ISA Model and Serialization Specifications

Release 1.0

ISA Community

Apr 29, 2021

Contents

Status: ISA Model and Serialization Specifications 1.0 (28 October 2016)

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described by RFC 2119.

The ISA Model and Serialization Specifications are licensed under CC BY-SA 4.0.

The ISA Model and Serialization Specifications are maintained by Susanna-Assunta Sansone[1], Philippe Rocca-Serra[1], Alejandra Gonzalez-Beltran[1] and David Johnson[1] on behalf of the ISA Community.

If you wish to make comments regarding these specifications, please see the page on *how to contribute*.

---

[1] Oxford e-Research Centre, University of Oxford, UK.

Contents   **1**

Introduction

ISA is a metadata framework to manage an increasingly diverse set of life science, environmental and biomedical experiments that employ one or a combination of technologies. Built around the **Investigation** (the project context), **Study** (a unit of research) and **Assay** (analytical measurements) concepts, ISA helps you to provide rich descriptions of experimental metadata (i.e. sample characteristics, technology and measurement types, sample-to-data relationships) so that the resulting data and discoveries are reproducible and reusable.

Note: For an introduction to ISA, please read the paper, Towards interoperable bioscience data published in *Nature Genetics*. For more details on the ISA framework and supported tools, please see https://isa-tools.org

The ISA Model and Serialization Specifications define an Abstract Model of the metadata framework. The ISA Abstract Model has been implemented in two format specifications, ISA-Tab and ISA-JSON, both of which have supporting tools and services associated with them. The format specifications are also available for additional tooling to take advantage of ISA-formatted content.

These specifications are primarily aimed at software engineers to facilitate the development of automated export from databases, or import into analytical or other tools.

1.1 ISA Abstract Model

This ISA specification defines an Abstract Model of the metadata framework. The ISA Abstract Model has been implemented in two format specifications, *ISA-Tab* and *ISA-JSON*, both of which have supporting tools and services associated with them. The format specifications are also available for additional tooling to take advantage of ISA-formatted content.

The concept map below shows the ISA objects/entities and their relation to one another:
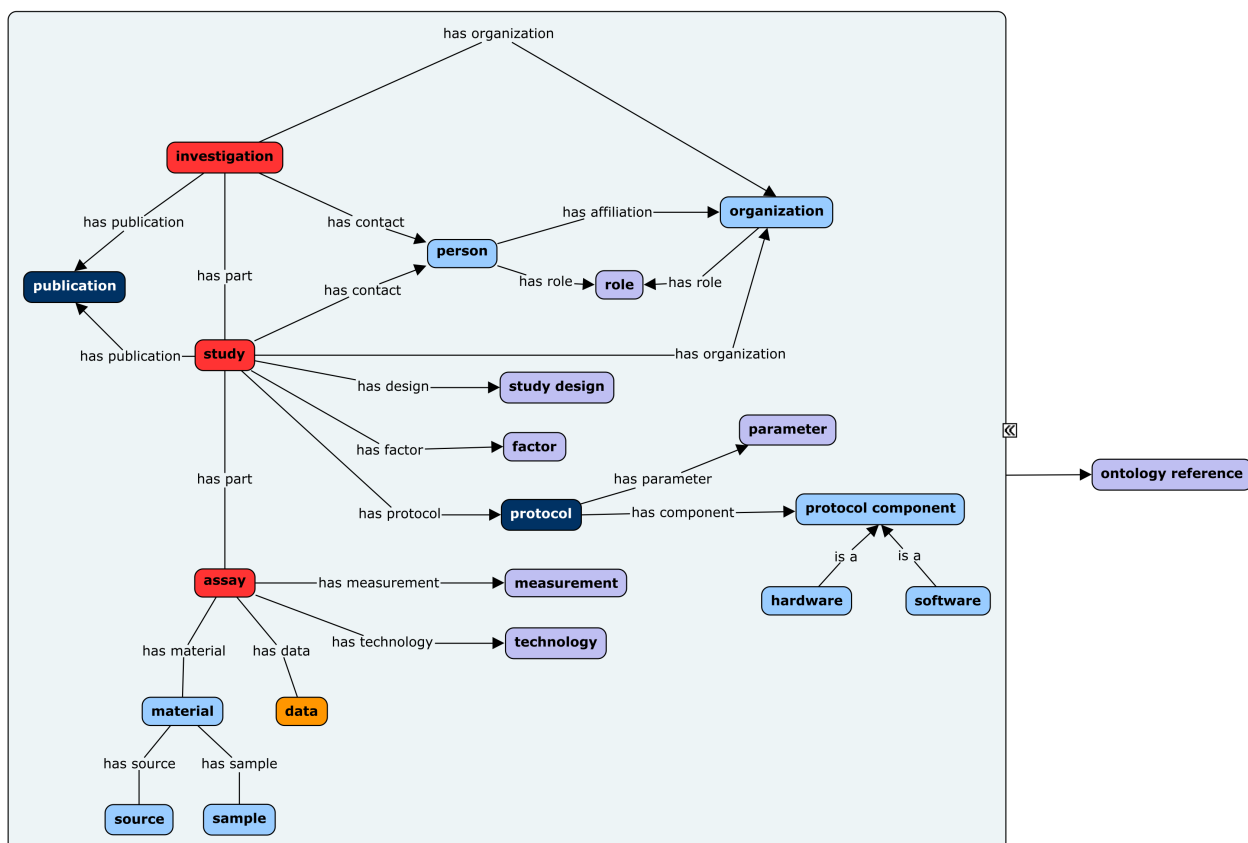
Note: The concept `ontology reference` depicted above refers to a combination of the `Ontology Annotation` and `Ontology Source` concepts as described below.

1.1.1 Investigation, Study, Assay

The ISA model consists of three core entities to capture experimental metadata:

- `Investigation`
- `Study`
- `Assay`

An `Investigation` contains all the information needed to understand the overall goals and means used in an experiment; experimental steps (or sequences of events) are described in a `Study` and `Assay` . For each `Investigation` there may be one or more `Study` associated with it; for each `Study` there may be one or more `Assay`.

Investigation

An `Investigation` is intended to:

1. to record metadata relating to a given investigation

2. to link related `Study` objects under an `Investigation` (this only becomes necessary when two or more `Study` objects need to be grouped)

An `Investigation` is used to record metadata relating to the description of the investigation context, such as the title and description of the investigation as well as about related people and scholarly publications. `Study` and `Assay` objects are grouped within an `Investigation` to record other metadata within the relevant contexts.

An `Investigation` SHOULD record the following:

Property	Datatype	Description
Identifier	String	A identifier or an accession number provided by a repository. This SHOULD be locally unique.
Title	String	A concise name given to the investigation.
Description	String	A textual description of the investigation.
Submission Date	Representation of a ISO8601 date	The date on which the investigation was reported to the repository.
Public Release Date	Representation of a ISO8601 date	The date on which the investigation was released publicly.
Publications	A list of Publication	A list of Publications relating to the investigation.
Contacts	A list of Contact	A list of Contacts relating to the investigation.

Study

A `Study` is a central concept containing information on the subject under study, its characteristics and any treatments applied.

A `Study` contains contextualising information for one or more `Assay`. Metadata about the study design, study factors used, and study protocols are recorded in `Study` objects, as well as information similarly to the `Investigation` including title and description of the study, and related people and scholarly publications.

A `Study` SHOULD record the following:

Property	Datatype	Description
Identifier	String	A identifier or an accession number provided by a repository. This SHOULD be locally unique.
Title	String	A concise name given to the investigation.
Description	String	A textual description of the investigation.
Submission Date	Representation of a ISO8601 date	The date on which the investigation was reported to the repository.
Public Release Date	Representation of a ISO8601 date	The date on which the investigation was released publicly.
Publications	A list of Publication	A list of Publications relating to the study.
Contacts	A list of Contact	A list of Contacts relating to the study.
Design Type	Ontology Annotation	A classifier of the study based on the overall experimental design, e.g cross-over design or parallel group design.
Factor Name	String	The name of one factor used in the Study and/or Assay files. A factor corresponds to an independent variable manipulated by the experimentalist with the intention to affect biological systems in a way that can be measured by an assay. The value of a factor is given in the Study or Assay file, accordingly.
Factor Type	Ontology Annotation	An classification of this factor into categories.

In a `Study` object we record the provenance of biological samples, from source material through a collection process to sample material, represented with directed acyclic graphs (direct graphs with no loops/cycles). The pattern of nodes is usually formed of a source material node, followed by a sample collection process node, followed by a sample material node.

For example:

```
(source material)->(sample collection)->(sample material)
```

These study graphs MAY split and pool depending on how the samples are collected.

In a splitting example, multiple samples might be derived from the same source:

```
(source material 1)->(sample collection)->(sample material 1)
(source material 1)->(sample collection)->(sample material 2)
```

In a pooling example, multiple sources may be used to create a single sample:

```
(source material 1)->(sample collection)->(sample material 1)
(source material 2)->(sample collection)->(sample material 1)
```

Assay

An `Assay` represents a test performed either on material taken from a subject or on a whole initial subject, producing qualitative or quantitative measurements.

An `Assay` groups descriptions of provenance of sample processing for related tests. Each test typically follows the steps of one particular experimental workflow described by a particular protocol.

`Assay`-related metadata includes descriptions of the measurement type and technology used, and a link to what study protocol is applied. Where an assay produces data files, links to the data are recorded here.

An `Assay` SHOULD record the following:

Property	Datatype	Description
Measurement Type	Ontology Annotation	An Ontology Annotation to qualify the endpoint, or what is being measured (e.g. gene expression profiling or protein identification).
Technology Type	Ontology Annotation	An Ontology Annotation to identify the technology used to perform the measurement, e.g. DNA microarray, mass spectrometry.
Technology Platform	String	The manufacturer and platform name, e.g. Bruker AVANCE, of the technology used.

In an `Assay` we record the provenance of biological samples, from sample material through an experimental workflow, represented with directed acyclic graphs. `Assay` graphs usually follow the pattern of a sample material, followed by a series of process and material/data nodes.

For example, to show a sample that goes through some extraction process (e.g. nucleic acid extraction) through to producing some sequenced data, we might produce something like:

```
(sample material)->(extraction process)->(extract)->(sequencing process)->(raw data␣
↪file)
```

Like with the study graphs, splitting and pooling can occur where appropriate in assay graphs.

Study and Assay graphs

Experimental graphs relating to `Study` and `Assay` objects are made up of specific types of nodes.

Experimental graphs MUST be directed and acyclic (i.e. MUST NOT contain loops/cycles).

All nodes in `Study` and `Assay` graphs MUST be uniquely identifiable. User-defined identifiers MAY also be used.

Experimental graphs MUST be composed of the following node types

Material nodes

`Material` nodes can also be used as a generic structure to describe materials consumed or produced during an experimental workflow. `Material` nodes SHOULD record the following:

Property	Datatype	Description
Characteristics	A list of Characteristic	A list of material characteristics that may be qualitative or quantitative in description. Qualitative values MAY be Ontology Annotations, while quantitative values MAY be qualified with a Unit definition.
Material Type	Ontology Annotation	An Ontology Annotation describing the material.

`Source` nodes are a special kind of `Material` node and are considered as the starting biological material used in a study. `Source` nodes SHOULD be followed by a `Process` node describing a sample collection process, and SHOULD only appear in `Study` graphs.

`Sample` nodes are a special kind of `Material` node and represent major outputs resulting from a protocol application. `Sample` nodes in the `Study` graphs SHOULD be preceded by a `Process` node describing a sample collection process. `Sample` nodes in the `Assay` graphs SHOULD be followed by a `Process` node and SHOULD NOT be preceded by any node.

