elements in slot list. Update scheme URI to 1.0 2023-06-13 SMR add project property on SamplingEvent, with text or URI value. 2024-01-17 SMR run linkml-lint from current linkml version before rebuilding JSON schema. 2024-04-19. DV ran yaml linter on the file to clean up formatting, put hard return in text bodies. Update schema URI to dereference with w3id.
license: https://creativecommons.org/publicdomain/zero/1.0/
version: "20240117"
prefixes:
  linkml: https://w3id.org/linkml/
  isam: https://w3id.org/isample/schema/1.0/
  dct: http://purl.org/dc/terms/
  mat: https://w3id.org/isample/vocabulary/material/1.0/
  sf: https://w3id.org/isample/vocabulary/sampledfeature/1.0/
  skos: http://www.w3.org/2004/02/skos/core#
  spec: https://w3id.org/isample/vocabulary/sampleobjecttype/1.0/
  w3cpos: http://www.w3.org/2003/01/geo/wgs84_pos#
  rdfs: http://www.w3.org/2000/01/rdf-schema#
  xsd: http://www.w3.org/2001/XMLSchema#
  sdo: http://schema.org/
  skos_concept: http://www.w3.org/2004/02/skos/core#Concept
  datacite: http://datacite.org/schema/kernel-4/
default_prefix: isam

imports:
  - linkml:types

classes:
  MaterialSampleRecord:
    description: This is a data object that is a digital representation of a material sample, and thus shares the same
        identifier as the physical object. It provides descriptive properties for any iSamples material sample, URI for
        the metadata record is same as URI for material sample-- digital object is considered twin of physical object,
        a representation. IGSN is recommended. Must be a URI that can be dereferenced on the web.
    slots:
      - sample_identifier
      - label
      - description
      - alternate_identifiers
      - produced_by
      - sampling_purpose
      - has_context_category
      - has_material_category
      - has_sample_object_type
      - keywords
      - related_resource
      - complies_with
      - dc_rights
```

```
- curation
- registrant

  slot_usage:
    label:
      required: true
    sample_identifier:
      required: true
```

Agent:

```
  description: object to represent a person who plays a role relative to sample collection or curation.
  slots:
    - name
    - affiliation
    - contact_information
    - identifier
    - role
```

Keyword:

```
  description: a simple string (tag) useful for guiding discovery of the sample, or a scoped name that associates
      the keyword string with a scheme name, and optional scheme URI and value URI. The scoped name
      keywords can be used to categorize the sample using a formal vocabulary. This mechanism provides the
      hook to use extension vocabularies for domain specific concepts extending the context, material type and
      object type high-level vocabularies.
  slots:
    - keyword
    - keyword_uri
    - scheme_name
    - scheme_uri
```

IdentifierObject:

```
  description: object to represent an identified concept; property values from vocabularies with resolvable
      identifiers for vocabulary terms. Either a label or identifier is required, to allow for cases that a URI is not
      available.  Data validation process will need to be used for properties that require an identified term value
      (e.g. has_context_category, has_material_category, has_sample_object_type).
  slots:
    - label
    - identifier
    - scheme_name
```

SamplingSite:

```
  description: documentation of the site where the sample was collected, with place name(s) and a geospatial
      location
  slots:
    - identifier
    - description
    - label
    - sample_location
    - place_name
    - is_part_of
```

SamplingEvent:

```
  description: Information about the event resulting in the creation of the material sample. Include information
      about permitting in the authorized_by property. The sampling procedure should be described in the
```

description. If any special protocols were followed in the sampling procedure, they should be documented using the MaterialSampleRecord/complies_with property.
slots:
- identifier
- label
- description
- has_feature_of_interest
- project
- responsibility
- result_time
- sampling_site
- authorized_by

GeospatialCoordLocation:
description: A physical location in decimal degrees using EPSG:4326. Could be sample point location, or the centroid of a sampling area. Elevation is specified as a string that should include the measure, units of measure, and the vertical reference system (e.g. 'above mean sea level', 'below ground surface', 'below sea floor'...)
slots:
- elevation
- latitude
- longitude
- obfuscated

MaterialSampleCuration:
description: Information about the current storage of sample, access to sample, and events in curation history. Curation as used here starts when the sample is removed from its original context, and might include various processing steps for preservation. Processing related to analysis preparation such as crushing, dissolution, evaporation, filtering are considered part of the sampling method for the derived child sample.
slots:
- identifier
- access_constraints
- curation_location
- description
- label
- responsibility

SampleRelation:
description: semantic link to other samples or related resources.
slots:
- description
- label
- relationship
- target

slots:
alternate_identifiers:
range: IdentifierObject
description: one or more identifiers used to identify the sample in other contexts. In this context, the identifier property and scheme_name should be required.
multivalued: true

access_constraints:
range: string

description: Cultural, legal, or other policy issues that bear on access to view, borrow, or subsample a sample or visit a sampling site.
multivalued: true

affiliation:
range: string
domain: Agent
description: Name of an organization that employes or sponsors a person for the role they play relative to sample collection or curation.

authorized_by:
range: string
domain: MaterialSampleRecord
description: a list of permits or other formal permission documents under which the sample was collected. Use to cite legal documents authorizing sample collection.
multivalued: true

contact_information:
range: string
domain: Agent
description: e-mail address, telephone number, mailing address, or web site through which an agent can be contacted. Something that should be stable and reliable.

curation:
range: MaterialSampleCuration
description: Information about the current storage of sample, access to sample, and events in curation history. Curation as used here starts when the sample is removed from its original context, and might include various processing steps for preservation. Processing related to analysis preparation such as crushing, dissolution, evaporation, filtering are considered part of the sampling method for the derived child sample.

curation_location:
range: string
description: Information about where and how the sample is currently stored.

complies_with:
range: string
domain: MaterialSampleRecord
description: a list of policies, recommendations, best practices (etc.) that have been followed in the collection and curation of the sample. Use to address CARE principles.
multivalued: true

dc_rights:
range: string
description: a statement about various property rights associated with the resource, including intellectual property rights. Recommended practice is to refer to a rights statement with a URI. If this is not possible or feasible, a literal value (name, label, or short text) may be provided.
close_mappings: [dct:rights]

description:
range: string
description: Free text description of the subject of a triple.
close_mappings: [dct:description]

elevation:
range: string
domain: GeospatialCoordLocation
description: Should be a number and Unit of measure, and the vertical reference system. e.g. 401 m above mean sea level.

identifier:
range: uriorcurie
domain: IdentifierObject

description: a string that identifies some resource. Expectation is that it is a URI that is resolvable to some useful representation of the resource.

has_context_category:
range is an instance of at least one skos:Concept
with skos:inScheme sf:sampledfeaturevocabulary ;
range: IdentifierObject
domain: MaterialSampleRecord
description: Top level context, based on the kind of feature sampled. Specific identification of the sampled feature of interest is done through the SamplingEvent/Feature of Interest property. At least one value is an instance of skos:Concept from the iSamples Sampled Feature vocabulary.
multivalued: true

has_feature_of_interest:
range: string
domain: SamplingEvent
description: What does the sample represent.

