ISA-TAB-Nano Guidelines

Overview

The purpose of the ISA-TAB-Nano Guidelines are to provide users with guidance and best practices on the creation of ISA-TAB-Nano files. This includes general guidelines associated with creating all ISA-TAB-Nano files as well as guidelines for creating each ISA-TAB-Nano file: Investigation File, Material File, Assay File, and Study File. In general, ISA-TAB-Nano follows recommendations from the ISA-TAB community for guidance and best practices associated with the capture of information not specific to ISA-TAB-Nano extensions.

The sections below provide guidelines on ISA-TAB-Nano that were developed based on feedback from ISA-TAB-Nano user community. The ISA-TAB-Nano guidelines are intended to promote interoperability across organizations in support of material comparison. Regardless of whether an organization chooses to adopt ISA-TAB-Nano guidelines, it is recommended that organizations use a consistent approach for creating ISA-TAB-Nano files within their organization.

General Guidelines

Providing Reasons for Missing Values

Many ontologies and standardized vocabularies contain terminology that can be used to indicate why certain values are not being reported in a study or assay (such as unknown, not applicable, not available, not provided, etc.). Terminologies that have coded values for missing data include:

- Logical Observation Identifier Names and Codes (LOINC), which can be found on the Bioontology LOINC Site, NCI Term Browser LOINC PAGE, or at LOINC
- Systematized Nomenclature of Medicine - Clinical Terms (SNOMEDCT), which can be found on the Bioontology SNOMEDCT Site, NCI Term Browser SNOMEDCT Page
- The International Health Terminology Standards Development Organisation
- The Health Level Seven Reference Implementation Model, Version 3 (HL7), which can be found on the Bioontology HL7 Site, NCI Term Browser HL7 Page, or HL7 site

Additionally, some ontologies present these values in discrete sets (such as value sets, code lists or hierarchical branches); Radiology Lexicon (RadLex) has a hierarchical branch under the term certainty descriptor (Code RID29, see the Relationships tab at RadLex), NCI Thesaurus (NCIt) has a hierarchical branch under the term Missing Value Reason (Code C48655, see the Relationships tab at NCI Thesaurus), and the Clinical Data Interchange Standards Consortium Terminology Study Data Tabulation Model (CDISC SDTM) terminology has a code list in a value CDISC SDTM Yes No Unknown or Not Applicable Response Terminology (see NCI Thesaurus).

Referencing Ontology Terms

ISA-TAB-Nano recommends using the JavaScript Object Notation (JSON) convention drafted by the ISA-TAB team for referencing Ontology Term Sources and Ontology Term Accession Numbers, which utilizes Uniform Resource Identifiers (URI) to unambiguously reference both the sources and individual term accession numbers. This draft schema is on the ISA-tools GitHub page (https://github.com/ISA-tools/isa-api/blob/master/isatools/schemas/isa_model_version_2_0_schemas/core/resource_annotation_schema.json). For referencing an Ontology Term Source, the full source name (Term Source Description), an abbreviated name (Term Source Name) and source homepage URI (Term Source File) should be included in the Ontology Source Reference section of the Investigation file. Then to reference an Ontology Term Accession Number, the URI for the term, which contains the abbreviated name of the source and the term accession number, should be present in the Term Accession Number fields used in the Study, Assay and Material files.

Investigation File Guidelines

Organizing Investigations and Studies

ISA-TAB provides the flexibility to allow researchers to organize their investigations and studies as appropriate. ISA-TAB-Nano employs the same principles in support of nanotechnology investigations and studies. It is recommended that as a best practice, the number of publications associated with an investigation should be minimized insofar as this does not result in multiple Material files being created for identical material samples. For example, if multiple publications refer to exactly the same nanomaterials, they should be part of the same Investigation; otherwise, an Investigation should refer to a single Investigation to facilitate tracking of provenance.

Material File Guidelines

Support for Nominal Particle Characteristics vs. Experimental

The Characteristics field in the Material File captures "nominal" particle characteristics which refer to the initial composition of the particle. "Nominal" particle characteristics are those that are inherent to the particle prior to performing an experiment or assay to characterize the particle. For example, an organization may receive a *single walled* carbon nanotube with a specific *diameter* from a vendor. The organization would record the carbon nanotube *wall type* and *diameter* as characteristics of the particle in the Material File. The organization may then perform characterization assays to determine the change in the particle *diameter* under certain experimental conditions (t*emperature, Ph, media solvent*, etc.). The organization would record the particle *dia meter* measured after performing the assay as Measurement Values in the Assay File along with the conditions which may be recorded as factors or parameters depending on the intent of the assay.