Data nodes

`Data` nodes represent outputs resulting from a protocol application that corresponds to some process that produces data, typically in the form of data files. `Data` nodes SHOULD record the following:

Property	Datatype	Description
File name	String	A file name or full path referencing a data file produced by the related process that MAY be packaged with, or is accessible via, the ISA reference implementation content.

`Data` nodes SHOULD be preceded by a `Process` node describing a data-producing process, such as NMR scanning or DNA sequencing.

Process nodes

`Process` nodes represent the application of a protocol to some input material (e.g. a `Source`) to produce some output (e.g.a `Sample`).

`Process` nodes SHOULD record the following:

Property	Datatype	Description
Parameter Values	A list of Parameter Value	Reporting on the values taken by parameters when applying a protocol. A protocol description in the Study SHOULD declare the required parameters, where here the values applied are recorded.
Performer	String	Name of the operator who carried out the protocol. This allows account to be taken of operator effects and can be part of a quality control data tracking.
Date	Representation of an ISO8601 date	The date on which a protocol is performed. This allows account to be taken of day effects and can be part of a quality control data tracking.

`Process` nodes SHOULD be preceded by zero or more `Material` or `Data` nodes, and followed by zero or more `Material` or `Data` nodes.

1.1.2 Ontology Annotation

For a given value, an `Ontology Annotation` SHOULD qualify this value with an accession number taken from an `Ontology Source`.

An `Ontology Annotation` SHOULD record the following:

Property	Datatype	Description
Accession Number	String or URI	The accession number or reference from the Ontology Source associated with the selected term.

1.1.3 Ontology Source

An `Ontology Source` describes the resource from which the value of an `Ontology Annotation` is derived from. An `Ontology Source` SHOULD be referenced by an `Ontology Annotation`. An `Ontology`

`Source` should contain enough information on which to be able to ascertain the provenance of an `Ontology Source`.

An `Ontology Source` SHOULD record the following:

Property	Datatype	Description
Name	String	The name of the source of a term; i.e. the source controlled vocabulary or ontology. These names will be used to reference the Ontology Source from an Ontology Annotation.
File	String	A file name or a URI of an official resource.
Version	String	The version number of the Term Source to support terms tracking.

1.1.4 Unit

A `Unit` is used to classify dimensional data, and used accordingly with relevant values.

A `Unit` SHOULD be implemented as an `Ontology Annotation`.

1.1.5 Publication

A `Publication` SHOULD record the following:

Property	Datatype	Description
PubMed ID	Representation of a PubMed ID	The PubMed IDs of the described publication(s) associated with this investigation.
DOI	Representation of a DOI	A Digital Object Identifier (DOI) for that publication (where available).
Author List	A list of Strings	The list of authors associated with that publication.
Title	String	The title of publication associated with the investigation.
Status	Ontology Annotation	An Ontology Annotation describing the status of that publication (i.e. submitted, in preparation, published).

1.1.6 Contact

A `Contact` SHOULD record the following:

Property	Datatype	Description
Name	String	The name of a person.
Email	Representation of an email	The email address of a person.
Phone	Representation of a phone number	The telephone number of a person.
Address	Multi-line string	The address of a person.
Affiliation	String	The organization affiliation for a person.
Roles	A list of Ontology Annotations	Ontology Annotations to classify the roles performed by this person in the context of an Investigation or Study.

1.2 ISA-Tab format

Important: As a pre-requisite to reading this specification, please make sure you have read and understood the *ISA Abstract Model* that the ISA-Tab format is based on.

For detail on ISA framework terminology, please read the *ISA Abstract Model specification*.

This document describes the ISA Abstract Model reference implementation specified in the ISA-Tab format. ISA-Tab files are tab separated value (tsv) files, with specific labeled column structures specified below.

Below we provide the schemas and the content rules for valid ISA-Tab documents. Full examples of ISA content as ISA-Tab can be found in the ISA datasets repository, here https://git.io/vD1vC We recommend that you study these examples to better understand the structure of ISA-Tab documents.

1.2.1 Format

ISA-Tab uses three types of file to capture the experimental metadata:

- Investigation file
- Study file
- Assay file (with associated data files)

The Investigation file contains all the information needed to understand the overall goals and means used in an experiment; experimental steps (or sequences of events) are described in the Study and in the Assay file(s). For each Investigation file there may be one or more Studies defined with a corresponding Study file; for each Study there may be one or more Assays defined with corresponding Assay files.

Files SHOULD be encoded using UTF-8.

Column delimiters SHOULD be the Unicode Horizontal Tab character (Unicode U+0009).

In order to facilitate identification of ISA-Tab component files, specific naming patterns SHOULD follow:

- `i_*.txt` for identifying the Investigation file, e.g. `i_investigation.txt`

- `s_*.txt` for identifying Study file(s), e.g. `s_gene_survey.txt`

- `a_*.txt` for identifying Assay file(s), e.g. `a_transcription.txt`

All labels are case-sensitive:

- In the Investigation file, section headers MUST be completely written in upper case (e.g. STUDY), field headers MUST have the first letter of each word in upper case (e.g. Study Identifier); with the exception of the referencing label (REF).

- In the Study and Assay files, column headers MUST also have the first letter of each word in upper case, with the exception of the referencing label (REF).

Dates SHOULD be supplied in the ISO8601 format.

All values of cells MAY be enveloped with the Unicode Quotation Mark, Unicode U+0022 (the `"` character).

For maximal portability file names should only contain only ASCII characters not excluded already (that is `A-Za-z0-9._!#$%&+,;=@^(){}'[]` - we exclude space as many utilities do not accept spaces in file paths): non-English alphabetic characters cannot be guaranteed to be supported in all locales. It would be good practice to avoid the shell metacharacters `(){}'[]$.".`

1.2.2 Investigation File

The Investigation file fulfils four needs:

1. to declare key entities, such as factors, protocols, which may be referenced in the other files

2. to track provenance of the terminologies (controlled vocabularies or ontologies) there are used, where applicable

3. to relate Assay files to Studies

4. to relate each Study file to an Investigation (this only becomes necessary when two or more Study files need to be grouped).

An Investigation file is structured as a table with vertical headings along the first column, and corresponding values in the subsequent columns. The following section headings MUST appear in the Investigation file (in order), and the study block (headings from `STUDY` to `STUDY CONTACTS`) can be repeated, one block per study associated with the investigation.

- `ONTOLOGY SOURCE REFERENCE`

- `INVESTIGATION`

- `INVESTIGATION PUBLICATIONS`

- `INVESTIGATION CONTACTS`

- `STUDY`

- `STUDY DESIGN DESCRIPTORS`

- `STUDY PUBLICATIONS`

- `STUDY FACTORS`

- `STUDY ASSAYS`

- `STUDY PROTOCOLS`

- `STUDY CONTACTS`

In the following sections, examples of each section block are given beside the specification of each section.

For a full example of a complete Investigation File, please see https://git.io/vD1va.

Attention: Rows in which the first character in the first column is Unicode U+0023 (the # character) MUST be interpreted as comments, where reference implementation parsers SHOULD ignore those lines entirely.

Rows where the label `Comment[<comment name>]` appear can also appear within any of the section blocks. Where these appear, the comment name must be unique within the context of a single block (e.g. you cannot have multiple occurences of `Comment[external DB REF]` within `STUDY ASSAYS`. Also, the value cells MUST match the number of values indicated by the rest of the section in context.

Ontology Source Reference section

The Ontology Source section of the Investigation file is used to declare Ontology Sources used elsewhere in the ISA-Tab files within the context of an Investigation.

Where a row labelled with `Term Source REF` suffixed in the Investigation file, the value of the cell SHOULD match one of the `Term Source Name` value declared in this section.

Where a column labelled with `Term Source REF` in a Study file or Assay file associated with the Investigation, the value of the cell SHOULD match one of the `Term Source Name` value declared in this section.

This section implements a list of `Ontology Source` from the ISA Abstract Model.

This section MUST contain zero or more values.

ONTOLOGY SOURCE REFERENCE

This section MUST contain the following labels, with the specified datatypes for values supported:

Label	Datatype	Description
Term Source Name	String	The name of the source of a term; i.e. the source controlled vocabulary or ontology. These names will be used in all corresponding Term Source REF fields that occur elsewhere.
Term Source File	String (file name or URI)	A file name or a URI of an official resource.
Term Source Version	String	The version number of the Term Source to support terms tracking.
Term Source Description	String	Use for disambiguating resources when homologous prefixes have been used.

For example, the `ONTOLOGY SOURCE REFERENCE` section of an ISA-Tab `i_*.txt` file may look as follows:

```
ONTOLOGY SOURCE REFERENCE
Term Source Name        "CHEBI"        "EFO"        "OBI"        "NCBITAXON"
↪"PATO"
Term Source File        "http://data.bioontology.org/ontologies/CHEBI"        "http://
↪data.bioontology.org/ontologies/EFO"        "http://data.bioontology.org/ontologies/
↪OBI"        "http://data.bioontology.org/ontologies/NCBITAXON"        "http://data.
↪bioontology.org/ontologies/PATO"
Term Source Version        "78"        "111"        "21"        "2"        "160"
Term Source Description        "Chemical Entities of Biological Interest Ontology
↪"        "Experimental Factor Ontology"        "Ontology for Biomedical␣
↪Investigations"        "National Center for Biotechnology Information (NCBI)␣
↪Organismal Classification"        "Phenotypic Quality Ontology"
```

Investigation section

This section is organized in several subsections, described in detail below. The Investigation section provides a flexible mechanism for grouping two or more Study files where required. When only one Study is created, the values in this section SHOULD be left empty and the relevant metadata values recorded in the Study section only.

These sections implement an `Investigation` from the ISA Abstract Model.

INVESTIGATION

This section MUST contain zero or one values.

This section MUST contain the following labels, with the specified datatypes for values supported:

Label	Datatype	Description
Investigation Identifier	String	A identifier or an accession number provided by a repository. This SHOULD be locally unique.
Investigation Title	String	A concise name given to the investigation.
Investigation Description	String	A textual description of the investigation.
Investigation Submission Date	String formatted as ISO8601 date YYYY-MM-DD	The date on which the investigation was reported to the repository.
Investigation Public Release Date	String formatted as ISO8601 date YYYY-MM-DD	The date on which the investigation was released publicly.

For example, the `INVESTIGATION` section of an ISA-Tab `i_*.txt` file may look as follows:

```
INVESTIGATION
Investigation Identifier        "BII-I-1"
Investigation Title       "Growth control of the eukaryote cell: a systems biology
↪study in yeast"
Investigation Description       "Background Cell growth underlies many key cellular
↪and developmental processes, yet a limited number of studies have been carried out
↪on cell-growth regulation. Comprehensive studies at the transcriptional, proteomic
↪and metabolic levels under defined controlled conditions are currently lacking.
↪Results Metabolic control analysis is being exploited in a systems biology study of
↪the eukaryotic cell. Using chemostat culture, we have measured the impact of
↪changes in flux (growth rate) on the transcriptome, proteome, endometabolome and
↪exometabolome of the yeast Saccharomyces cerevisiae. Each functional genomic level
↪shows clear growth-rate-associated trends and discriminates between carbon-
↪sufficient and carbon-limited conditions. Genes consistently and significantly
↪upregulated with increasing growth rate are frequently essential and encode
↪evolutionarily conserved proteins of known function that participate in many
↪protein-protein interactions. In contrast, more unknown, and fewer essential, genes
↪are downregulated with increasing growth rate; their protein products rarely
↪interact with one another. A large proportion of yeast genes under positive growth-
↪rate control share orthologs with other eukaryotes, including humans. Significantly,
↪ transcription of genes encoding components of the TOR complex (a major controller
↪of eukaryotic cell growth) is not subject to growth-rate regulation. Moreover,
↪integrative studies reveal the extent and importance of post-transcriptional
↪control, patterns of control of metabolic fluxes at the level of enzyme synthesis,
↪and the relevance of specific enzymatic reactions in the control of metabolic
↪fluxes during cell growth. Conclusion This work constitutes a first comprehensive
↪systems biology study on growth-rate control in the eukaryotic cell. The results
↪have direct implications for advanced studies on cell growth, in vivo regulation of
↪metabolic fluxes for comprehensive metabolic engineering, and for the design of
↪genome-scale systems biology models of the eukaryotic cell."
```

```
Investigation Submission Date          "2007-04-30"
Investigation Public Release Date         "2009-03-10"
```

INVESTIGATION PUBLICATIONS

This section MUST contain zero or more values.