has_material_category:
range is an instance of at least one skos:Concept
with skos:inScheme = mat:materials Vocabulary
domain: MaterialSampleRecord
range: IdentifierObject
description: The kind of material that constitutes the sample. At least one value is an instance of skos:Concept from the iSamples Material Type Vocabulary; extension vocabularies can be used for more precise categorization.
multivalued: true

has_sample_object_type_:
range is an instance at least one skos:Concept
with skos:inScheme = spec:material_sample_type vocabulary
domain: MaterialSampleRecord
description: The kind of object the material sample is. At least one value is an instance of skos:Concept from the iSamples Material Sample Object Type Vocabulary; extension vocabularies can be used for more precise categorization.
range: IdentifierObject
multivalued: true

is_part_of:
range: uriorcurie
domain: SamplingSite
description: A link from a SamplingSite to a larger sampling site that spatially contains the source SamplingSite. The container site is specified by its identifier.
close_mappings: [sdo:isPartOf]
multivalued: true

keywords:
range: Keyword
domain: MaterialSampleRecord
description: free text terms or formal categories associate with sample to support discovery. As in DataCite metadata, each keyword is a separate element. Multiple keywords should NOT be included as a comma-delimited list.
close_mappings: [sdo:keywords, datacite:subjects]
multivalued: true

keyword:
range: string
domain: Keyword
description: string label for a concept that characterizes the sample in some respect
multivalued: false

keyword_uri:
range: string
domain: Keyword
description: an identifier for the keyword concept
multivalued: false

label:
range: string
description: a human intelligible string used to identify a thing, i.e. the name to use for the thing; should be unique in the scope of a sample collection or dataset.
close_mappings: [rdfs:label]

latitude:
range: decimal
domain: GeospatialCoordLocation
description: angular coordinate measured positive north from the equator.

longitude:
range: decimal
domain: GeospatialCoordLocation
description: angular coordinate measured positive eastward from the prime meridian.

sample_location:
range: GeospatialCoordLocation
domain: SamplingSite
description: geospatial location of site; required default is WGS84 latitude, longitude in decimal degrees. Elevation as a string with number, unit of measure, and datum.

name:
range: string
domain: Agent
description: string, name of a person. Format {firstName MI lastName}.

obfuscated:
range: boolean
domain: GeospatialCoordLocation
description: Flag to indicate that accuracy of provided coordinate location has been reduced (obfuscated) because the exact location is restricted information.
ifabsent: "False"

place_name:
range: string
domain: SamplingSite
description: one or more names by which the sampling site is known.
multivalued: true

project:
range: string
domain: SamplingEvent
description: the name or identifier for a project that is the sponsor for a samplingEvent. Use to link samples collected in the context of a particular project. We do not have a preferred project model at this point; the RAID effort (ISO23527, https://www.raid.org.au/) will hopefully mature into a registry we can use for project identification and metadata. Property value should be a name or a resolvable URI.

produced_by:
range: SamplingEvent
domain: MaterialSampleRecord
description: object that documents the sampling event--who, where, when the material sample was obtained

registrant:
range: Agent
domain: MaterialSampleRecord

description: identification of the agent that registered the sample, with contact information. Should include person name and affiliation, or position name and affiliation, or just organization name. e-mail address is preferred contact information.

related_resource:
range: SampleRelation
multivalued: true
description: link to related resource with relationship property to indicate nature of connection. Target should be identifier for a resource.

relationship:
range: string
description: term to identify realationship between host sample and the sample relation target. Should be controlled vocabulary (ScopedName). for now start with string, 'derivedFrom'.

role:
range: string
domain: Agent
description: term that specifies how a person is related to a sample curation or collection activity.

sample_identifier:
range: string
domain: MaterialSampleRecord
description: URI that identifies the material sample described by this record
close_mappings: [sdo:identifier]

sampling_purpose:
range: string
domain: MaterialSampleRecord
description: term to specify why a sample was collection.

scheme_name:
range: string
description: string label for the concept scheme or vocabulary within which the keyword concept is defined. The scheme name can be used to scope keyword to particular categories, e.g. 'CGI rock classification', 'GCMD keywords', 'Getty AAT'
multivalued: false

scheme_uri:
range: string
domain: Keyword
description: Identifier for the concept scheme or vocabulary within which the keyword concept is defined.
multivalued: false

target:
range: uriorcurie
description: identifier for the target resource in the relationship. Start with string, should be Identifier object.

responsibility:
range: Agent
description: Agent object includes person name, person identifier, affiliation, or organization name, their role relative to the parent element.
multivalued: true

result_time:
range: date
domain: SamplingEvent
description: Date on which the sample was collected.

sampling_site:
range: SamplingSite
domain: SamplingEvent
description: object that identifies the place where the sample was collected

Schema mapping appendices

Mapping to several schema for describing material samples that are in use or in development is included in appendices 2 to 6. Rather than present these in one large and rather sparse table, each mapping is in a separate table in its own file, and only includes the elements in the target scheme that are mapped to the iSamples scheme. Where there is content in a target scheme that does not fit well into the iSamples scheme, the field is highlighted in gray.

Appendix 2

Mapping from iSamples to Scheme.org. Attached file. Source is draft schema.org implementation from the iSamples project.[56] See note about schema.org implementation at end of table.

iSamplesSchemaCore1.0	iSamples notes	Schema.org	Schema.org note
-- affiliation identifier	Add in iSamples Agent/affiliation value string.	sdpublisher/affiliation/ identifier	
-- digital object identifier	identifier for the metadata record	@id	
-- metadata license		sdLicense	if dc_rights statement is about the metadata
-- Agent identifier scheme URI	no iSamples property	about/event/participant [roleName=collector]/identifier@propertyID	
$schema	identifier for the schema for this metadata ('digital sample' document). iSamples sample descriptions are implemented with a JSON serialization, so this should be an identifier for a JSON schema	dcterms:conformsTo	
alternate_identifiers	other identifiers can be provided with label, scheme_name (authority), and identifier string.	about/identifier	alternate identifiers for the sample described by this record. Note that about/id is the URI for the sample.
alternate_identifiers/ identifier		about/identifier/value	
alternate_identifiers/label		about/identifier/ description	

[56] https://github.com/isamplesorg/metadata/blob/schema.orgTest/notes/schemaOrg/iSamplesSchema.orgScheme.json

iSamplesSchemaCore1.0	iSamples notes	Schema.org	Schema.org note
alternate_identifiers/scheme_name		about/identifier/ propertyID	
complies_with	a list of policies, recommendations, best practices (etc.) that have been followed in the collection and curation of the sample. If any special protocols were followed in the sampling procedure, they should be documented here, with prefix 'Sampling protocol:'.	about/ethicsPolicy	
curation	Information about the current storage of sample, access to sample, and events in curation history. Curation as used here starts when the sample is removed from its original context, and might include various processing steps for preservation. Processing related to analysis preparation such as crushing, dissolution, evaporation, filtering are considered part of the sampling method for the derived child sample.	about/curation	
curation/access_constraints	cultural, legal or other policy issues that bear on access to view, borrow, or subsample a sample (Curation)	about/curation/ conditionsOfAccess	
curation/curation_location	Information about where and how the sample is currently stored.	about/curation/location	
curation/description		about/curation/ description	If the schema.org curation event includes subEvents, concatenate information from those into isample:curation/description
curation/label		about/curation/name	
curation/responsibility [role='classification']	unique identifier for the person, people, groups, or organizations responsible for assigning the scientific name to the subject. Include this as a curation/responsibility	about/curation/ maintainer[roleName = classification]	
curation/responsibility.Agent	Agents associated with sample curation, Need mapping from iSample roles to DataCite contributorType	about/curation/ maintainer/Role	person or organization, implemented as sdo:Role
curation/responsibility/ affiliation	If a person, can have an organization affiliation	about/curation/ maintainer/participant/ affiliation	if the participant is a Person.

