

ICAT Schema Suggestions

Rolf Krahl

ICAT F2F Meeting, RAL, November 2017

- 1 Persistent Identifiers
- 2 Data Publications
- 3 Miscellaneous

- As a general rule: every item that needs to be unambiguously referenced from outside of ICAT should have a persistent identifier.
- At least, there should be a field in the schema to record one.

Already present in the schema:

- Investigation: name, visitId, doi.
(Some kind of standardization on the usage might be helpful though.)
- Dataset, Datafile, DataCollection: doi.
(Don't need to take the attribute name too strict, could also be used for different types of PID other than DOI.)
- User: orcidId.

Missing:

- Instrument.
- ParameterType.
(Relates to the discussion on parameter ontologies.)
- Sample.
(Could use a SampleParameter though.)

- We want to manage the publication workflow for data with ICAT.
- A data publication in general comprises one or more Datasets or Datafiles that may or may not belong to a single Investigation. The content of a publication would best be represented by a DataCollection.
- There is no suitable place in the schema to store the publication metadata related to a DataCollection.

Taking DataCite 4.0 as orientation, the following publication metadata would be needed:

- Doi.
- User (creator and contributor).
- Title.
- Publisher (facility).
- Dates: date of creation and date of publication. (Date of creation may be derived from the related datasets.)
- Resource type (“Dataset”).
- Subject, keyword, classification code, or key phrase.
- Related identifiers. Relevant relation types may include: “IsCitedBy”, “Cites”, “IsSupplementTo”, “IsPartOf”, “IsCompiledBy”, “IsDerivedFrom”, “IsSourceOf”.
- Description, e.g. abstract.
- Funding information.

Suggestion: Add New Class DataPublication

Constraint: facility, doi

Relationships:

Card	Class	Field
1,1	Facility	facility
1,1	DataCollection	content
0,*	DataPublicationUser	dataPublicationUsers
0,*	RelatedIdentifier	relatedIdentifiers
0,*	FundingReference	fundingReferences

Other fields:

Field	Type
doi	String [255] NOT NULL
title	String [255] NOT NULL
publicationDate	Date NOT NULL
subject	String [1023]
description	String [4000]

Suggestion: Add New Class DataPublicationUser

Constraint: publication, user, contributorType

Relationships:

Card	Class	Field
1,1	DataPublication	publication
1,1	User	user

Other fields:

Field	Type
contributorType	String [255] NOT NULL
key	String [255]

(key is only needed to define an order.)

Suggestion: Add New Class RelatedIdentifier

Constraint: publication, identifier, relationType

Relationships:

Card	Class	Field
1,1	DataPublication	publication

Other fields:

Field	Type
identifier	String [255] NOT NULL
relationType	String [255] NOT NULL

Suggestion: Add New Class FundingReference

Constraint: publication, funderName, awardNumber

Relationships:

Card	Class	Field
1,1	DataPublication	publication

Other fields:

Field	Type
funderName	String [255] NOT NULL
funderIdentifier	String [255]
awardNumber	String [255] NOT NULL
awardTitle	String [255]

Note: we might also want to add funding information to Investigation. In this case, the publication relationship would need to be replaced by two many-to-many relationships with Investigation and DataPublication.

- For the purpose of data publication, we would need family name and given name separately and the affiliation for users.
- Suggestion: add optional fields `familyName`, `givenName`, and `affiliation` to `User`.

- HZB needs to register beamtimes for Investigations in ICAT.
- Currently use Shift.
- Drawback: there is no relation to Instrument.
- Suggestion: add an optional relationship with Instrument (“0,1” cardinality) to Shift.