This section MUST contain the following labels, with the specified datatypes for values supported:

Label	Datatype	Description
Investigation PubMed ID	String formatted as valid PubMed ID	The PubMed IDs of the described publication(s) associated with this investigation.
Investigation Publication DOI	String formatted as valid DOI	A Digital Object Identifier (DOI) for that publication (where available).
Investigation Publication Author List	String	The list of authors associated with that publication.
Investigation Publication Title	String	The title of publication associated with the investigation.
Investigation Publication Status	String, or Ontology Annotation by providing accompanying Term Accession Number and Term Source REF	A term describing the status of that publication (i.e. submitted, in preparation, published).
Investigation Publication Status Term Accession Number	String or URI	The accession number from the Term Source associated with the selected term.
Investigation Publication Status Term Source REF	String	Identifies the controlled vocabulary or ontology that this term comes from. The Source REF has to match one the Term Source Name declared in the in the Ontology Source Reference section.

For example, the INVESTIGATION PUBLICATIONS section of an ISA-Tab i_*.txt file may look as follows:

```
INVESTIGATION PUBLICATIONS
Investigation PubMed ID        "17439666"
Investigation Publication DOI        "doi:10.1186/jbiol54"
Investigation Publication Author List        "Castrillo JI, Zeef LA, Hoyle DC, Zhang⌴
↪N, Hayes A, Gardner DC, Cornell MJ, Petty J, Hakes L, Wardleworth L, Rash B, Brown⌴
↪M, Dunn WB, Broadhurst D, O'Donoghue K, Hester SS, Dunkley TP, Hart SR, Swainston N,
↪ Li P, Gaskell SJ, Paton NW, Lilley KS, Kell DB, Oliver SG."
Investigation Publication Title        "Growth control of the eukaryote cell: a⌴
↪systems biology study in yeast."
Investigation Publication Status        "indexed in Pubmed"
Investigation Publication Status Term Accession Number        ""
Investigation Publication Status Term Source REF        ""
```

INVESTIGATION CONTACTS

This section MUST contain zero or more values.

This section MUST contain the following labels, with the specified datatypes for values supported:

Label	Datatype	Description
Investigation Person Last Name	String	The last name of a person associated with the investigation.
Investigation Person First Name	String	Investigation Person Name
Investigation Person Mid Initials	String	The middle initials of a person associated with the investigation.
Investigation Person Email	String formatted as email	The email address of a person associated with the investigation.
Investigation Person Phone	String	The telephone number of a person associated with the investigation.
Investigation Person Fax	String	The fax number of a person associated with the investigation.
Investigation Person Address	String	The address of a person associated with the investigation.
Investigation Person Affiliation	String	The organization affiliation for a person associated with the investigation.
Investigation Person Roles	String or Ontology Annotation if accompanied by Term Accession Numbers and Term Source REFs	Term to classify the role(s) performed by this person in the context of the investigation, which means that the roles reported here need not correspond to roles held withing their affiliated organization. Multiple annotations or values attached to one person can be provided by using a semicolon (";") Unicode (U0003+B) as a separator (e.g.: submitter;funder;sponsor) .The term can be free text or from, for example, a controlled vocabulary or an ontology. If the latter source is used the Term Accession Number and Term Source REF fields below are required.
Investigation Person Roles Term Accession Number	String	The accession number from the Term Source associated with the selected term.
Investigation Person Roles Term Source REF	String	Identifies the controlled vocabulary or ontology that this term comes from. The Source REF has to match one of the Term Source Names declared in the Ontology Source Reference section.

For example, the `INVESTIGATION CONTACTS` section of an ISA-Tab `i_*.txt` file may look as follows:

```
INVESTIGATION CONTACTS
Investigation Person Last Name      "Stephen"       "Castrillo"       "Zeef"
Investigation Person First Name      "Oliver"        "Juan"         "Leo"
Investigation Person Mid Initials     "G"         "I"         "A"
Investigation Person Email        ""          ""         ""
Investigation Person Phone        ""          ""         ""
Investigation Person Fax       ""         ""         ""
Investigation Person Address       "Oxford Road, Manchester M13 9PT, UK"
→"Oxford Road, Manchester M13 9PT, UK"        "Oxford Road, Manchester M13 9PT, UK"
Investigation Person Affiliation        "Faculty of Life Sciences, Michael Smith␣
→Building, University of Manchester"        "Faculty of Life Sciences, Michael Smith␣
→Building, University of Manchester"        "Faculty of Life Sciences, Michael Smith␣
→Building, University of Manchester"
Investigation Person Roles        "corresponding author"        "author"
→"author"
Investigation Person Roles Term Accession Number        ""         ""         ""
Investigation Person Roles Term Source REF        ""         ""         ""
```

Study section

This section is organized in several subsections, described in detail below. This section also represents a **repeatable block**, which is replicated according to the number of Studies to report (i.e. two Studies, two Study blocks are represented in the Investigation file). The subsections in the block are arranged vertically; the intent being to enhance readability and presentation, and possibly to help with parsing. These subsections MUST remain within this repeatable block, although their order MAY vary; the fields MUST remain within their subsection.

These sections implement the metadata for a `Study` from the ISA Abstract Model and a list of `Assay` (i.e. `Study` and `Assay` **without** graphs; graphs are implemented in ISA-Tab as table files).

STUDY

This section MUST contain zero or one values.

This section MUST contain the following labels, with the specified datatypes for values supported:

Label	Datatype	Description
Study Identifier	String	A unique identifier, either a temporary identifier supplied by users or one generated by a repository or other database. For example, it could be an identifier complying with the LSID specification.
Study Title	String	A concise phrase used to encapsulate the purpose and goal of the study.
Study Description	String	A textual description of the study, with components such as objective or goals.
Study Submission Date	String formatted as ISO8601 date	The date on which the study is submitted to an archive.
Study Public Release Date	String formatted as ISO8601 date	The date on which the study SHOULD be released publicly.
Study File Name	String formatted as file name or URI	A field to specify the name of the Study Table file corresponding the definition of that Study. There can be only one file per cell.

For example, the `STUDY` section of an ISA-Tab `i_*.txt` file may look as follows:

```
Study Identifier        "BII-S-3"
Study Title        "Metagenomes and Metatranscriptomes of phytoplankton blooms from
↪an ocean acidification mesocosm experiment"
Study Description        "Sequencing the metatranscriptome can provide information
↪about the response of organisms to varying environmental conditions. We present a
↪methodology for obtaining random whole-community mRNA from a complex microbial
↪assemblage using Pyrosequencing. The metatranscriptome had, with minimum
↪contamination by ribosomal RNA, significant coverage of abundant transcripts, and
↪included significantly more potentially novel proteins than in the metagenome. This
↪experiment is part of a much larger experiment. We have produced 4 454
↪metatranscriptomic datasets and 6 454 metagenomic datasets. These were derived from
↪4 samples."
Study Submission Date        "2008-08-15"
Study Public Release Date        "2008-08-15"
Study File Name        "s_BII-S-3.txt"
```

STUDY DESIGN DESCRIPTORS

This section MUST contain zero or more values.

This section MUST contain the following labels, with the specified datatypes for values supported:

Label	Datatype	Description
Study Design Type	String	A term allowing the classification of the study based on the overall experimental design, e.g cross-over design or parallel group design. The term can be free text or from, for example, a controlled vocabulary or an ontology. If the latter source is used the Term Accession Number and Term Source REF fields below are required.
Study Design Type Term Accession Number	String	The accession number from the Term Source associated with the selected term.
Study Design Type Term Source REF	String	Identifies the controlled vocabulary or ontology that this term comes from. The Study Design Term Source REF has to match one the Term Source Name declared in the Ontology Source Reference section.

For example, the `STUDY DESIGN DESCRIPTORS` section of an ISA-Tab `i_*.txt` file may look as follows:

```
STUDY DESIGN DESCRIPTORS
Study Design Type              "time series design"
Study Design Type Term Accession Number        "http://purl.obolibrary.org/obo/OBI_
↪0500020"
Study Design Type Term Source REF        "OBI"
```

STUDY PUBLICATIONS

This section MUST contain zero or more values.

This section MUST contain the following labels, with the specified datatypes for values supported:

Label	Datatype	Description
Study PubMed ID	String formatted as valid PubMed ID	The PubMed IDs of the described publication(s) associated with this study.
Study Publication DOI	String formatted as valid DOI	A Digital Object Identifier (DOI) for that publication (where available).
Study Publication Author List	String	The list of authors associated with that publication.
Study Publication Title	String	The title of publication associated with the investigation.
Study Publication Status	String, or Ontology Annotation by providing accompanying Term Accession Number and Term Source REF	A term describing the status of that publication (i.e. submitted, in preparation, published).
Study Publication Status Term Accession Number	String or URI	The accession number from the Term Source associated with the selected term.
Study Publication Status Term Source REF	String	Identifies the controlled vocabulary or ontology that this term comes from. The Source REF has to match one the Term Source Name declared in the in the Ontology Source Reference section.

For example, the `STUDY PUBLICATIONS` section of an ISA-Tab `i_*.txt` file may look as follows:

```
STUDY PUBLICATIONS
Study PubMed ID        "18725995"       "18783384"
Study Publication DOI        "10.1371/journal.pone.0003042"        "10.1111/j.1462-
↪2920.2008.01745.x"
Study Publication Author List        "Gilbert JA, Field D, Huang Y, Edwards R, Li W,␣
↪Gilna P, Joint I."        "Gilbert JA, Thomas S, Cooley NA, Kulakova A, Field D,␣
↪Booth T, McGrath JW, Quinn JP, Joint I."
Study Publication Title        "Detection of large numbers of novel sequences in the␣
↪metatranscriptomes of complex marine microbial communities."        "Potential for␣
↪phosphonoacetate utilization by marine bacteria in temperate coastal waters."
Study Publication Status        "indexed in PubMed"        "indexed in PubMed"
Study Publication Status Term Accession Number        ""        ""
Study Publication Status Term Source REF        ""        ""
```

STUDY FACTORS

This section MUST contain zero or more values.

This section MUST contain the following labels, with the specified datatypes for values supported:

Label	Datatype	Description
Study Factor Name	String	The name of one factor used in the Study and/or Assay files. A factor corresponds to an independent variable manipulated by the experimentalist with the intention to affect biological systems in a way that can be measured by an assay. The value of a factor is given in the Study or Assay file, accordingly. If both Study and Assay have a Factor Value, these must be different.
Study Factor Type	String	A term allowing the classification of this factor into categories. The term can be free text or from, for example, a controlled vocabulary or an ontology. If the latter source is used the Term Accession Number and Term Source REF fields below are required.
Study Factor Type Term Accession Number	String	The accession number from the Term Source associated with the selected term.
Study Factor Type Term Source REF	String	Identifies the controlled vocabulary or ontology that this term comes from. The Source REF has to match one of the Term Source Name declared in the Ontology Source Reference section.