There are some situations in which an organization may decide to record particle composition information (such as the “Material Chemical Name” for a given component) that is determined experimentally. ISA-TAB-Nano does not prevent organizations from recording this information in the Material File; however, if this information is recorded in the Material File, the assay information associated with the experiment will not be captured in the Material File. If an organization decides to record particle composition information that is determined experimentally in the Material File, ISA-TAB-Nano recommends that an Assay File also be created to record the experimental conditions in which the assay was performed to determine the composition information. Although this may result in some duplicate information, this approach would increase experimental reproducibility.

One example of this situation involves the capture of particle purity and impurities. For example, a particle with a certain % of iron oxide on the surface may be considered an impurity. Typically, particle impurities are included as characteristics associated with particle composition. Particle impurities can be included as characteristics in the Material File if using the same measurement technique. If using a different measurement technique for each impurity, it is recommended that this information be recorded in the Assay File as Measured Values to support the capture of the diverse experimental conditions. Additionally, the chemical associated with the surface (such as iron oxide), could be recorded as a component of the particle in a separate row in the Material File.

Linking Multiple External Files

ISA-TAB-Nano provides support for linking multiple external files (data, images, etc.) to a Material File. There are several ways in which multiple files can be referenced which include:

- Create a tar.gz file containing the multiple files and referencing the tar.gz file in the Material Data File column. If a tar.gz file is used, it is recommended that the tar.gz file contain a manifest file that describes the contents of the tar.gz file.
- Reference each file separated by a semi-colon in the Material Data File column (such as "file 1; file 2"). If using this approach, all columns associated with the file (such as file description) need to be modified to support multiplicity
- Replicate the Material File row that requires references to multiple files and reference one file per row

In discussions with ISA-TAB, the preferred way is to have multiple files appear on different rows; however, the file can grow quite large. ISA-TAB would accept having an archive file (zip, tar.gz) and semi-colon separated files (file 1; file 2); however the ISA-TAB software currently does not support this.

Support for Comments

ISA-TAB-Nano 1.3 has been modified to allow comments anywhere within the Material File. If comments occur in a Study or Assay files the comments are attached to the node or the closest field, for example after the Sample name and Protocol REF. Comments should always have brackets [].

Support for Sample Characteristics

Characteristics […] can be applied to the *Source Name* and *Sample Name* in the Study File. The Study File template has been modified to include this feature.

Assay File Guidelines

Measurement Value as an Attribute off of the Assay Node

The Measurement Value column was added as an extension to ISA-TAB to capture the assay endpoints for nanomaterial assays that do not require separate files. For example, in a size by DLS assay, the assay endpoint may be average size (nm) with separate columns for factors (such as temperature) and parameters (such as media solvent). Since the measurement value is associated with the assay being performed, the measurement value is an attribute off of the assay name node.

Standard Approach for Recording Data Specified in Ranges (Plan to discuss with ISA-TAB and the Nano WG, put off until future release)

- Standardized the approach for recording data specified in ranges (Challenge #3)
- •Discussion Points:
  –Discussed how best to record data specified in ranges (such as sonication strength between 30-40), or other operations (such as LOEL < 60) in Factor Values, Parameter Values, and Measurement Values
  •Discussed two options presented by Richard Marchese-Robinson:
  –1) Use the statistic term such as minimum or less than (such as minimum(Sonication Strength), Less Than(LOEL))
  •Egon Willighagen suggested using min_value and max_value
  •Liz Hahn-Dantona noted that there are ontologies for statistical terms
  •This approach may be easier to standardize
  –2) Standardize use of “30 – 40”, “< 60” in existing columns
  •Noted that 30 – 40 may be interpreted differently
  –Nina Jeliazkova questioned why ISA-TAB-Nano added Measurement Values which is not compliant with ISA-TAB
  •Characterization data is often a spreadsheet containing the sample ID, parameters, factors, and measurements (such as sample ID, media solvent, temperature, mean size). Adding a column for measurement values in ISA-TAB-Nano instead of referencing a separate file avoids creating multiple files with duplicate information.
  •Recommended ISA-TAB-Nano Improvement:
  –Agreed that a standard way for representing data in ranges or other operations needs to be determined and documented in ISA-TAB-Nano. The approach needs further discussion.
- Recommend use of STATO: the statistical methods ontology (https://www.ebi.ac.uk/ols/ontologies/stato or https://bioportal.bioontology.org/ontologies/STATO) as it appears to be the most complete ontology for statistics.
- Note: STATO supports Minimum_Value and Maximum_Value but does not contain Less Than, Greater Than, etc.

Definitions for Parameter Value and Factor Value

Parameters Values are intrinsic experimental conditions that may be varied or constant. *Factor Values* are variable experimental conditions manipulated by the investigator with the intention to affect the subject of study. The definitions for these fields have been modified to clarify the differences.