iSamplesSchemaCore1.0	iSamples notes	Schema.org	Schema.org note
curation/responsibility/ contact_information		about/curation/ maintainer/participant/ pointOfContact	
curation/responsibility/ identifier	Identifier for person or organization. only one identifier in iSamples. An identifier for the institution having custody of the object(s)	about/curation/ maintainer/participant/ identifier	
curation/responsibility/name	Person or organization name	about/curation/ maintainer/participant/ name	
curation/responsibility/role	roles should be restricted to those relevant to sample curation	about/curation/ maintainer/roleName	
dc_rights	A statement about various property rights associated with the resource, including intellectual property rights. Recommended practice is to refer to a rights statement with a URI. If this is not possible or feasible, a literal value (name, label, or short text) may be provided.	about/curation/ conditionsOfAccess or /sdLicense	if the rights pertain to the sample, goes in conditionsOfAccess; if the rights pertain to the sample metadata record, put in sdLicense.
description	free text description of the physical sample, and any related information. Include as much detail as will be useful for users to find the sample and understand its context.	about/description	
has_context_category/ scheme_name	https://w3id.org/isample/vocabulary/sampledfeature/1.0/sampledfeaturevocabulary	category/inDefinedTermSet/name = iSamples SampledFeature type	
has_context_category/ identifier	use URI from isamples sampledFeature vocabulary	category[inDefinedTermSet/ name = iSamples SampledFeature type]/id	
has_context_category/label	For context using controlled vocabulary, the dataCite subject is the identifier/label, for keywords, it is the keyword	category[inDefinedTermSet/ name = iSamples SampledFeature type]/name	
has_material_category/ identifier	use URI from iSamples material type vocabular	category[inDefinedTermSet/ name = iSamples Material type]/id	
has_material_category/ scheme_name	use https://w3id.org/isample/vocabulary/material/1.0/materialsvocabulary	category/inDefinedTermSet/name = iSamples material type]	
has_material_category/label	For material using controlled vocabulary, the dataCite subject is the identifier/label, for keywords, it is the keyword	category[inDefinedTermSet/ name = iSamples Material type]/name	

iSamplesSchemaCore1.0	iSamples notes	Schema.org	Schema.org note
has_material_sample_object_type	specify the kind of object that the material sample is, use iSamples Material Sample Object Type Vocabulary, with ResourceTypeGeneral = PhysicalObject. the label for the Material Sample Object Type is the text value for dataCite ResourceType	about/additionalType	implement with schema:additionalType
has_ material_sample_object_type / identifier		about/additionalType	The URI for term. At least one URI from iSamples vocabulary is expected
has_ material_sample_object_type /label	repeat label and uri in subjects for datacite	about/additionalType	Label associated with material sample object type.
keywords/keyword	An identifier for the nomenclatural (not taxonomic) details of a scientific name.	keywords/{string} or keywords/name	
keywords/keyword_uri		keywords/identifier	
keywords/scheme_name		keywords/ inDefinedTermSet/name	
keywords/scheme_uri	not included for controlled vocabulary terms	keywords/ inDefinedTermSet/ identifier	
label	a human intelligible string used to identify the sample; i.e. the name to use for the sample; should be unique in the scope of a sample collection. This will typically be a sample identiifer or label assigned by the original collector	about/name	
produced_by/SamplingEvent	object that documents the sampling event--who, where, when the material sample was obtained	about/event	object that documents the sampling event--who, where, when the material sample was obtained. Implements iSamples SamplingEvent
produced_by/SamplingEvent/ responsibility/affiliation		about/event/participant [roleName={role}]/affiliation	
produced_by/SamplingEvent/ responsibility/contact_information	any agent can have an e-mail or telephone number for contacting	about/event/participant [roleName={role}]/contactPoint	
produced_by/SamplingEvent/ responsibility/identifier		about/event/participant [roleName={role}]/identifier	
produced_by/SamplingEvent/ responsibility/name		about/event/participant [roleName={role}]/name	

iSamplesSchemaCore1.0	iSamples notes	Schema.org	Schema.org note
produced_by/SamplingEvent/ responsibility/role		about/event/participant [roleName={role}]	
produced_by/SamplingEvent/ responsibility[roles other than 'collector', 'funder'],	Agents associated with sampling event, agent/role is subclass of 'contributor'. Need mapping from iSample roles to DataCite contributorType	about/event/participant	
produced_by/SamplingEvent/authorized_by	a list of permits or other formal permission documents under which the sample was collected. Use to cite legal documents authorizing sample collection.	http://resource.isamples.org/schema/1.0/ authorized_by	no schema.org property
produced_by/SamplingEvent/description		about/event/description	
produced_by/SamplingEvent/has_feature_of_interest	The specific thing that was sampled; should be an instance of the type identified by the has_context_category property	about/event/about	What was sampled. implements iSample has_feature_of_interest. string or scopedName
produced_by/SamplingEvent/identifier		about/event/@id	A string that identifies the samplingEvent. Use to gather all samples from a single sampling event.
produced_by/SamplingEvent/label		about/event/name	
produced_by/SamplingEvent/project	the name or identifier for a project that is the sponsor for a samplingEvent. Use to link samples collected in the context of a particular project. Property value should be a name or a resolvable URI. If funding was through an identified grant/award/contract instrument, this is the label associated with that instrument, If a grant/contract/award was the instrument, this would be the award title.	about/event/organizer	string or scopedName

iSamplesSchemaCore1.0	iSamples notes	Schema.org	Schema.org note
produced_by/SamplingEvent/responsibility	the responsibility is an agent-- could be person or organization, with role = 'collector'. Other agents associated with the sampling event could be included, e.g. with roles like 'sponsor', 'funder'....	about/event/participant	Agent with person name and affiliation, or organization name, and their role relative to the sampling event.
produced_by/SamplingEvent/responsibility[role=collector]/ identifier	only one identifier associated with person in this role. iSamples identifier value is a string.	about/event/participant [roleName=collector]/identifier	
produced_by/SamplingEvent/responsibility[role=collector]/affiliation	only one affiliation associated with person in this role.	about/event/participant [roleName=collector]/affiliation/name	
produced_by/SamplingEvent/responsibility[role=collector]/name	string, name of a person. Format: {firstName MI lastName}.	about/event/participant [roleName=collector]/name	participant has a Role, Agent can be person or organization
produced_by/SamplingEvent/responsibility[Role=funder]/ identifier	identifier for funding agent	about/event/participant [roleName=funder]/identifier	
produced_by/SamplingEvent/responsibility[Role=funder]/ name	Name person or organization that supported the activity;	about/event/participant [roleName=funder]/name	
produced_by/SamplingEvent/result_time	Date on which the sample was collected.	about/event/endDate	when the sampling event was done.