For example, the STUDY FACTORS section of an ISA-Tab i_*.txt file may look as follows:

```
STUDY FACTORS
Study Factor Name        "dose"        "compound"        "collection time"
Study Factor Type        "dose"        "chemical substance"        "time"
Study Factor Type Term Accession Number        "http://www.ebi.ac.uk/efo/EFO_0000428
↪"        "http://purl.obolibrary.org/obo/CHEBI_59999"        "http://purl.
↪obolibrary.org/obo/PATO_0000165"
Study Factor Type Term Source REF        "EFO"        "CHEBI"        "PATO"
```

STUDY ASSAYS

This section MUST contain zero or more values.

This section MUST contain the following labels, with the specified datatypes for values supported:

Label	Datatype	Description
Study Assay Measurement Type	String	A term to qualify the endpoint, or what is being measured (e.g. gene expression profiling or protein identification). The term can be free text or from, for example, a controlled vocabulary or an ontology. If the latter source is used the Term Accession Number and Term Source REF fields below are required.
Study Assay Measurement Type Term Accession Number	String	The accession number from the Term Source associated with the selected term.
Study Assay Measurement Type Term Source REF	String	The Source REF has to match one of the Term Source Name declared in the Ontology Source Reference section.
Study Assay Technology Type	String	Term to identify the technology used to perform the measurement, e.g. DNA microarray, mass spectrometry. The term can be free text or from, for example, a controlled vocabulary or an ontology. If the latter source is used the Term Accession Number and Term Source REF fields below are required.
Study Assay Technology Type Term Accession Number	String	The accession number from the Term Source associated with the selected term.
Study Assay Technology Type Term Source REF	String	Identifies the controlled vocabulary or ontology that this term comes from. The Source REF has to match one of the Term Source Names declared in the Ontology Source Reference section.
Study Assay Technology Platform	String	Manufacturer and platform name, e.g. Bruker AVANCE
Study Assay File Name	String	A field to specify the name of the Assay Table file corresponding the definition of that assay. There can be only one file per cell.

For example, the STUDY ASSAYS section of an ISA-Tab `i_*.txt` file may look as follows:

```
STUDY ASSAYS
Study Assay File Name        "a_gilbert-assay-Gx.txt"        "a_gilbert-assay-Tx.txt"
Study Assay Measurement Type        "metagenome sequencing"        "transcription␣
↪profiling"
Study Assay Measurement Type Term Accession Number        ""        ""
Study Assay Measurement Type Term Source REF        "OBI"        "OBI"
Study Assay Technology Type        "nucleotide sequencing"        "nucleotide␣
↪sequencing"
Study Assay Technology Type Term Accession Number        ""        ""
Study Assay Technology Type Term Source REF        "OBI"        "OBI"
Study Assay Technology Platform        "454 GS FLX"        "454 GS FLX"
```

STUDY PROTOCOLS

This section MUST contain zero or more values.

This section MUST contain the following labels, with the specified datatypes for values supported:

Label	Datatype	Description
Study Protocol Name	String	The name of the protocols used within the ISA-Tab document. The names are used as identifiers within the ISA-Tab document and will be referenced in the Study and Assay files in the Protocol REF columns. Names can be either local identifiers, unique within the ISA Archive which contains them, or fully qualified external accession numbers.
Study Protocol Type	String	Term to classify the protocol. The term can be free text or from, for example, a controlled vocabulary or an ontology. If the latter source is used the Term Accession Number and Term Source REF fields below are required.
Study Protocol Type Term Accession Number	String	The accession number from the Term Source associated with the selected term.
Study Protocol Type Term Source REF	String	Identifies the controlled vocabulary or ontology that this term comes from. The Source REF has to match one of the Term Source Name declared in the Ontology Source Reference section.
Study Protocol Description	String	A free-text description of the protocol.
Study Protocol URI	String	Pointer to protocol resources external to the ISA-Tab that can be accessed by their Uniform Resource Identifier (URI).
Study Protocol Version	String	An identifier for the version to ensure protocol tracking.
Study Protocol Parameters Name	String	A semicolon-delimited (";") list of parameter names, used as an identifier within the ISA-Tab document. These names are used in the Study and Assay files (in the "Parameter Value []" column heading) to list the values used for each protocol parameter. Refer to section Multiple values fields in the Investigation File on how to encode multiple values in one field and match term sources
Study Protocol Parameters Term Accession Number	String	The accession number from the Term Source associated with the selected term.
Study Protocol Parameters Term Source REF	String	Identifies the controlled vocabulary or ontology that this term comes from. The Source REF has to match one of the Term Source Name declared in the Ontology Source Reference section.
Study Protocol Components Name	String	A semicolon-delimited (";") list of a protocol's components; e.g. instrument names, software names, and reagents names. Refer to section Multiple values fields in the Investigation File on how to encode multiple components in one field and match term sources.
Study Protocol Components Type	String	Term to classify the protocol components listed for example, instrument, software, detector or reagent. The term can be free text or from, for example, a controlled vocabulary or an ontology. If the latter source is used the Term Accession Number and Term Source REF fields below are required.
Study Protocol Components Type Term Accession Number	String	The accession number from the Source associated to the selected terms.
Study Protocol Components Type Term Source REF	String	Identifies the controlled vocabulary or ontology that this term comes from. The Source REF has to match a Term Source Name previously declared in the ontology section

For example, the STUDY PROTOCOLS section of an ISA-Tab i_*.txt file may look as follows:

```
STUDY PROTOCOLS
Study Protocol Name        "environmental material collection – standard procedure 1
→"        "nucleic acid extraction – standard procedure 2"        "mRNA extraction –␣
→standard procedure 3"        "genomic DNA extraction – standard procedure 4"
→"reverse transcription – standard procedure 5"        "library construction"
→"pyrosequencing – standard procedure 6"        "sequence analysis – standard␣
→procedure 7"
Study Protocol Type        "sample collection"        "nucleic acid extraction
→"        "nucleic acid extraction"        "nucleic acid extraction"        "reverse␣
→transcription"        "library construction"        "nucleic acid sequencing
→"        "data transformation"
Study Protocol Type Term Accession Number        ""        ""        ""        "
→"        ""        ""        ""        ""
Study Protocol Type Term Source REF        ""        ""        ""        ""        "
→"        ""        ""        ""
Study Protocol Description        "Waters samples were prefiltered through a 1.6 um␣
→GF/A glass fibre filter to reduce Eukaryotic contamination. Filtrate was then␣
→collected on a 0.2 um Sterivex (millipore) filter which was frozen in liquid␣
→nitrogen until nucelic acid extraction. CO2 bubbled through 11000 L mesocosm to␣
→simulate ocean acidification predicted conditions. Then phosphate and nitrate were␣
→added to induce a phytoplankton bloom."        "Total nucleic acid extraction was␣
→done as quickly as possible using the method of Neufeld et al, 2007."        "RNA␣
→MinElute + substrative Hybridization + MEGAclear For transcriptomics, total RNA was␣
→separated from the columns using the RNA MinElute clean-up kit (Qiagen) and checked␣
→for integrity of rRNA using an Agilent bioanalyser (RNA nano6000 chip). High␣
→integrity rRNA is essential for subtractive hybridization. Samples were treated␣
→with Turbo DNA-free enzyme (Ambion) to remove contaminating DNA. The rRNA was␣
→removed from mRNA by subtractive hybridization (Microbe Express Kit, Ambion), and␣
→absence of rRNA and DNA contamination was confirmed using the Agilent bioanalyser.␣
→The mRNA was further purified with the MEGAclearTM kit (Ambion). Reverse␣
→transcription of mRNA was performed using the SuperScript III enzyme (Invitrogen)␣
→with random hexamer primers (Promega). The cDNA was treated with RiboShredderTM␣
→RNase Blend (Epicentre) to remove trace RNA contaminants. To improve the yield of␣
→cDNA, samples were subjected to random amplification using the GenomiPhi V2 method␣
→(GE Healthcare). GenomiPhi technology produces branched DNA molecules that are␣
→recalcitrant to the pyrosequencing methodology. Therefore amplified samples were␣
→treated with S1 nuclease using the method of Zhang et al.2006."        ""
→"superscript+random hexamer primer"        ""        "1. Sample Input and␣
→Fragmentation: The Genome Sequencer FLX System supports the sequencing of samples␣
→from a wide variety of starting materials including genomic DNA, PCR products, BACs,
→ and cDNA. Samples such as genomic DNA and BACs are fractionated into small, 300-␣
→to 800-base pair fragments. For smaller samples, such as small non-coding RNA or␣
→PCR amplicons, fragmentation is not required. Instead, short PCR products amplified␣
→using Genome Sequencer fusion primers can be used for immobilization onto DNA␣
→capture beads as shown below."        ""
Study Protocol URI        ""        ""        ""        ""        ""        ""
→""        ""
Study Protocol Version        ""        ""        ""        ""        ""        "
→"        ""        ""
Study Protocol Parameters Name        "filter pore size"        ""        ""        "
→"        ""        "library strategy;library layout;library selection"
→"sequencing instrument"        ""
Study Protocol Parameters Name Term Accession Number        ""        ""        "
→"        ""        ";;"        ""        ""
Study Protocol Parameters Name Term Source REF        ""        ""        ""        "
→"        ""        ";;"        ""        ""
Study Protocol Components Name        ""        ""        ""        ""        "
→"        ""        ""        ""
```

```
Study Protocol Components Type          ""          ""          ""          ""          "
↪"          ""          ""          ""
Study Protocol Components Type Term Accession Number          ""          ""          "
↪"          ""          ""          ""          ""          ""
Study Protocol Components Type Term Source REF          ""          ""          ""          "
↪"          ""          ""          ""          ""
```

STUDY CONTACTS

This section MUST contain zero or more values.

This section MUST contain the following labels, with the specified datatypes for values supported:

abel	Datatype	Description
Study Person Last Name	String	The last name of a person associated with the study.
Study Person First Name	String	Study Person Name
Study Person Mid Initials	String	The middle initials of a person associated with the study.
Study Person Email	String formatted as email	The email address of a person associated with the study.
Study Person Phone	String	The telephone number of a person associated with the study.
IStudy Person Fax	String	The fax number of a person associated with the study.
Study Person Address	String	The address of a person associated with the study.
Study Person Affiliation	String	The organization affiliation for a person associated with the study.
Study Person Roles	String or Ontology Annotation if accompanied by Term Accession Numbers and Term Source REFs	Term to classify the role(s) performed by this person in the context of the study, which means that the roles reported here need not correspond to roles held withing their affiliated organization. Multiple annotations or values attached to one person can be provided by using a semicolon (";") Unicode (U0003+B) as a separator (e.g.: submitter;funder;sponsor) .The term can be free text or from, for example, a controlled vocabulary or an ontology. If the latter source is used the Term Accession Number and Term Source REF fields below are required.
Study Person Roles Term Accession Number	String	The accession number from the Term Source associated with the selected term.
Study Person Roles Term Source REF	String	Identifies the controlled vocabulary or ontology that this term comes from. The Source REF has to match one of the Term Source Names declared in the Ontology Source Reference section.

