Schema Changes in ICAT 6.2 and 7.0

Hints for a seamless upgrade

Rolf Krahl

ICAT F2F Meeting, 10 February 2026, Abingdon, UK

Schema Changes in ICAT 6.2

- Schema Changes in ICAT 6.2 (released 2025-11-28):
  - Add a new entity type `Subject` to add keywords to data publications #327
  - Add `startDate` and `endDate` fields to `Instrument` #349
  - Add `internalId` attribute to `DataPublication` #355
  - Add `acknowledgement` attribute to `FundingReference` #369
  - Increase size of `location` column for `Datafile` and `Dataset` #364
- Only incremental changes, no compatibility issues.
- Upgrade: recommend to run the SQL upgrade script to keep the size of location columns in the database consistent with the schema.
- Besides that: just start to use the new entity type and new attributes if you need them.

Migrating from DataPublication.subject to Subject

- Assuming you store a semicolon separated list in
  DataPublication.subject, you may migrate to the new
  Subject entity:

```
for pub in client.search("DataPublication"):
    for subj in pub.subject.split(';'):
        subject = client.new("Subject", dataPublication=pub)
        subject.name = subj.strip()
        subject.create()
```

- You still need to update your scripts to use
  DataPublication.subjects in place of
  DataPublication.subject
- You will need appropriate access permissions for class Subject.
- You may now use the new attributes in class Subject.

Schema Changes in ICAT 7.0

- Schema Changes in ICAT 7.0 (planned):
    - Make the relationship between `Sample` and `Investigation` many-to-many #231
    - Add a `pid` property to `SampleType` and change the uniqueness constraint #326
- Incompatible changes. You must run the SQL upgrade script!
- It is recommended to prepare the upgrade with respect to the `Sample` table while the old ICAT server is still running, see subsequent slides.

Changes in SampleType

- Changes to `SampleType` are:
  - drop the relation to `Facility`,
  - add new `pid` and `description` attributes,
  - allow NULL in `molecularFormula`,
  - change the uniqueness constraint to be just `pid`.
- Caveat for the upgrade: the new `pid` attribute must be populated with unique values. The upgrade script cares about that, setting dummy values.
- If you don't care about PIDs for sample types, you can just keep using dummy values and disregard them.

Changes in SampleType (cont.)

- You may need to update your scripts, mostly for creating new
  SampleType:

```
sampletype = client.new("SampleType")
if 'pid' in sampletype.InstAttr:
    # ICAT schema >= 7.0
    sampletype.pid = str(uuid.uuid4())
sampletype.name = "Some SampleType name"
sampletype.molecularFormula = "Some formula"
if 'facility' in sampletype.InstRel:
    # ICAT schema < 7.0
    sampletype.facility = client.assertedSearch("Facility")[0]
sampletype.create()
```

Changes in Sample

- Changes to `Sample` are:
    - drop the many-to-one relation from `Sample` to `Investigation`,
    - add class `InvestigationSample` to establish a many-to-many relation from `Sample` to `Investigation`,
    - change the uniqueness constraint in `Sample` to be just `pid`.
- Caveat for the upgrade: the `pid` attribute must be populated with unique values.
- Additional challenge here: the `pid` attribute did exist in the old schema. Depending on the existing values, there may be obstacles for the upgrade: non-unique values, values that may conflict with the dummy values that the upgrade script tries to set.
- Recommended strategy: prepare the upgrade in the running ICAT, populate the `pid` attribute before trying to upgrade.
- We provide a script helping with that, see next slide.

Preparing the Upgrade

- Helper script `check-samples`, see icatproject-contrib/scripts#1, implementing four subcommands:

  stats Display some statistics about the values in `Sample.pid` and raises warnings if there are any obstacles for the upgrade,

  lsdup List all duplicate `Sample.pid` values,

  setpids populate `Sample.pid`, e.g. set unique dummy values for all `Sample` where pid is NULL,

  dedup deduplicate `Sample.pid` values, e.g. change duplicate values so that they are unique by appending a suffix.

- If `check-samples stats` does not issue a warning, the upgrade via the SQL script should work.

- Still, it is recommended to run `check-samples setpids` if there are any NULL values, just to be safe.

Changes in Sample (cont.)

- You may need to update your scripts, for instance creating a new `Sample` related to a given `Investigation`:

```python
sample = client.new("Sample")
sample.name = "Some name"
sample.pid = str(uuid.uuid4())
if 'investigationSamples' in sample.InstMRel:
    # ICAT schema >= 7.0
    inv_samp = client.new("InvestigationSample",
                          investigation=investigation)
    sample.investigationSamples.append(inv_samp)
else:
    # ICAT schema < 7.0
    sample.investigation = investigation
sample.type = sampletype
sample.create()
```

- You will need appropriate access permissions for class `InvestigationSample`.