produced_by/SamplingEvent/sampling_site	one WGS 84 Decimal Degree lat /long point location is required. Schema doesn't accomodate other coordinate locations. This corresponds to DataCite geoLocationPoint. Multiple places might be specified; iSamples is limited to a single point location -- doesn't make sense for a sample to come from more than one point location...	about/event/location	
produced_by/SamplingEvent/sampling_site/sample_location/latitude	sample location point coordinate	about/event/location/geo/latitude	
produced_by/SamplingEvent/sampling_site/sample_location/longitude	sample location point coordinate	about/event/location/geo/longitude	

iSamplesSchemaCore1.0	iSamples notes	Schema.org	Schema.org note
produced_by/SamplingEvent/sampling_site/description	description of location	about/event/location/ description	
produced_by/SamplingEvent/sampling_site/label	text to identify a location for human users;	about/event/location/ name	
produced_by/SamplingEvent/sampling_site/sample_location/elevation	Should be a number and Unit of measure, and the vertical reference system. e.g. 401 m above mean sea level.	about/event/location/geo/elevation	
produced_by/SamplingEvent/sampling_site/sample_location/is_part_of	A link from a SamplingSite to a larger sampling site that spatially contains the source SamplingSite. The container site is specified by its identifier.	about/event/location/ containedInPlace/ identifier or name	
produced_by/SamplingEvent/sampling_site/sample_location/obfuscated	Flag to indicate that accuracy of provided coordinate location has been reduced (obfuscated) because the exact location is restricted information. Default is 'false'	--	no schema.org property
registrant [role=registrant]/name/	identification of the agent that registered the sample, with contact information. Role = 'registrant'	sdpublisher/name	
registrant/ contact_information		sdpublisher/affiliation/ contactPoint	
registrant/affiliation		sdpublisher/affiliation/ name	
registrant/identifier		sdpublisher/identifier	
related_resource	link to another resource via identifier for the target resource, with a relationship type, label, and description. Label and description do not have matching properties in DataCite.	about/relatedLink	
related_resource/description		about/relatedLink/ description	

iSamplesSchemaCore1.0	iSamples notes	Schema.org	Schema.org note
related_resource/label	a human intelligible string used to identify a thing, i.e. the name to use for the thing; should be unique in the scope of a sample collection or dataset.	about/relatedLink/name	
related_resource/relationship	nature of relationship between sample and resource that is the target of the relationship. See vocabulary in development: https://docs.google.com/spreadsheets/d/1yk3ZEWzEc1cXBOwb-foCtq3isaBBa95Rv8eA6TIyieM	about/relatedLink/ linkRelationship/name	string or scopedName
related_resource/target		about/relatedLink/ linkRelationship/target/url	
sample_identifier	unique identifier for the physical object, ideally a URI that is physically attached to the material sample object, an IGSN or ARK	about/@id	@id identifies the subject of the triples; should be a resolvable URI.

The mapping assumes a structure for the Schema.org JSON-LD like this:


```
{
  "@context": {
     "@vocab": "https://schema.org/",  "dcterms": "http://purl.org/dc/terms/", "id": "@id", "type": "@type",
          curation":"http://resource.isamples.org/schema/1.0/MaterialSampleCuration"    },
  "type": "DigitalDocument",
  "id": "https://isamples.org/igsn-id/metadata",
  "dcterms:conformsTo": "http://resource.isamples.org/schema/1.0/schema.org",
  "sdPublisher": {
     "type": "Organization",
     "id": "https://ror.org/someorg7523824",
     "name": "This is the metadata publisher—where the record was published",
     "contactPoint": {
        "type": "ContactPoint",
        "email": "curator@gsample.org"
     }
  },
  "about": {
     "type":  "Thing",
```

```
    "id": "https://igsn.org/slk222-046537",
    "additionalType": ["https://w3id.org/isample/vocabulary/materialsampleobjecttype/1.0/solidmaterialsample"],
    "name": "name for the material sample",
    "description": "description of material sample.",
    "identifier": [
      {
        "type": "PropertyValue",
        "name": "other identifiers for sample",
        "value": "smr2010-398"
      }
    ],
….. rest of record omitted.
```


In the first part of the record, the subject of the statements (type, dcterms:conformsTo, sdPublisher) is the metadata record, which as a type 'Digital Document'. In the 'about' object, the subject of the statements is the sample with identifier in the 'id' property. The body of the sample description is in this 'about' object. Since schema.org does not include a type for material object, the type 'Thing' is assigned, and the actual sample type is assigned from the iSamples Material Sample Object Type vocabulary in the 'additonalType' property. In many implementations, the part of the record about the metadata object is omitted, and only the 'about' object is provided.

Appendix 3

Mapping from iSamples to original IGSN registration and description schemes, based on IGSN github repository.[57]

iSamplesSchemaCore1.0	iSamples notes	**IGSN Element**	Definition
-- affiliation identifier	organization in an affiliation does not have an identifier in iSamples, could put in affiliation.	registrant/ affiliation/identifier	has identifier type
-- geometry spatial reference system	Not applicable, iSamples requires WGS84	geometry/sridType	controlled vocabulary; only value is '4326', apparently requiring use of WGS84?
-- location URI	An identifier for the geographical locality where the material sample was collected.	toponym/identifier	
-- metadata update date	Update date not currently included	logElement [event = updated] @timeStamp	
-- name identifier scheme name	no iSamples property.	collector/identifier/ identifierType	

[57] https://github.com/IGSN/metadata

iSamplesSchemaCore1.0	iSamples notes	IGSN Element	Definition
-- publication year	Time stamp for when the sample registration information is made public through the identifier authority; not currently included in iSamples metadata, use year of sample registration; If an embargo period has been in effect, use the date when the embargo period ends.	logElement [eventtype = registered] @timeStamp	Log of events relevant to the object. Events have eventType, timeStamp and comments. eventType in {submitted, registered, updated, deprecated, destroyed}
-- related identifier type	no iSamples equivalent. Default to PURL, unless you can determine its one of the others in the DataCite vocabulary.	relatedIdentifierType	Has relatedIdentifierType in {doi, handle, lsid, url, urn, igsn}
alternate_identifiers/ identifier		alternateIdentifier	
alternate_identifiers/ scheme_name		alternateIdentifier/ identifierType	
curation/ access_constraints	cultural, legal or other policy issues that bear on access to view, borrow, or subsample a sample (Curation)	sampleAccess	
description	free text description of the physical sample, and any related information. Include as much detail as will be useful for users to find the sample and understand its context.	description	Free text, anything else that might be useful to know about the sample at its 'birth'.
has_material_sample_object_type / label	repeat label and uri, from iSamples Material Sample Object Type vocabulary.	alternateResourceType	map to iSamples vocabulary
keywords/keyword	An identifier for the nomenclatural (not taxonomic) details of a scientific name.	alternateResourceType | alternateMaterial	include verbatim as keywords; map resourceType and material to iSamples vocabulary if possible.
keywords/keyword [scheme_name= IGSN_description]		resourceType, material	controlled vocabulary {feature, sample, collection} and {uses ODM2 medium vocabulary}
label	a human intelligible string used to identify the sample; i.e. the name to use for the sample; should be unique in the scope of a sample collection. This will typically be a sample identifier or label assigned by the original collector	name	Text string for people to understand what is identified. What would typically be presented in a user interface.