For example, the STUDY CONTACTS section of an ISA-Tab i_*.txt file may look as follows:

```
Study Person Last Name          "Gilbert"          "Field"          "Huang"          "Edwards
↪"          "Li"          "Gilna"          "Joint"
Study Person First Name          "Jack"          "Dawn"          "Ying"          "Rob"
↪"Weizhong"          "Paul"          "Ian"
Study Person Mid Initials          "A"          ""          ""          ""          ""          "
↪"          ""
Study Person Email          "jagi@pml.ac.uk"          ""          ""          ""          "
↪"          ""          ""
```

```
Study Person Phone        ""        ""        ""        ""        ""        ""
↪""
Study Person Fax          ""        ""        ""        ""        ""        ""        ""
Study Person Address          "Prospect Place, Plymouth, United Kingdom"        "CEH␣
↪Oxford, Oxford, United Kingdom"        "San Diego State University, San Diego,␣
↪California, United States of America"        "Argonne National Laboratory, Argonne,␣
↪Illinois, United States of America"        "San Diego State University, San Diego,␣
↪California, United States of America"        "San Diego State University, San Diego,␣
↪ California, United States of America"        "Prospect Place, Plymouth, United␣
↪Kingdom"
Study Person Affiliation        "Plymouth Marine Laboratory"        "NERC Centre for␣
↪Ecology and Hydrology"        "California Institute for Telecommunications and␣
↪Information Technology"        "Department of Computer Science, Mathematics and␣
↪Computer Science Division,"        "California Institute for Telecommunications and␣
↪Information Technology"        "California Institute for Telecommunications and␣
↪Information Technology"        "Plymouth Marine Laboratory"
Study Person Roles        "principal investigator role;SRA Inform On Status;SRA␣
↪Inform On Error"        "principal investigator role"        "principal␣
↪investigator role"        "principal investigator role"        "principal␣
↪investigator role"        "principal investigator role"        "principal␣
↪investigator role"
Study Person Roles Term Accession Number        ";;"        ""        ""        "
↪"        ""        ""        ""
Study Person Roles Term Source REF        ";;"        ""        ""        ""        "
↪"        ""        ""
```

1.2.3 Study and Assay files

`Study` and `Assay` Table files are structure with fields organized on a per-row basis. The first row MUST be used for column headers. Generally, objects such as Materials and Processes are indicated with `<entity> Name`, for example `Sample Name` to indicate a sample, or `Assay Name` to indicate a named instance of a process that has been applied. Object properties MUST follow this column, where materials MAY have Characteristics and Processes have MAY have Parameter Values. Both `Characteristics` and `Parameter Values` MUST be of type string, numeric, or an `Ontology Annotation`. `<entity> File` MAY be used to indicate a data file node.

> **Attention:** Comments are also allowed in Study and Assay files, in a similar fashion to how they are used in the Investigation file. Columns headed with `Comment[<comment name>]` MAY appear after any named node in the Study and Assay files (e.g. if `Comment[ORCID ID]` appears **after** the `Source Name` column, we know that the comment regarding `ORCID ID` applies to the relevant Source node based on the row.

Specific types of nodes are specified in the Assay Table file section below.

Ontology Annotations

Where a value is an `Ontology Annotation` in a table file, `Term Accession Number` and `Term Source REF` fields MUST follow the column cell in which the value is entered. For example, a characteristic type `Organism` with a value of `Homo sapiens` can be qualified with an `Ontology Annotation` of a term from NCBI Taxonomy as follows:

Characteristics[Organism]	Term Source REF	Term Accession Number
Homo sapiens	NCBITaxon	http://.../NCBITAXON/9606

An `Ontology Annotation` MAY be applied to any appropriate `Characteristics` or `Parameter Value`.

This implements `Ontology Annotation` from the ISA Abstract Model.

Unit

Where a value is numeric, a `Unit` MAY be used to qualify the quantity. In this case, following the column in which a `Unit` is used, a `Unit` heading MUST be present, and MAY be further annotated as an `Ontology Annotation`.

For example, to qualify the value `300` with a `Unit Kelvin` qualified as an `Ontology Annotation` from the Units Ontology declared in the Ontology Sources with `UO`:

Parameter Value[Temperature]	Unit	Term Source REF	Term Accession Number
300	Kelvin	UO	http://.../obo/UO_0000012

Processes

A `Process` MUST be indicated with the column heading `Protocol REF`. The value of `Protocol REF` cells MUST reference a `Protocol` declared in the investigation file.

Characteristics

`Characteristics` are used as an attribute column following `Source Name`, `Sample Name`. This column contains terms describing each material according to the characteristics category indicated in the column header in the pattern `Characteristics [<category term>]`. For example, a column header `Characteristics [organ part]` would contain terms describing an organ part. `Characteristics` SHOULD be used as an attribute column following `Source Name`, or `Sample Name`. The value MUST be free text, numeric, or an `Ontology Annotation`.

For example, a characteristic type Organism with a value of Homo sapiens can be qualified with an Ontology Annotation of a term from NCBI Taxonomy as follows:

Characteristics[organ part]	Term Source REF	Term Accession Number
Liver	MeSH	D008099

Factor Value

A factor is an independent variable manipulated by an experimentalist with the intention to affect biological systems in a way that can be measured by an assay. This field holds the actual data for the `Factor Value` named between the square brackets (as declared in the Investigation file) so MUST match; for example, `Factor Value [compound]`. The value MUST be free text, numeric, or an `Ontology Annotation`.

Factor Value[Gender]	Term Source REF	Term Accession Number
Male	MeSH	D008297

Study Table file

The `Study` file contains contextualizing information for one or more assays, for example; the subjects studied; their source(s); the sampling methodology; their characteristics; and any treatments or manipulations performed to prepare the specimens.

For a full example of a complete Study Table file, please see https://git.io/vD1vi

Study Table files SHOULD have file names corresponding to the pattern `s_*.txt`, e.g. `s_Study01.txt`

In Study files, there are two types of `Material` nodes implemented: `Source` and `Sample`.

These are linked with a Process node, incidcated with a value under a column headed `Protocol REF` that MUST be of a Protocol type that is of a type `sample collection` declared in the Investigation file.

A `Source` MUST be indicated with the column heading `Source Name`.

The protocol referenced MUST be of protocol type `sample collection`.

A `Sample` MUST be indicated with the column heading `Sample Name`.

For example, a simple source to sample may be represented as:

Source Name	Protocol REF	Sample Name
source1	sample collection	sample1

Where a graph splits or pools, we use the `Name` column to represent the same nodes.

For example, if we split a source into two samples, we might represent this as:

Source Name	Protocol REF	Sample Name
source1	sample collection	sample1
source1	sample collection	sample2

If we pool two sources into a single sample, we might represent this as:

Source Name	Protocol REF	Sample Name
source1	sample collection	sample1
source2	sample collection	sample1

Node properties, such as `Characteristics` (for `Material` nodes), `Parameter Value` (for `Process` nodes) and additional `Name` columns for special cases of `Process` node to disambiguate `Protocol REF` entries of MUST follow the named node of context.

For example,

```
"Source Name"        "Characteristics[organism]"        "Term Source REF"
↪"Term Accession Number"        "Characteristics[strain]"        "Term Source REF"
↪"        "Term Accession Number"        "Characteristics[genotype]"        "Term␣
↪Source REF"        "Term Accession Number"        "Characteristics[mating type]
↪"        "Term Source REF"        "Term Accession Number"        "Protocol REF
↪"        "Sample Name"
"Saccharomyces cerevisiae FY1679 "        "Saccharomyces cerevisiae (Baker's yeast)
↪"        "NEWT"        ""        "FY1679"        ""        ""        "KanMx4 MATa/
↪MATalpha ura3-52/ura3-52 leu2-1/+trp1-63/+his3-D200/+ hoD KanMx4/hoD"        "
↪"        ""        "mating_type_alpha"        ""        ""        "growth"
↪"NZ_0hrs_Grow_1"
"Saccharomyces cerevisiae FY1679 "        "Saccharomyces cerevisiae (Baker's yeast)
↪"        "NEWT"        ""        "FY1679"        ""        ""        "KanMx4 MATa/
↪MATalpha ura3-52/ura3-52 leu2-1/+trp1-63/+his3-D200/+ hoD KanMx4/hoD"        "
↪"        ""        "mating_type_alpha"        ""        ""        "growth"
↪"NZ_0hrs_Grow_2"
```

The Study Table file implements the Study graphs from the ISA Abstract Model.

Assay Table file

The `Assay` file represents a portion of the experimental graph (i.e., one part of the overall structure of the workflow); each `Assay` file must contain assays of the same type, defined by the type of measurement (e.g. gene expression) and the technology employed (e.g. DNA microarray). Assay-related information includes protocols, additional information relating to the execution of those protocols and references to data files (whether raw or derived).

For a full example of a complete Assay Table file, please see https://git.io/vD1vy.

Assay Table files SHOULD have file names corresponding to the pattern `a_*.txt`, e.g. `a_Assay01.txt`

A `Sample` MUST be provided as the first node in the experimental graph, indicated with the column heading `Sample Name`.

`Protocol REF` columns MUST be used to indicate `Process` nodes, with values referencing protocols declared in the Investigation file. The `Protocol REF` column MAY be qualified with `Parameter Value [<parameter term>]`, ``Performer` and `Date`. The `Parameter Value [<parameter term>]` field allows reporting the values taken by the parameter when applying in a protocol. Note that the term between [ ] must map to one (and only one) of parameters defined in the investigation file. Values can be qualitative or quantitative. The `Performer` field reports the name of the operator who carried out the protocol. This allows account to be taken of operator effects and can be part of a quality control data tracking. `Date` is the date on which a protocol is performed. This allows account to be taken of day effects and can be part of a quality control data tracking. Dates should be reported in ISO8601 format.

`Extract Name` MUST be used as an identifier for a Extract Material node within an `Assay` file. This column contains user-defined names for each portion of extracted material. Extracts MAY be qualified with `Characteristics`, `Material Type` and `Description`.

`Labeled Extract` Name MUST be used as an identifier for a Labeled Extract Material node within an `Assay` file. Labeled Extracts MAY be qualified with `Label`, `Characteristics`, `Material Type`, `Description`.

`Assay Name` MUST be used as an identifier for user-defined names for each assay.

`Image File`, `Raw Data File` or `Derived Data File` column heading MUST correspond to a relevant `Data` node to provide names or URIs of file locations. For submission or transfer, files MAY be packed with ISA-Tab files.

`Data Transformation Name` MUST be used as an identifier for a user-defined name for each data transformation `Process` applied.

`Normalization Name` MUST be used as an identifier for a user-defined name for each normalization `Process` applied.

Splitting and pooling is allowed as per the examples given in Study Table file.

For example,

```
"Sample Name"        "Protocol REF"         "Protocol REF"          "Extract Name
→"        "Material Type"        "Term Source REF"         "Term Accession Number
→"        "Protocol REF"        "Parameter Value[library strategy]"
→"Parameter Value[library selection]"        "Parameter Value[library layout]
→"        "Protocol REF"        "Parameter Value[sequencing instrument]"
→"Assay Name"        "Raw Data File"        "Comment[TraceDB]"
"GSM255770"        "nucleic acid extraction – standard procedure 2"        "genomic␣
→DNA extraction – standard procedure 4"        "GSM255770.e1"
→"deoxyribonucleic acid"        "CHEBI"        "http://purl.obolibrary.org/obo/CHEBI_
→16991"        "library construction"        "WGS"        "RANDOM"        "SINGLE
→"        "pyrosequencing – standard procedure 6"        "454 GS FLX"        "assay1
→"        "EWOEPZA01.sff"        "ftp://ftp.ncbi.nih.gov/pub/TraceDB/ShortRead/
→SRA000266/EWOEPZA01.sff"
```

<div align="right">(continues on next page)</div>

```
"GSM255771"          "nucleic acid extraction - standard procedure 2"          "genomic␣
↪DNA extraction - standard procedure 4"        "GSM255771.e1"
↪"deoxyribonucleic acid"        "CHEBI"          "http://purl.obolibrary.org/obo/CHEBI_
↪16991"        "library construction"          "WGS"        "RANDOM"        "SINGLE
↪"          "pyrosequencing - standard procedure 6"          "454 GS FLX"          "assay2
↪"        "EWOEPZA02.sff"          "ftp://ftp.ncbi.nih.gov/pub/TraceDB/ShortRead/
↪SRA000266/EWOEPZA02.sff"
"GSM255772"          "nucleic acid extraction - standard procedure 2"          "genomic␣
↪DNA extraction - standard procedure 4"        "GSM255772.e1"
↪"deoxyribonucleic acid"        "CHEBI"          "http://purl.obolibrary.org/obo/CHEBI_
↪16991"        "library construction"          "WGS"        "RANDOM"        "SINGLE
↪"          "pyrosequencing - standard procedure 6"          "454 GS FLX"          "assay3.
↪1"        "EXHS9OF01.sff"          "ftp://ftp.ncbi.nih.gov/pub/TraceDB/ShortRead/
↪SRA000266/EXHS9OF01.sff"
"GSM255772"          "nucleic acid extraction - standard procedure 2"          "genomic␣
↪DNA extraction - standard procedure 4"        "GSM255772.e1"
↪"deoxyribonucleic acid"        "CHEBI"          "http://purl.obolibrary.org/obo/CHEBI_
↪16991"        "library construction"          "WGS"        "RANDOM"        "SINGLE
↪"          "pyrosequencing - standard procedure 6"          "454 GS FLX"          "assay3.
↪2"        "EX398L102.sff"          "ftp://ftp.ncbi.nih.gov/pub/TraceDB/ShortRead/
↪SRA000266/EX398L102.sff"
"GSM255773"          "nucleic acid extraction - standard procedure 2"          "genomic␣
↪DNA extraction - standard procedure 4"        "GSM255773.e1"
↪"deoxyribonucleic acid"        "CHEBI"          "http://purl.obolibrary.org/obo/CHEBI_
↪16991"        "library construction"          "WGS"        "RANDOM"        "SINGLE
↪"          "pyrosequencing - standard procedure 6"          "454 GS FLX"          "assay4.
↪1"        "EXHS9OF02.sff"          "ftp://ftp.ncbi.nih.gov/pub/TraceDB/ShortRead/
↪SRA000266/EXHS9OF02.sff"
"GSM255773"          "nucleic acid extraction - standard procedure 2"          "genomic␣
↪DNA extraction - standard procedure 4"        "GSM255773.e1"
↪"deoxyribonucleic acid"        "CHEBI"          "http://purl.obolibrary.org/obo/CHEBI_
↪16991"        "library construction"          "WGS"        "RANDOM"        "SINGLE
↪"          "pyrosequencing - standard procedure 6"          "454 GS FLX"          "assay4.
↪2"        "EX398L101.sff"          "ftp://ftp.ncbi.nih.gov/pub/TraceDB/ShortRead/
↪SRA000266/EX398L101.sff"
```

The Assay Table file implements the `Assay` graphs from the ISA Abstract Model.

Special cases

Assay with technology type: DNA microarray hybridization

If an Assay being described has a technology type of DNA microarray hybridization, the following additional nodes MAY apply.

Hybridization Assay Name (in place of Assay Name) Used as an identifier within the Assay file. This column contains an user-defined name for each hybridization. Qualifying headers for Hybridization Assay Name item include Array Design REF or Array Design File.

Scan Name Used as an identifier within the Assay file. This column contains a user-defined name for each Scan event.

Array Data File (in place of Raw Data File) Column to provide name (or URI) of raw array data files.

Derived Array Data File (in place of Derived Data File) Column to provide name (or URI) of data files resulting from data transformation or processing.

Array Data Matrix File Column to provide name (or URI) of raw data matrix files.

Derived Array Data Matrix File Column to provide name (or URI) of processed data matrix files, resulting from data transformation or processing. Where data from multiple hybridizations is stored in a single file, the data should be mapped to the appropriate hybridization (or scan, or normalization) via the Data Matrix format itself

Array Design File Column to provide name of file containing the array design, used for a particular hybridization. For submission or transfer, ADF files can be packaged with ISA-TAB files into an ISArchive, see section 2.4.

Array Design REF This column is used to reference the identifier (or accession number) of an existing array design.

Assay file with technology type: Gel electrophoresis

If an Assay being described has a technology type of Gel electrophoresis, the following additional nodes MAY apply.

Gel Electrophoresis Assay Name (in place of Assay Name) Used as an identifier within the Assay file. This column contains user-defined names for each electrophoresis gel assay. For 2-dimensional gels, the following qualifying headers can be used instead:

First Dimension The term can be free text or from, for example, a controlled vocabulary or an ontology. If the latter source is used the Term Accession Number and Term Source REF fields are required.

Second Dimension The term can be free text or from, for example, a controlled vocabulary or an ontology. If the latter source is used the Term Accession Number and Term Source REF fields are required.

Scan Name Used as an identifier within the Assay file. This column contains user-defined names for each Scan event.

Spot Picking File Column to provide name (or URI) of files file holding protein spot coordinates and metadata for use by spot picking instruments.

Assay file with technology type: Mass Spectrometry (MS)

If an Assay being described has a technology type of Mass Spectrometry, the following additional nodes MAY apply.

MS Assay Name (in place of Assay Name) Used as an identifier within the Assay file. This column contains user-defined names for each MS Assay.

Raw Spectral Data File (in place of Raw Data File) Column to provide name (or URI) of 'raw' spectral data files.

Derived Spectral Data File (in place of Derived Data File) Column to provide name (or URI) of derived spectral data files, resulting from data transformation or processing.

When Mass Spectrometry is used in proteomics the following data files are required, according to PSI specifications and Pride submission requirements (6, 10):

Peptide Assignment File Column to provide name (or URI) of file(s) containing peptide assignments.

Protein Assignment File Column to provide name (or URI) of file(s) containing protein assignments.

Post Translational Modification Assignment File Column to provide name (or URI) of file(s) containing posited post-translational modifications.

Capturing data resulting from the use of mass spectrometry in metabol/nomics requires a settled definition for a Metabolite Assignment File (inter alia); such a file is currently under development in collaboration with the Metabolomics Standards Initiative (MSI).

Data Files

ISA-Tab focuses on structuring experimental metadata; raw and derived data files are considered as external files. The Assay file can refer to one or more of these external data files. For guidelines on how to format these data files, users should refer to the relevant standards group or reference repository.

For submission or transfer, ISA-Tab files and associated data files MAY be packaged into an ISArchive, a zip file containing all the files together.

1.3 ISA-JSON format

Important: As a pre-requisite to reading this specification, please make sure you have read and understood the *ISA Abstract Model* that the ISA-Json format is based on.

For detail on ISA framework terminology, please read the *ISA Abstract Model specification*.

This document describes the ISA Abstract Model reference implementation specified in the JSON format [RFC7159]. The JavaScript Object Notation (JSON) [RFC7159] is a text format for serializing structured data. Objects are rendered as an unordered collection of name-value pairs. The JSON Schema (see [JSON Schema], [JSON Schema Core], and [JSON Schema Validation]) defines a JSON format for describing JSON formats.

Below we provide the *schemas* and the *content rules* for valid ISA-JSON documents. Full examples of ISA content as ISA-JSON can be found in the ISA datasets repository, here https://git.io/vD1vx.

We recommend that you study these to better understand the structure of ISA-JSON documents.

1.3.1 Format

Files SHOULD be encoded using UTF-8.

All ISA-JSON content regarding multiple `Study` and `Assay` should fall under one `Investigation` JSON structure, therefore should be recorded in a single JSON file. The JSON file SHOULD have a `.json` extension.

Dates SHOULD be supplied in the ISO8601 format `YYYY-MM-DD`.

For maximal portability file names should only contain only ASCII characters not excluded already (that is `A-Za-z0-9._!#$%&+,;=@^(){}'[]` - we exclude space as many utilities do not accept spaces in file paths): non-English alphabetic characters cannot be guaranteed to be supported in all locales. It would be good practice to avoid the shell metacharacters `(){}'[]$."`.

1.3.2 Schemas

The ISA-JSON schemas define the structure of the ISA-JSON objects that implement the ISA Abstract Model. Here we list the JSON schemas with their corresponding model entity, and provide show the schema implemented.

You can also find these schemas in Github at https://git.io/fARJs

investigation_schema.json

This schema implements Investigation from the ISA Abstract Model.

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA investigation schema",
    "description" : "JSON-schema representing an investigation in the ISA model",
    "type" : "object",
    "properties" : {
        "@id": { "type": "string", "format": "uri" },
        "filename": { "type" : "string"},
        "identifier" : { "type" : "string" },
        "title" : { "type" : "string"},
        "description" : { "type" : "string"},
        "submissionDate" : { "type" : "string", "format" : "date-time"},
        "publicReleaseDate" : { "type" : "string", "format" : "date-time"},
        "ontologySourceReferences" : {
            "type" : "array",
            "items" : {
                "$ref": "ontology_source_reference_schema.json#"
            }
        },
        "publications" : {
            "type" : "array",
            "items" : {
                "$ref": "publication_schema.json#"

            }
        },
        "people" : {
            "type" : "array",
            "items" : {
                "$ref": "person_schema.json#"

            }
        },
        "studies" : {
            "type" : "array",
            "items" : {
                "$ref": "study_schema.json#"

            }
        },
        "comments" : {
            "type": "array",
            "items": {
                "$ref": "comment_schema.json#"
            }
        }
    },
    "additionalProperties": false
}
```

study_schema.json

This schema implements Study from the ISA Abstract Model.

Schema:

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "title": "Study JSON Schema",
  "description": "JSON Schema describing an Study",
  "@context": {
    "@base": "http://purl.org/isaterms/",
    "xsd": "http://www.w3.org/2001/XMLSchema#"
  },
  "type": "object",
  "properties": {
    "@id": { "type": "string", "format": "uri" },
    "filename" : { "type" : "string"},
    "identifier" : { "type" : "string" },
    "title" : { "type" : "string"},
    "description" : { "type" : "string"},
    "submissionDate" : { "type" : "string", "format" : "date-time"},
    "publicReleaseDate" : { "type" : "string", "format" : "date-time"},
    "publications" : {
      "type" : "array",
      "items" : {
        "$ref": "publication_schema.json#"
      }
    },
    "people" : {
      "type" : "array",
      "items" : {
        "$ref": "person_schema.json#"

      }
    },
    "studyDesignDescriptors":{
      "type": "array",
      "items" : {
        "$ref": "ontology_annotation_schema.json#"
      }
    },
    "protocols" : {
      "type": "array",
      "items" : {
        "$ref": "protocol_schema.json#"
      }
    },
    "materials": {
      "type": "object",
      "properties": {
        "sources": {
          "type": "array",
          "items": {
            "$ref": "source_schema.json#"
          }
        },
        "samples": {
          "type": "array",
          "items": {
            "$ref": "sample_schema.json#"
          }
        },
```

```
        "otherMaterials": {
          "type": "array",
          "items": {
            "$ref": "material_schema.json#"
          }
        }
      }
    },
    "processSequence": {
      "type": "array",
      "items" : {
        "$ref" : "process_schema.json#"
      }
    },
    "assays" : {
      "type": "array",
      "items" : {
        "$ref": "assay_schema.json#"
      }
    },
    "factors": {
      "type": "array",
      "items": {
        "$ref": "factor_schema.json#"
      }
    },
    "characteristicCategories": {
      "description": "List of all the characteristics categories (or material␣
→attributes) defined in the study, used to avoid duplication of their declaration␣
→when each material_attribute_value is created. ",
      "type": "array",
      "items": {
        "$ref": "material_attribute_schema.json#"
      }
    },
    "unitCategories": {
      "description": "List of all the unitsdefined in the study, used to avoid␣
→duplication of their declaration when each value is created. ",
      "type": "array",
      "items": {
        "$ref": "ontology_annotation_schema.json#"
      }
    },
    "comments" : {
      "type": "array",
      "items": {
        "$ref": "comment_schema.json#"
      }
    }
  },
  "additionalProperties": false
}
```

assay_schema.json

This schema implements Assay from the ISA Abstract Model.