produced_by/ SamplingEvent/ responsibility	the responsibility is an agent-- could be person or organization, with role = 'collector'. Other agents associated with the sampling event could be included, e.g. with roles like 'sponsor', 'funder'....	collector	Who collected the sample. has identifier, identifierType, name, affiliation that map to iSamples Agent

iSamplesSchemaCore1.0	iSamples notes	IGSN Element	Definition
produced_by/ SamplingEvent/ responsibility [role=collector]/ affiliation	only one affiliation associated with person in this role.	collector/affiliation/ name & affiliation/identifier	has affiliation identifier
produced_by/ SamplingEvent/ responsibility [role=collector]/ identifier	only one identifier associated with person in this role. iSamples identifier value is a string.	collector/identifier	
produced_by/ SamplingEvent/ responsibility [role=collector]/name	string, name of a person. Format: {firstName MI lastName}.	collector/name	
produced_by/ SamplingEvent/ responsibility/identifier		contributor/identifier	
produced_by/ SamplingEvent/ responsibility/name		contributor/name	
produced_by/ SamplingEvent/ responsibility/role		contributor/ contributorType	Controlled Vocabulary
produced_by/ SamplingEvent/ responsibility[roles other than 'collector', 'funder'],	Agents associated with sampling event, agent/role is subclass of 'contributor'. Need mapping from iSample roles to DataCite contributorType	contributors	
produced_by/ SamplingEvent/ sampling_site/ place_name	iSamples can have multiple values.	toponym/name	
produced_by/ SamplingEvent/ sampling_site/ sample_location or place_name	for DCAT, use whichever is there	geoLocations	Where was the sample acquired relative to the Earth (or another celestial body...). Some samples might be 'non-geographic': mineral specimen, synthetic material. geometry types allowed: {LineString, MultiLineString, point, multipoint, polygon, multipolygon}. Allows different spatial reference systems
produced_by/ SamplingEvent/description		collectionMethod | alternateCollectionMethod	collectionMethod has controlled vocabulary, looks similar to SESAR
produced_by/ SamplingEvent/result_time	Date on which the sample was collected.	collectionTime	When was the sample collected. instant or interval.
produced_by/label && produced_by/ description && produced_by/ authorized_by	concatenate iSamples SamplingEvent label and description; include information in 'authorized_by'	collectionMethod	Term to categorize the process through which the sample was acquired as an independent object.

iSamplesSchemaCore1.0	iSamples notes	IGSN Element	Definition
registrant [role=registrant]/name/	identification of the agent that registered the sample, with contact information. Role = 'registrant'	registrant/ registrantName	allocating agent. appears in Registration and description, but labels are slightly different.
registrant/affiliation		registrant/ affiliation	
registrant/identifier		registrant/ nameIdentifier	name identifier has scheme in {ORCID, ISNI, VIAF, researcherID}
related_resource [relationship= 'hasOtherMetadata']		supplementalMetadata/ record	
related_resource/ relationship	nature of relationship between sample and resource that is the target of the relationship. See vocabulary in development: https://docs.google.com/spreadsheets/d/1yk3ZEWzEc1cXBOwb-foCtq3isaBBa95Rv8eA6TIyieM	relatedResourceIdentifier @relationType (relatedIdentifier/ relationType in Description schema)	controlled vocabulary
related_resource/target	identifier for the target resource in the relationship. Should be a resolvable URI.	relatedResourceIdentifier (relatedIdentifier in Description schema)	Identifiers of related resources. Has relatedIdentifierType in {doi, handle, lsid, url, urn, igsn}
sample_identifier	unique identifier for the physical object, ideally a URI that is physically attached to the material sample object, an IGSN or ARK	sampleNumber	The Identifier is a unique string that identifies a resource. At present, the only allowed value is an IGSN handle.

Appendix 4

Mapping from iSamples to DataCite v4.3 metadata scheme, based on DataCite IGSN recommendations.[58]

iSamplesSchemaCore1.0	iSamples notes	DataCite 4.3	Sub element	Definition
		affiliation		array
-- affiliation identifier			.	datacite also has affiliationIdentifier, affiliationIdentifierScheme, and schemeURI

[58] https://support.datacite.org/docs/igsn-id-metadata-recommendations

iSamplesSchemaCore1.0	iSamples notes	DataCite 4.3	Sub element	Definition
produced_by/SamplingEvent/ responsibility [role=collector]/affiliation	only one affiliation associated with person in this role.		name	string value
alternate_identifiers	other identifiers can be provided with label, scheme_name (authority), and identifier string. Label doesn't map to DataCite	alternateIdentifiers	.	
alternate_identifiers/ identifier			alternateIdentifier	
alternate_identifiers/ scheme_name			alternateIdentifierType	
produced_by/responsibility[roles other than 'collector', 'funder'],	Agents associated with sampling event, agent/role is subclass of 'contributor'. Need mapping from iSample roles to DataCite contributorType	Contributor	.	The institution or person responsible for managing, distributing, or otherwise contributing to the collection or curation of the sample
produced_by/SamplingEvent/ responsibility/name			name	
produced_by/SamplingEvent/ responsibility/affiliation			affiliation	
produced_by/SamplingEvent/ responsibility/identifier			nameIdentifier	
produced_by/SamplingEvent/ responsibility/role			contributorType	
curation/responsibility.Agent	Agents associated with sample curation, Need mapping from iSample roles to DataCite contributorType	contributor	.	
curation/responsibility/role	roles should be restricted to those relevant to sample curation		contributorType	DataCite terms for contributor roles
curation/responsibility/name	Person or organization name		name	
curation/responsibility/affiliation	If a person, can have an organization affiliation		affiliation	

iSamplesSchemaCore1.0	iSamples notes	DataCite 4.3	Sub element	Definition
curation/responsibility/identifier	Identifier for person or organization. only one identifier in iSamples. An identifier for the institution having custody of the object(s)		nameIdentifiers	
produced_by/SamplingEvent/responsibility	the responsibility is an agent-- could be person or organization, with role = 'collector'. Other agents associated with the sampling event could be included, e.g. with roles like 'sponsor', 'funder'….	creators	.	Array; The main researchers involved in producing the data, or the authors of the publication, in priority order.
produced_by/SamplingEvent/result_time	Date on which the sample was collected.	Date[dateType = 'Collected']	.	
description	free text description of the physical sample, and any related information. Include as much detail as will be useful for users to find the sample and understand its context.	Description [descriptionType = 'Abstract']	.	All additional information that does not fit in any of the other categories. May be used for technical information.
produced_by/label && produced_by/ description && produced_by/ authorized_by	concatenate iSamples SamplingEvent label and description; include information in 'authorized_by'	Description [descriptionType = 'Method']	.	The methodology employed for the study or research.