Schema:

```json
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "title": "Assay JSON Schema",
  "name": "Assay JSON Schema",
  "description": "JSON Schema describing an Assay",
  "@context": {
    "@base": "http://purl.org/isaterms/",
    "xsd": "http://www.w3.org/2001/XMLSchema#"
  },
  "type": "object",
  "properties": {
    "@id": { "type": "string", "format": "uri" },
    "comments" : {
      "type": "array",
      "items": {
        "$ref": "comment_schema.json#"
      }
    },
    "filename" : { "type" : "string" },
    "measurementType" : {
      "$ref": "ontology_annotation_schema.json#"
    },
    "technologyType" : {
      "type" : "object",
      "properties": {
        "ontologyAnnotation" : {
          "$ref": "ontology_annotation_schema.json#"
        }
      }
    },
    "technologyPlatform" : { "type" : "string"},
    "dataFiles" : {
      "type": "array",
      "items" : {
        "$ref": "data_schema.json#"
      }
    },
    "materials": {
      "type": "object",
      "properties": {
        "samples": {
          "type": "array",
          "items": {
            "$ref": "sample_schema.json#"
          }
        },
        "otherMaterials": {
          "type": "array",
          "items": {
            "$ref": "material_schema.json#"
          }
        }
      }
    },
    "characteristicCategories": {
      "description": "List of all the characteristics categories (or material␣
↪attributes) defined in the study, used to avoid duplication of their declaration␣
↪when each material_attribute_value is created. ",
```

```json
      "type": "array",
      "items": {
        "$ref": "material_attribute_schema.json#"
      }
    },
    "unitCategories": {
      "description": "List of all the unitsdefined in the study, used to avoid
→duplication of their declaration when each value is created. ",
      "type": "array",
      "items": {
        "$ref": "ontology_annotation_schema.json#"
      }
    },
    "processSequence": {
      "type": "array",
      "items" : {
        "$ref" : "process_schema.json#"
      }
    }
  },
  "additionalProperties": false
}
```

comment_schema.json

This schema implements the ability to annotate objects with user-defined comments.

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title": "ISA comment schema - it corresponds to ISA Comment[] construct",
    "description": "JSON-schema representing a comment in the ISA model",
    "type": "object",
    "properties": {
        "@id": { "type": "string", "format": "uri" },
        "name": {
            "type": "string"
        },
        "value": {
            "type": "string"
        }
    },
    "additionalProperties": false
}
```

data_schema.json

This schema implements Data from the ISA Abstract Model.

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
```

```
    "title": "ISA data schema",
    "description": "JSON-schema representing a data file in the ISA model",
    "type": "object",
    "properties": {
        "@id": { "type": "string", "format": "uri" },
        "name": {
            "type": "string"
        },
        "type": {
            "type": "string",
            "enum": [
                "Raw Data File",
                "Derived Data File",
                "Image File"
            ]
        },
        "comments" : {
            "type": "array",
            "items": {
                "$ref": "comment_schema.json#"
            }
        }
    },
    "additionalProperties": false
}
```

factor_schema.json

This schema implements Study factor from the ISA Abstract Model.

Schema:

```
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title": "ISA factor schema",
    "name": "ISA factor schema",
    "description": "JSON-schema representing a factor value in the ISA model",
    "type": "object",
    "properties": {
        "@id": { "type": "string", "format": "uri" },
        "factorName": {
            "type": "string"
        },
        "factorType": {
            "$ref": "ontology_annotation_schema.json#"
        },
        "comments" : {
            "type": "array",
            "items": {
                "$ref": "comment_schema.json#"
            }
        }
    },
    "additionalProperties": false
}
```

factor_value_schema.json

This schema implements Factor value given to a node corresponding to a declared Factor.

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title": "ISA factor value schema",
    "description": "JSON-schema representing a factor value in the ISA model",
    "type": "object",
    "properties": {
        "@id": { "type": "string", "format": "uri" },
        "category" : {
            "$ref": "factor_schema.json#"
        },
        "value": {
            "anyOf" : [
                { "$ref": "ontology_annotation_schema.json#"},
                { "type": "string"},
                { "type": "number"}
                ]
        },
        "unit": {
            "$ref": "ontology_annotation_schema.json#"
        }
    },
    "additionalProperties": false
}
```

material_attribute_schema.json

This schema is used in a Material node to declare an attribute (Characteristic).

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA material attribute schema",
    "description" : "JSON-schema representing a characteristics category (what
↪appears between the brackets in Charactersitics[]) in the ISA model",
    "type" : "object",
    "properties" : {
        "@id": { "type": "string", "format": "uri" },
        "characteristicType": {
            "$ref": "ontology_annotation_schema.json#"
        }
    },
    "additionalProperties": false
}
```

material_attribute_value_schema.json

This schema is used in a Material node to hold an attribute value (value of a Characteristic).

Schema:

```
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA material attribute schema",
    "description" : "JSON-schema representing a material attribute (or
↪characteristic) value in the ISA model",
    "type" : "object",
    "properties" : {
        "@id": { "type": "string", "format": "uri" },
        "category" : {
            "$ref": "material_attribute_schema.json#"
        },
        "value": {
            "anyOf" : [
                { "$ref": "ontology_annotation_schema.json#"},
                { "type": "string"},
                { "type": "number"}
                ]
        },
        "unit": {
            "$ref": "ontology_annotation_schema.json#"
        }
    },
    "additionalProperties": false
}
```

material_schema.json

This schema implements Material nodes from the ISA Abstract Model.

Schema:

```
{
  "$schema": "http://json-schema.org/draft-04/schema",
  "title" : "ISA material node schema",
  "description" : "JSON-schema representing a material node in the ISA model, which
↪is not a source or a sample (as they have specific schemas) – this will correspond
↪to 'Extract Name', 'Labeled Extract Name'",
  "type" : "object",
  "properties" : {
    "@id": { "type": "string", "format": "uri" },
    "name" : { "type" : "string" },
    "type": {
      "type": "string",
      "enum": [
        "Extract Name",
        "Labeled Extract Name"
      ]
    },
    "characteristics" : {
      "type" : "array",
      "items" :  {
        "$ref": "material_attribute_value_schema.json#"
      }
    },
    "derivesFrom": {
      "type" : "array",
```

```json
      "items" : {
        "$ref": "material_schema.json#"
      }
    }
  },
  "additionalProperties": false
}
```

ontology_annotation_schema.json

This schema implements Ontology from the ISA Abstract Model.

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA ontology reference schema",
    "name" : "ISA ontology reference schema",
    "description" : "JSON-schema representing an ontology reference or annotation in
→the ISA model (for fields that are required to be ontology annotations)",
    "type" : "object",
    "properties" : {
        "@id": { "type": "string", "format": "uri" },
        "annotationValue": {
            "anyOf": [
                { "type": "string" },
                { "type": "number"}
            ]
        },
        "termSource" : {
            "type" : "string",
            "description" : "The abbreviated ontology name. It should correspond to
→one of the sources as specified in the ontologySourceReference section of the
→Investigation."
        },
        "termAccession" : {
            "type" : "string",
            "format" : "uri"
        },
        "comments" : {
            "type": "array",
            "items": {
                "$ref": "comment_schema.json#"
            }
        }
    },
    "additionalProperties": false
}
```

ontology_source_reference_schema.json

This schema implements Ontology from the ISA Abstract Model.

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA ontology source reference schema",
    "name" : "ISA ontology source reference schema",
    "description" : "JSON-schema representing an ontology reference in the ISA model",
    "type" : "object",
    "properties" : {
        "comments" : {
            "type": "array",
            "items": {
                "$ref": "comment_schema.json#"
            }
        },
        "description" : { "type" : "string" },
        "file" : { "type" : "string" },
        "name": {"type": "string"},
        "version": { "type": "string"}
    },
    "additionalProperties": false
}
```

person_schema.json

This schema implements Contact from the ISA Abstract Model.

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA person schema",
    "description" : "JSON-schema representing a person in the ISA model",
    "type" : "object",
    "properties" : {
        "@id": { "type": "string", "format": "uri" },
        "lastName" : { "type" : "string"},
        "firstName" : { "type" : "string"},
        "midInitials" : { "type" : "string" },
        "email" : { "type" : "string", "format" : "email"},
        "phone" : { "type": "string"},
        "fax" : { "type" : "string" },
        "address" : { "type" : "string" },
        "affiliation" : { "type" : "string" },
        "roles" : {
            "type" : "array",
            "items" : {
                "$ref": "ontology_annotation_schema.json#"
            }
        },
        "comments" : {
            "type": "array",
            "items": {
                "$ref": "comment_schema.json#"
            }
        }
    },
    "additionalProperties": false
}
```

process_parameter_value_schema.json

This schema is used in a Process node to hold a parameter value (value of a Protocol parameter).

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA process parameter value schema",
    "description" : "JSON-schema representing a Parameter Value (associated with a
↪Protocol REF) in the ISA model",
    "type" : "object",
    "properties" : {
        "category" : {
            "$ref": "protocol_parameter_schema.json#"
        },
        "value": {
            "anyOf" : [
                { "$ref": "ontology_annotation_schema.json#"},
                { "type": "string"},
                { "type": "number"}
                ]
        },
        "unit": {
            "$ref": "ontology_annotation_schema.json#"
        }
    },
    "additionalProperties": false
}
```

process_schema.json

This schema implements Process nodes from the ISA Abstract Model.

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title": "ISA process or protocol application schema, corresponds to 'Protocol REF
↪' columns in the study and assay files",
    "description": "JSON-schema representing a protocol application in the ISA model",
    "type": "object",
    "properties": {
        "@id": { "type": "string", "format": "uri" },
        "name": {
            "type": "string"
        },
        "executesProtocol": {
            "$ref": "protocol_schema.json#"
        },
        "parameterValues": {
            "type": "array",
            "items": {
                "$ref" : "process_parameter_value_schema.json#"
            }
        },
        "performer": {
```

```json
                "type": "string"
        },
        "date": {
                "type": "string",
                 "format": "date-time"
        },
        "previousProcess" : {
                "$ref" : "process_schema.json#"
        },
        "nextProcess": {
                "$ref" : "process_schema.json#"
        },
        "inputs" : {
                "type": "array",
                "items": {
                        "anyOf": [
                          {
                            "$ref": "source_schema.json#"
                          },
                          {
                            "$ref": "sample_schema.json#"
                          },
                          {
                            "$ref": "data_schema.json#"
                          },
                          {
                            "$ref": "material_schema.json#"
                          }
                        ]
                }
        },
        "outputs" : {
                "type": "array",
                "items": {
                        "anyOf": [
                                {
                                    "$ref": "sample_schema.json#"
                                },
                                {
                                    "$ref": "data_schema.json#"
                                },
                                {
                                "$ref": "material_schema.json#"
                                }
                        ]
                }
        },
        "comments" : {
                "type": "array",
                "items": {
                        "$ref": "comment_schema.json#"
                }
        }
    },
    "additionalProperties": false
}
```

protocol_parameter_schema.json

This schema is used in a Protocol to describe a protocol parameter.

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA protocol parameter schema",
    "description" : "JSON-schema representing a parameter for a protocol (category␣
↪declared in the investigation file) in the ISA model",
    "type" : "object",
    "properties" : {
        "@id": { "type": "string", "format": "uri" },
        "parameterName": {
            "$ref": "ontology_annotation_schema.json#"
        }
    },
    "additionalProperties": false
}
```

protocol_schema.json

This schema implements Protocol from the ISA Abstract Model.

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title": "ISA protocol schema",
    "name": "ISA protocol schema",
    "description": "JSON-schema representing a protocol in the ISA model",
    "type": "object",
    "properties": {
        "@id": { "type": "string", "format": "uri" },
        "comments" : {
            "type": "array",
            "items": {
                "$ref": "comment_schema.json#"
            }
        },
        "name": {
            "type": "string"
        },
        "protocolType": {
            "$ref": "ontology_annotation_schema.json#"
        },
        "description": {
            "type": "string"
        },
        "uri": {
            "type": "string",
            "format": "uri"
        },
        "version": {
            "type": "string"
        },
```

```json
        "parameters": {
            "type": "array",
            "items": {
                "$ref": "protocol_parameter_schema.json#"
            }
        },
        "components": {
            "type": "array",
            "items": {
                "type": "object",
                "properties": {
                    "componentName": {
                        "type": "string"
                    },
                    "componentType": {
                        "$ref": "ontology_annotation_schema.json#"
                    }
                }
            }
        }
    },
    "additionalProperties": false
}
```

publication_schema.json

This schema implements Publication from the ISA Abstract Model.

Schema:

```json
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA investigation schema",
    "name" : "ISA investigation schema",
    "description" : "JSON-schema representing an investigation in the ISA model",
    "type" : "object",
    "properties" : {
        "comments" : {
            "type": "array",
            "items": {
                "$ref": "comment_schema.json#"
            }
        },
        "pubMedID" : { "type" : "string" },
        "doi" : { "type" : "string"},
        "authorList" : { "type" : "string" },
        "title" : { "type" : "string" },
        "status" : {
            "$ref": "ontology_annotation_schema.json#"
        }
    },
    "additionalProperties": false
}
```

sample_schema.json

This schema implements Sample from the ISA Abstract Model.

Schema:

```
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA sample schema",
    "description" : "JSON-schema representing a sample in the ISA model. A sample␣
↪represents a major output resulting from a protocol application other than the␣
↪special case outputs of Extract or a Labeled Extract.",
    "type": "object",
    "properties" : {
        "@id": { "type": "string", "format": "uri" },
        "name" : { "type" : "string" },
        "characteristics" : {
            "type" : "array",
            "items" :  {
                "$ref": "material_attribute_value_schema.json#"
            }
        },
        "factorValues" : {
            "type" : "array",
            "items" : {
                "$ref" : "factor_value_schema.json#"
            }
        },
        "derivesFrom": {
            "type" : "array",
            "items" : {
                "$ref": "source_schema.json#"
            }
        }
    },
    "additionalProperties": false
}
```

source_schema.json

This schema implements Source from the ISA Abstract Model.

Schema:

```
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA source schema",
    "description" : "JSON-schema representing a source in the ISA model. Sources are␣
↪considered as the starting biological material used in a study.",
     "properties" : {
        "@id": { "type": "string", "format": "uri" },
        "name" : { "type" : "string" },
        "characteristics" : {
            "type" : "array",
            "items" :  {
                "$ref": "material_attribute_value_schema.json#"
                }
```

(continues on next page)

```
        }
    },
    "additionalProperties": false
}
```

1.3.3 Content rules

The rules described here define the content and relationship rules that the ISA-JSON objects must adhere to to implement ISA Abstract Model.

1. Files SHOULD be encoded using UTF-8.

2. ISA-JSON content MUST be well-formed JSON.

3. ISA-JSON content MUST validate against the ISA-JSON schemas.

4. ISA-JSON files SHOULD be suffixed with a `.json` extension.

5. Dates SHOULD be supplied in the ISO8601 format "YYYY-MM-DD".

6. DOIs SHOULD conform to the standard format ISO 26324 DOI format "10.NN/xxxNNNNNN".

7. PubMed IDs SHOULD be a string of eight numbers (e.g. 12345678), optionally prefixed with PMC (e.g. PMC12345678).

8. Characteristic Categories declared SHOULD be referenced by at least one Characteristic.

9. Characteristics MUST reference a Characteristic Category declaration.

10. Unit Categories declared SHOULD be referenced by at least one Unit.

11. Units MUST reference a Unit Category declaration.

12. All Sources and Samples MUST be declared in the Study-level materials section.

13. All other materials (Extracts etc.) and DataFiles MUST be declared in the Assay-level material and data sections respectively.

14. Each Process in a Process Sequence MUST link with other Processes forwards or backwards, unless it is a starting or terminating Process (i.e. Beginning or end of the experimental graph).

15. Protocols declared SHOULD be referenced by at least one Protocol REF.

16. Protocol REFs MUST reference a Protocol declaration.

17. Study Factors declared SHOULD be referenced by at least one Factor Value.

18. Factor Values MUST reference a Study Factor declared in the Study-level factors section.

19. Protocols SHOULD have a name (in order to be referenced in ISA-Tab).

20. Protocol Parameters SHOULD have a name (in order to be referenced in ISA-Tab).

21. Study Factors SHOULD have a name (in order to be referenced in ISA-Tab).

22. Sources and Samples declared SHOULD be referenced by at least one Process at the Study-level.

23. Samples, other materials, and DataFiles declared SHOULD be used in at least one Process at the Assay-level.

24. Study and Assay filenames SHOULD be present (in order to be referenced in ISA-Tab).

25. Ontology Source References declared SHOULD be referenced by at least one Ontology Annotation.

26. Ontology Annotations MUST reference a Ontology Source Reference declaration.

27. Ontology Source References MUST contain a Term Source Name.

28. Ontology Annotations with a term and/or accession MUST provide a Term Source REF pointing to a declared Ontology Source Reference.

29. Publication metadata SHOULD match that of publication record in PubMed corresponding to the provided PubMed ID.

30. Comments MUST have a name.

1.4 Software tools

The ISA Model Specification has two Reference Implementations as data formats (*ISA-Tab* and *ISA-JSON*) with supporting software tools. Below is a summary list of tools and supported formats.

- **Active**: In active development and fully supported.

- **Maintenance mode**: No new features are being developed or planned, and only basic support and bug fixes will be supported.

- **Unsupported**: Not in development and no support available.

Software tools supported by the ISA Team

Tool	Description	Format	Development Status	Platform
ISA API	Python API for ISA conversions, validation and content creation	ISA-Tab, ISA-JSON	Active (pre-release)	Python 3+
ISA Explorer	Visualization and search over collections of ISA-Tabs (browser)	ISA-Tab, ISA-JSON	Not released - see preview	Python 3+
linkedISA	Convert ISA-Tab to OWL	ISA-Tab	Active	Java 1.6
OntoMaton	Annotation of ISA-Tab spreadsheets	ISA-Tab	Active	Google Spreadsheets Add-on
rISA	Parse ISA-Tab into R data structures	ISA-Tab	Active	R/Bioconductor
biopy-isatab	Python Parser for ISA-Tab	ISA-Tab	Active	Python 2.7+
ISA creator	Used for creating ISA-Tab files	ISA-Tab	Maintenance mode	Java 1.6
ISA-Tab Viewer	Visualizer for ISA-Tabs (browser)	ISA-Tab	Maintenance mode	JavaScript / HTML / CSS
ISA configurator	Used with ISAcreator to develop ISA-Tab XML Configurations that are used as ISA-Tab templates and used for validating against domain-specific requirements	ISA-Tab	Maintenance mode	Java 1.6
ISA validator	Used with ISA XML Configurations to validate ISA-Tab files against domain-specific requirements	ISA-Tab	Maintenance mode	Java 1.6
ISA converter	Convert ISA-Tab files into other formats	ISA-Tab	Maintenance mode	Java 1.6
BII (Bio Investigation Index)	Web application and DB	ISA-Tab	Maintenance mode	Java 1.6
MAGE to ISA converter	Converter which can pull from ArrayExpress (by an accession number) or read local files and convert them to ISAtab.	ISA-Tab	Unsupported	Java 1.6

To find these tools please visit http://www.isa-tools.org and http://www.github.com/ISA-tools

Other software tools

Tool	Description	For-mat	Maintained by	Plat-form
ISA4J	isa4j: a scalable Java library for creating ISA-Tab meta-data	ISA-Tab	Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, 06466, Germany	Java
ISA to RDF	Convert from ISA-Tab to RDF	ISA-Tab	ToxBank project	Java 1.6
Bio-Parser-ISATab	PERL Parser for ISA-Tab	ISA-Tab	Unknown	PERL

If you are a developer, user, or are aware of other software tools that implement ISA formats that you think should be listed here, please contact the ISA Team.

1.5 Contributing

If you wish to make comments regarding these specifications, please report using the ISA Model and Serialization Specifications issue tracker or send them to isatools@googlegroups.com. All comments are welcome.

1.5.1 License

> **Attention:** The ISA Model and Serialization Specifications are licensed under CC BY-SA 4.0.
>
> Please feel free to **share** and **adapt** the ISA Model and Serialization Specifications but you **must give appropriate attribution**, and you **must apply the same license in any redistribution** of the specifications, even if it incorporates your own contributions.

1.5.2 Contributors

The ISA Model and Serialization Specifications are maintained by Susanna-Assunta Sansone[1], Philippe Rocca-Serra[1], Alejandra Gonzalez-Beltran[1] and David Johnson[1] on behalf of the ISA Community.

The ISA-Tab RC1 specification authored in 2008, on which the ISA Model and Serialization Specifications 1.0 is based on, was initially drafted by Philippe Rocca-Serra, Susanna-Assunta Sansone and Marco Brandizi[2] and subsequently incorporates input from David Hancock[3], Stephen Harris[4], Allyson Lister[1], Michael Miller[5], Kieran O'Neill[6], Chris Taylor[7], Weida Tong[3], and contributors from the wider ISA Community.

Further feedback on ISA was also gratefully received from the ISA Community during the ISA as a FAIR research object workshop, specifically from Scott Edmunds[8], Peter Li[8], Rob Davidson[9], Chris Hunter[8], Nina Jeliazkova[10], Reza

---

[1] Oxford e-Research Centre, University of Oxford, UK.

[2] EMBL-EBI The European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK.

[3] NERC Bioinformatics Center (NEBC), Centre for Ecology and Hydrology, University of Manchester, School of Computer Science, Manchester, UK.

[4] FDA's National Center for Toxicological Research (NCTR), Center for Toxicoinformatics, Jefferson, AR, USA.

[5] Rosetta Biosoftware, Seattle, WA, USA.

[6] Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada.

[7] Earlham Institute, Norwich Research Park, Norwich, UK.

[8] GigaScience Journal, BioMed Central, London, UK.

[9] Office for National Statistics, Newport, Wales, UK

[10] ToxBank Consortium, European Union

Salek[2], Ken Haug[2], Pablo Conesa[2], Rob Davey[7], Ralf Weber[11], Norman Morrison[3], Marco Roos[12], Egon Willighagen[13] and Jildau Bouwman[14].

[11] Environmental Metabolomics Research Laboratory, University of Birmingham, UK
[12] Leiden University Medical Centre, Leiden, Netherlands
[13] Department of Bioinformatics, Maastricht University, Netherlands
[14] Netherlands Organisation for Applied Scientific Research (TNO), The Hague, Netherlands

Revision History

Version	Date	Description
1.0	2016-10-28	Final release of ISA Model and Serialization specifications[2]
1.0	2009-01-13	Final release of ISA-Tab specification[3]
1.0RC1	2008-11-26	First release candidate of ISA-Tab specification[4]

[2] Sansone, Susanna-Assunta, Rocca-Serra, Philippe, Gonzalez-Beltran, Alejandra, Johnson, David & ISA Community. (2016, October 28). ISA Model and Serialization Specifications 1.0. Zenodo. http://doi.org/10.5281/zenodo.163640

[3] Rocca-Serra, Philippe, Sansone, Susanna-Assunta, & Brandizi, Marco. (2009, January 13). Specification documentation: ISA-TAB 1.0. Zenodo. http://doi.org/10.5281/zenodo.161355

[4] Rocca-Serra, Philippe, Sansone, Susanna-Assunta, & Brandizi, Marco. (2008, November 24). Specification documentation: release candidate 1, ISA-TAB 1.0. Zenodo. http://doi.org/10.5281/zenodo.161350