--- family name	no iSamples property	familyName	.	string
N.A.	not applicable for physical samples	Format	.	
produced_by/SamplingEvent/responsibility[Role=funder]	Could also be a funder for curation in Curation/responsibilities. If the information is available, the funder should be considered to be the particular award, grant, contract, etc. that supported the sample collection or curation activity.	fundingReferences	.	
produced_by/SamplingEvent/responsibility[Role=funder]/name	Name person or organization that supported the activity;		funderName	
produced_by/SamplingEvent/responsibility[Role=funder]/identifier	identifier for funding agent		funderIdentifier	
--- funder award number	award identifiers should go in isam:project		awardNumber	

iSamplesSchemaCore1.0	iSamples notes	DataCite 4.3	Sub element	Definition
--- funder award URI	identifier for the particular funding instrument that supported the activity. award identifiers should go in isam:project		awardURI	
produced_by/SamplingEvent/ project	a name or a resolvable URI for a project that is the sponsor for a samplingEvent. If funding was through an identified grant/award/ contract instrument, this is the label associated with that instrument, If a grant/contract/award was the instrument, this would be the award title.	fundingReferences/ awardTitle, awardNumber, awardURI	awardTitle	concatenate award information to identify a funded project
produced_by/SamplingEvent/ sampling_site	one WGS 84 Decimal Degree lat /long point location is required. Schema doesn't accommodate other coordinate locations. This corresponds to DataCite geoLocationPoint. Multiple places might be specified; iSamples is limited to a single point location -- doesn't make sense for a sample to come from more than one point location...	GeoLocation	.	Spatial region or named place where the data was gathered or about which the data is focused.
produced_by/SamplingEvent/ sampling_site/label	text to identify a location for people;		geoLocationPlac e	
produced_by/SamplingEvent/ sampling_site/description	description of location		geoLocationPlac e	
produced_by/SamplingEvent/ sampling_site/place_name	iSamples can have multiple values.		geoLocationPlac e	
produced_by/SamplingEvent/ sampling_site/sample_locatio n/longitude	sample location point coordinate		geoLocationPoin t/ pointLongitude	
produced_by/SamplingEvent/ sampling_site/sample_locatio n/latitude	sample location point coordinate		geoLocationPoin t/ pointLatitude	
--- given name	no iSamples property	givenName	.	string

iSamplesSchemaCore1.0	iSamples notes	DataCite 4.3	Sub element	Definition
sample_identifier	unique identifier for the physical object, ideally a URI that is physically attached to the material sample object, an IGSN or ARK	Identifier	.	The Identifier is a unique string that identifies a resource.
-- keyword language	no iSamples property	lang, ClassificationCode		DataCite has ClassificationCode and xml:lang attributes for subject
-- metadata language	not in isamples model; English is assumed	Language	.	
produced_by/SamplingEvent/ responsibility [role=collector]/name	string, name of a person. Format: {firstName MI lastName}.	name (creatorName in XML)	.	string
--- name type	no iSamples property; DataCite allows 'organizational' or 'personal' name		nameType	controlled
		nameIdentifiers	.	an array
produced_by/SamplingEvent/ responsibility [role=collector]/identifier	only one identifier associated with person in this role. iSamples identifier value is a string.	nameIdentifiers	nameIdentifier	string
-- name identifier scheme name	no iSamples property	nameIdentifiers	nameIdentifierScheme	string
-- name identifier scheme URI	no iSamples property	nameIdentifiers	schemeURI	URI
--- publication year	Time stamp for when the sample registration information is made public through the identifier authority; not currently included in iSamples metadata, use year of sample registration; If an embargo period has been in effect, use the date when the embargo period ends.	PublicationYear	.	The year when the data was or will be made publicly available. If an embargo period has been in effect, use the date when the embargo period ends.
registrant [role=registrant]/name/	identification of the agent that registered the sample, with contact information. Role = 'registrant'	Publisher	.	String. The name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the resource.
related_resource	link to another resource via identifier for the target resource, with a relationship type,	RelatedIdentifiers	.	

iSamplesSchemaCore1.0	iSamples notes	DataCite 4.3	Sub element	Definition
	label, and description. Label and description do not have matching properties in DataCite.			
related_resource/relationship	nature of relationship between sample and resource that is the target of the relationship. See vocabulary in development: https://docs.google.com/spreadsheets/d/1yk3ZEWzEc1cXBOwb-foCtq3isaBBa95Rv8eA6TIyieM		relationType	DataCite has a vocabulary focused on datasets, some likely values are IsDescribedBy, HasMetadata, IsPartOf (for link to parent sample where relation is a split or physical separate), IsDerivedFrom (link to parent sample where derivation is chemical process). IsDocumentedBy not recommended because of lack of clarity on distinction with 'IsDescribedBy'.
related_resource/target	identifier for the target resource in the relationship. Should be a resolvable URI.		relatedIdentifier	Identifiers of related resources. These must be globally unique identifiers.
--- related identifier type	no iSamples equivalent. Default to PURL, unless you can determine its one of the others in the DataCite vocabulary.		relatedIdentifierType	
--- related resource type general	no iSamples equivalent. If relation is to another sample, use PhysicalObject. Other likely values would be Dataset, DataPaper, Collection, Image, Text. Use Other if nothing matches.		resourceTypeGeneral	
has_material_sample_object_type	specify the kind of object that the specimen is, use iSamples Material Sample Object Type vocabulary, with ResourceTypeGeneral = PhysicalObject. the label for the Material_Sample_Object_Type is the text value for dataCite ResourceType	ResourceType	.	
-- constant		ResourceTypeGeneral	.	ResourceTypeGeneral = PhysicalObject, a constant for all physical samples.
dc_rights	A statement about various property rights associated with the resource, including intellectual property rights. Recommended practice is to refer to a rights statement with	rightsList/rights	rightsList/rights	

iSamplesSchemaCore1.0	iSamples notes	DataCite 4.3	Sub element	Definition
	a URI. If this is not possible or feasible, a literal value (name, label, or short text) may be provided.			
keywords	keywords.	Subjects	.	Subject, keyword, classification code, or key phrase describing the resource.
keywords/keyword	An identifier for the nomenclatural (not taxonomic) details of a scientific name.		subject	
keywords/keyword_uri			valueURI	
keywords/scheme_name			subjectScheme	
keywords/scheme_uri	not included for controlled vocabulary terms		schemeURI	
has_context_category/label	For context using controlled vocabulary, the DataCite subject is the identifier/label, for keywords, it is the keyword		subject	
has_context_category/identifier	use URI from isamples sampledFeature vocabulary		valueURI	single object value. Examples add keys for types from othervocabs, e.g. schemaorg, citeproc,bibtex
has_context_category/scheme_name	https://w3id.org/isample/vocabulary/sampledfeature/1.0/sampledfeaturevocabulary		subjectScheme	
has_context_category	Top level context, based on the kind of feature sampled. Specific identification of the sampled feature of interest is done through the SamplingEvent/Feature of Interest property. At least one value a skos:Concept from the iSamples sampledFeature vocabulary.	Subjects [subjectScheme=isamplesContext]	.	
has_material_category	iSamples top level classifications for material	Subjects[subjectScheme=isamplesMaterial]	.	
has_material_category/label	For material using controlled vocabulary, the DataCite subject is the identifier/label, for keywords, it is the keyword		subject	

iSamplesSchemaCore1.0	iSamples notes	DataCite 4.3	Sub element	Definition
has_material_category/identifier	use URI from iSamples material type vocabular		valueURI	
has_material_category/scheme_name	use https://w3id.org/isample/vocabulary/material/1.0/materialsvocabulary		subjectScheme	
has_material_sample_object_type/label	repeat label and uri in subjects for DataCite	Subjects[subjectScheme=isamplesSpecimenType]	subject	
has_material_sample_object_type/identifier			valueURI	
has_material_sample_object_type/scheme_name	https://w3id.org/isample/vocabulary/materialsampleobjecttype/1.0/conceptscheme	Subjects[subjectScheme=isamplesSpecimenType]	subjectScheme	
label	a human intelligible string used to identify the material sample; i.e. the name to use for the sample; should be unique in the scope of a sample collection. This will typically be a sample identifier or label assigned by the original collector	Title	.	A name or title by which a resource is known. May be the title of a dataset or the name of a piece of software.
N.A.	Samples don't have versions. metadata version should be specified by the schema property or the update date	Version	.	

Appendix 5

Mapping from iSamples to Biodiversity Information Standards (TDWG) Minimum Information about a Digital Specimen (MIDS), based on MIDS Github issue tracker.[59] Levels in MIDS scheme are indicated. Some of the fields are proposed in MIDS but as of this writing are 'not accepted'.

iSamplesSchemaCore1.0	iSamples notes	**MIDS**	MIDS note
-- metadata update date	Update date not currently included	Modified (level 1)	date/time of first creation or subsequent modification

[59] https://github.com/tdwg/mids/labels/status%3A%20accepted%20in%20specification

iSamplesSchemaCore1.0	iSamples notes	**MIDS**	MIDS note
-- sample mass	MIDS includes Mass in MIDS-3. iSamples, concatenate in description	Mass (level 3)	quantity of matter in a specimen, particularly for minerals, phases and meteorites
alternate_identifiers/label		collectingNumber (level 2)	identifier given to the specimen at the time it was recorded [collected]
curation/curation_location	Information about where and how the sample is currently stored.	InstitutionID (level 3)	identifier for the institution having custody of the object(s)
curation/description	Concatenate information from MIDS	PreparationType, PreservationMethod (not accepted)	proposed, not accepted; this information in iSamples curation description
curation/responsibility [role='classification']	unique identifier for the person, people, groups, or organizations responsible for assigning the scientific name to the subject. Include this as a curation/responsibility	IdentifiedByID (level 3)	list (concatenated and separated) of the globally unique identifier for the person, people, groups, or organizations responsible for assigning the scientific name to the subject
curation/responsibility/name	Person or organization name	Organization (level 0)	term to indicate in which institution the specimen is held. This may include an institution code and an institution identifier.
dc_rights	Statement of legal requirements and rights for accessing, using, or sharing information about the material sample.	License (level 1)	License under which the specimen data are published
has_material_category	Map to iSamples top level classifications for material. Include verbatim values if different as keywords.	MaterialType (not accepted)	
has_sample_object_type	specify the kind of object that the specimen is. Map MIDS terms to iSamples Material Sample Object Type vocab	ObjectType (level 1)	term to describe the kind of specimen. In combination with SpecimenType - hierarchical; a more specific classification than described by SpecimenType
keywords/keyword	An identifier for the nomenclatural (not taxonomic) details of a scientific name.	SpecimenType (level 1)	High-level term to delimit and define specimens. For example: preserved specimen, fossil specimen, as opposed to observation. [if there is a controlled vocabulary, map to iSamplesMaterialType category where logically consistent]
-- keywords/scheme_name = GeologicAge	included in MIDS-2. implement as keyword in iSamples	GeologicAge (level 2)	

iSamplesSchemaCore1.0	iSamples notes	**MIDS**	MIDS note
-- keywords/scheme_name = TypeStatus	included in MIDS-2. implement as keyword in iSamples	TypeStatus (level 2)	nomenclatural type status of the specimen; a null value means "Assumed not to be a type". Examples: Holotype, Isotype, Syntype, Cotype, Epitype, Neotype, Lectotype
keywords/keyword [scheme_uri = ICS time scale]	Formal time ordinal era terms and identifier in keywords. Summary of details about an age estimation for temporal extent of sample origin goes is iSamples sample description.	GeologicAge (level 2)	geological age of a Earth Science specimen (i.e. Fossil, Rock, Mineral or Meteorite) and can be any kind of stratigraphic age, isotopically determined age or structural age [numeric age should be reported in the description; this field should be consistent with a term for a time-ordinal era]
keywords/keyword_uri		ScientificNameID (level 3)	identifier for the nomenclatural (not taxonomic) details of a scientific name. [note that the associated keyword should be the nomenclatural detail label associated with the identifier.]
label	a human intelligible string used to identify the sample; i.e. the name to use for the sample; should be unique in the scope of a sample collection. This will typically be a sample identifier or label assigned by the original collector	Name (level 1)	string of characters and/or numbers by which the object is referenced within a collection
produced_by/SamplingEvent/ responsibility	the responsibility is an agent-- could be person or organization, with role = 'collector'. Other agents associated with the sampling event could be included, e.g. with roles like 'sponsor', 'funder'....	CollectingAgent (level 2)	list (concatenated and separated) of names of people, groups, or organizations responsible for recording the original Occurrence [i.e. sample collection]
produced_by/SamplingEvent/ responsibility[role=collector]/ identifier	only one identifier associated with person in this role. iSamples identifier value is a string.	CollectorID (level 3)	list (concatenated and separated) of the globally unique identifier for the person, people, groups, or expeditions responsible for responsible for collecting the specimen
produced_by/SamplingEvent/ result_time	Date on which the sample was collected.	dateCollected (level 2)	date/time when the [sample collection] event was recorded
produced_by/SamplingEvent/ sampling_site/identifier/identifier	An identifier for the geographical locality where the material sample was collected.	GeographicalLocalityID (level 3)	identifier for the geographical locality where the specimen was collected
produced_by/SamplingEvent/ sampling_site/ sample_location/latitude	sample location point coordinate	quantitativeLocation (level 2)	A quantitative measure that would include coordinate or shape data, an identifier, or data that can be easily converted into a quantitative measure

iSamplesSchemaCore1.0	iSamples notes	**MIDS**	MIDS note
produced_by/SamplingEvent/sampling_site/ sample_location/longitude	sample location point coordinate	quantitativeLocation (level 2)	quantitative measure that would include coordinate or shape data, an identifier, or data that can be easily converted into a quantitative measure
produced_by/SamplingEvent/sampling_site/place_name	iSamples can have multiple values.	qualitativeLocation (level 2)	term [or text] to describe where the specimen was collected; A human readable location
related_resource/label	a human intelligible string used to identify a thing, i.e. the name to use for the thing; should be unique in the scope of a sample collection or dataset.	Media (level2)	list (concatenated and separated) of media associated with the specimen. [not clear if this are expected to be labels or identifiers]
related_resource/target	identifier for the target resource in the relationship. Should be a resolvable URI.	AssociatedMediaID (level 3)	list (concatenated and separated) of identifiers (publication, global unique identifier, URI) of media associated with the specimen.
sample_identifier	unique identifier for the physical object, ideally a URI that is physically attached to the material sample object, an IGSN or ARK	PhysicalSpecimenID (Level 0)	unique identity [identifier] for the specimen within the curating institution

Appendix 6

Mapping from iSamples to Distributed System of Scientific Collections (DiSSCo) Open Digital Specimen (OpenDS) schema, based on Digital specimens JSON schema 0.1.0.[60] Note that OpenDS term 'specimen' is interpreted to be equivalent to iSamples 'material sample'.

iSamplesSchemaCore1.0	iSamples notes	**OpenDS**	**OpenDS notes**
-- digital object create date	No mapping	ods:created	The timestamp that the object version was created in DiSSCo
-- digital object identifier	identifier for the metadata record	ods:id	identifier for digital object
-- digital object type	correspond to JSON-LD @type; something like 'metadata record'. This should be a determinant for the schema of the digitalObject, but midsLevel property is more granular schema determinant.	ods:type	

[60] https://github.com/DiSSCo/openDS/blob/master/data-model/digitalobjects/0.1.0/digital-specimens/schema/digital-specimen.json

iSamplesSchemaCore1.0	iSamples notes	OpenDS	OpenDS notes
-- digital object version	No mapping	ods:version	The version of the object, each change generates a new version
-- has Media	could be determined by looking at the related_resource content	ods:hasMedia	boolean flag, Indicates if there are any media objects attached to this specimen
-- keywords/scheme_name = TypeStatus	included in MIDS-2. implement as keyword in iSamples	ods:markedAsType	The specimen is marked as a type specimen
-- living or preserved	map to keyword, schemeName='livingOrPreserved'	ods:livingOrPreserved	Whether the specimen is living or preserved, enum "living", "preserved". Seems to overlap with Basis of record, for which relevant terms for material samples are 'living', 'preserved' and 'fossil'.
-- material entity	Need to clarify OpenDS semantics to determine mapping	ods:materialEntity	links to material entity object that seems to duplicate the properties in the host record; perhaps it is to build child sample descriptions into a record for a compound sample? [use iSample related resource]
-- metadata license	all iSamples metadata is currently CC0. Have to review if we will harvest licensed metadata.	dcterms:license	License for the metadata of the physical specimen
-- metadata source	Source of metadata record harvest; could map to registrant.	ods:sourceSystem	The handle to the source system object which retrieved the data from the CMS
-- metadata update date	Update date not currently included in iSamples	dcterms:modified	Modification date for specimen information
-- topic origin	could map from isamples sampled feature/material type/material sample type to these terms, but likely not vice-versa...	ods:topicOrigin	"The topic origin of the specimen", "enum": ["Natural", "Human-made", "Mixed origin", "Unclassified"].
$schema	identifier for the schema for this metadata ('digital sample' document). iSamples sample descriptions are implemented with a JSON serialization, so this should be an identifier for a JSON schema	ods:midsLevel	The MIDS level of the object, see https://www.tdwg.org/community/cd/mids/. [interpret this to denote the schema for the 'digital object']
alternate_identifiers	other identifiers can be provided with label, scheme_name (authority), and identifier string.	identifiers	no explanation, assume is for alternate identifiers for the sample; include identifier string, lose scheme_name
curation/access_constraints	cultural, legal or other policy issues that bear on access to view, borrow, or subsample a sample (Curation)	dcterms:accessRights	
curation/curation_location	Information about where and how the sample is currently stored.	dwc:disposition | dwc:institutionName	Examples: in collection, missing, on loan, used up, destroyed, deaccessioned. Could go in curation descriptions? Prefix on location place name?

iSamplesSchemaCore1.0	iSamples notes	OpenDS	OpenDS notes
curation/responsibility/ identifier	Identifier for person or organization. only one identifier in iSamples. An identifier for the institution having custody of the object(s)	dwc:institutionId	ROR or Wikidata identifier, identifier for the institution having custody of the object(s)
curation/responsibility/name	Person or organization name	dwc:institutionCode | dwc:institutionName	name (or acronym) in use by the institution having custody of the object(s)
curation/responsibility[role= 'owner']	if this is different from the institution that has custody	dwc:ownerInstitutionId	ROR or Wikidata identifier for the owning institution
curation/responsibility[role= 'rightsHolder']	if this is different from the institution that has custody	dcterms:rightsHolder	
curation/responsibility[role='classification']	unique identifier for the person, people, groups, or organizations responsible for assigning the scientific name to the subject. Include this as a curation/responsibility	dwc:identification, dwc:identifiedBy, dwc:identifiedById	openDS allows multiple identifications to be attached to the samples, with a bunch of metadata; simple iSample record with one identifier agent can be handled like this. if there is a full identification object, attach as a related resource. See https://github.com/DiSSCo/openDS/blob/master/data-model/digitalobjects/0.1.0/digital-specimens/schema/identifications.json
has_context_category	Top level context, based on the kind of feature sampled. Specific identification of the sampled feature of interest is done through the SamplingEvent/Feature of Interest property. At least one value a skos:Concept from the iSamples sampledFeature vocabulary.	ods:topicDomain	The topic domain of the specimen", "enum": ["Life", "Environment", "Earth System", "Extraterrestrial", "Cultural Artefacts", "Archive Material", "Unclassified"]. Map to appropriate sampled feature class.
has_material_sample_object_type/label	repeat label and uri. Use labels from iSamples Material_Sample_Object_Type vocabulary.	dwc:basisOfRecord | dwc: preparations	relevant 'basis of record' terms are 'living', 'preserved' or 'fossil'. preparations is a list of "preparations and preservation methods" for the sample. For isamples preservation methods would be in curation. In the dwc documentation, most of the examples correspond to isamples material sample types; would need a vocabulary mapping to map dwc strings to iSamples concepts.
keywords/keyword	An identifier for the nomenclatural (not taxonomic) details of a scientific name.	ods:topicDiscipline	The topic discipline of the specimen","enum": ["Anthropology", "Botany", "Astrogeology","Geology", "Microbiology", "Palaeontology", "Zoology", "Ecology", "Other Biodiversity", "Other Geodiversity", "Unclassified"],

iSamplesSchemaCore1.0	iSamples notes	OpenDS	OpenDS notes
label	a human intelligible string used to identify the sample; i.e. the name to use for the sample; should be unique in the scope of a sample collection. This will typically be a sample identifier or label assigned by the original collector	ods:specimenName	The accepted specimen name of the digital specimen
produced_by/SamplingEvent/sampling_site/sample_location/obfuscated	Flag to indicate that accuracy of provided coordinate location has been reduced (obfuscated) because the exact location is restricted information. Default is 'false'	dwc:informationWithheld | dwc:dataGeneralizations	location information not given for endangered species
registrant [role=registrant]/name/	identification of the agent that registered the sample, with contact information. Role = 'registrant'	dwc:recordedBy | ???:recordedByAgent	name or possibly a name/role/identifier
registrant/identifier		dwc:recordedById	
related_resource[relationship= 'hasAgeDetermination']	link to sample 'age' determination detail object	chronometric age	OpenDS has object with details about an age estimation for temporal extent of sample origin, link to that
related_resource[relationship= 'hasAssertionObject']	link to related assertion/observation result	ods:assertion	link to openDS assertion object that ?reports some observation result about the sample. See https://github.com/DiSSCo/openDS/blob/master/data-model/digitalobjects/0.1.0/shared-models/schema/assertions.json
related_resource[relationship= 'isCitedBy' or 'isDocumentedBy']	link to publications	ods:citations	target is a draft object in OpenDS json. Assume the intention is link to publications that reference the sample
related_resource[relationship= 'isRelatedTo']	generic resource link	ods:entityRelationships	openDS entityRelationship object mapping is pretty clear
related_resource[relationship='isMemberOf']/label		dwc:collectionCode	name, acronym, coden, or initialism identifying the collection that includes the sample
related_resource[relationship='isMemberOf']/target	identifier for the target resource in the relationship. Should be a resolvable URI.	dwc:collectionId	

iSamplesSchemaCore1.0	iSamples notes	OpenDS	OpenDS notes
sample_identifier	unique identifier for the physical object, ideally a URI that is physically attached to the material sample object, an IGSN or ARK	ods:normalisedPhysicalSpecimenId | ods:physicalSpecimenId | ods:physicalSpecimenIdType	physical specimen id might be a local identifier; normalized id has namespace (source-system-id) qualifier to make globally unique; The IdType indicate if the physical identifier is globally unique or locally unique, values are enum ["Resolvable", "Global", "Local"], applies to physicalSpecimenID; since iSamples sample_identifier is expected to be globally unique, provide the normalized id, thus the IdType is not relevant.