iGest

Elettra Sincrotrone Trieste

- Sample data generation
- Metadata ingestion
- Parallel cross-search

+ Experiments

Release: SV7, PaNdata Review

Ingestion

Elettra Sincrotrone Trieste

- Custom
- Hard-coded order
  - i.e. “where do I put the chemical formula?”

[image]

Heuristics? Querying for structure.

iGest Recursion

Elettra Sincrotrone Trieste

WSDL -> ['datafile']

1. '**`None`**' i.e. `datafile.name=sample.h5`
2. '**`list`**' i.e. `investigations`
3. '**`Instance`**' i.e. `facility`

recursive function (condition *isInstance*)

Reveals a structure

Acyclic Directed (kinda) – Beyond ER diagram

iGest Recursion

Elettra Sincrotrone Trieste

[python image]

iGest Recursion

Elettra Sincrotrone Trieste

Beyond visualisation

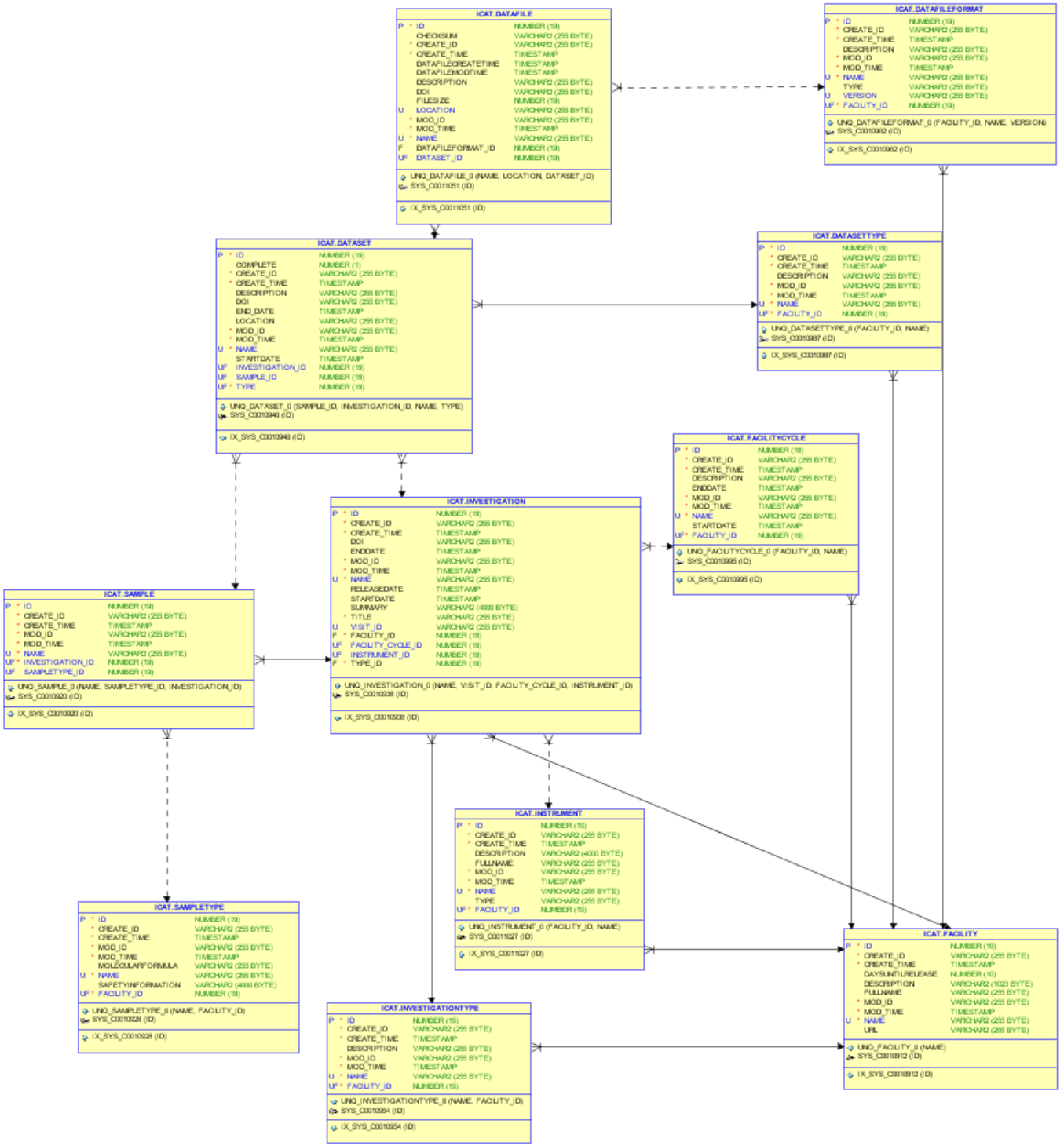
- `ingestIcatElement('datafile')`

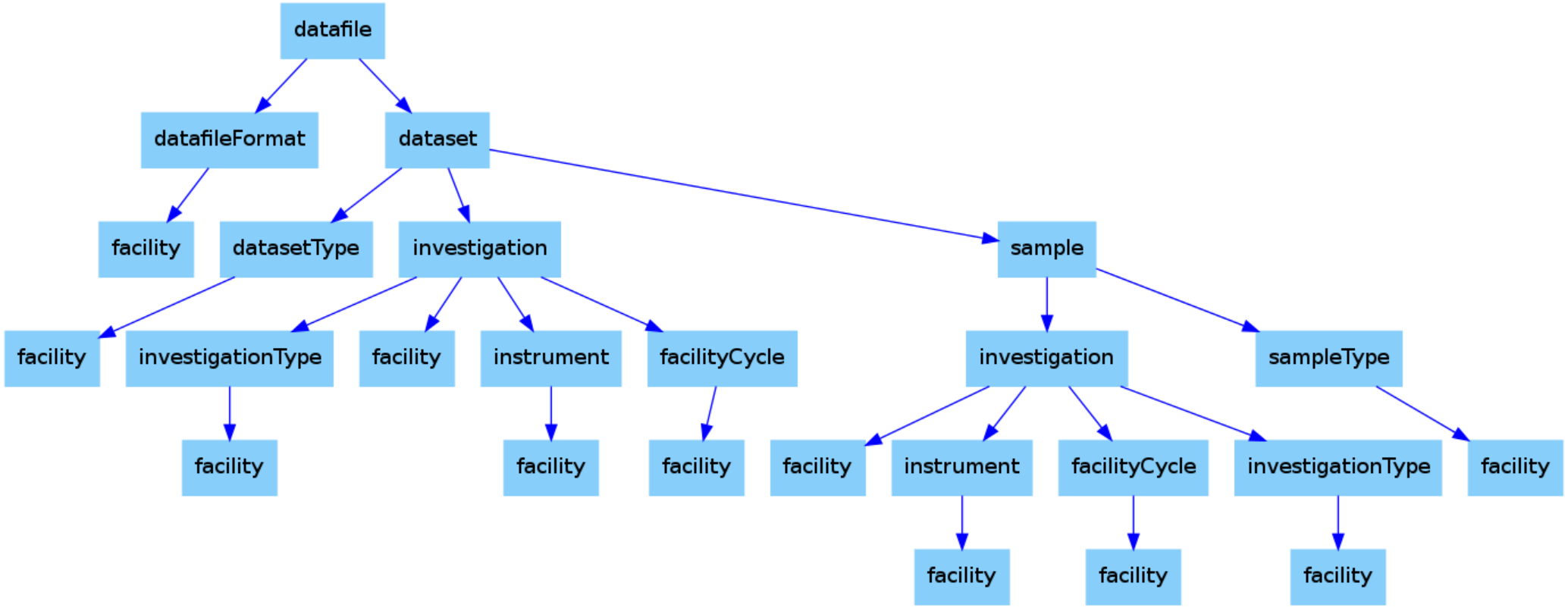
instead of

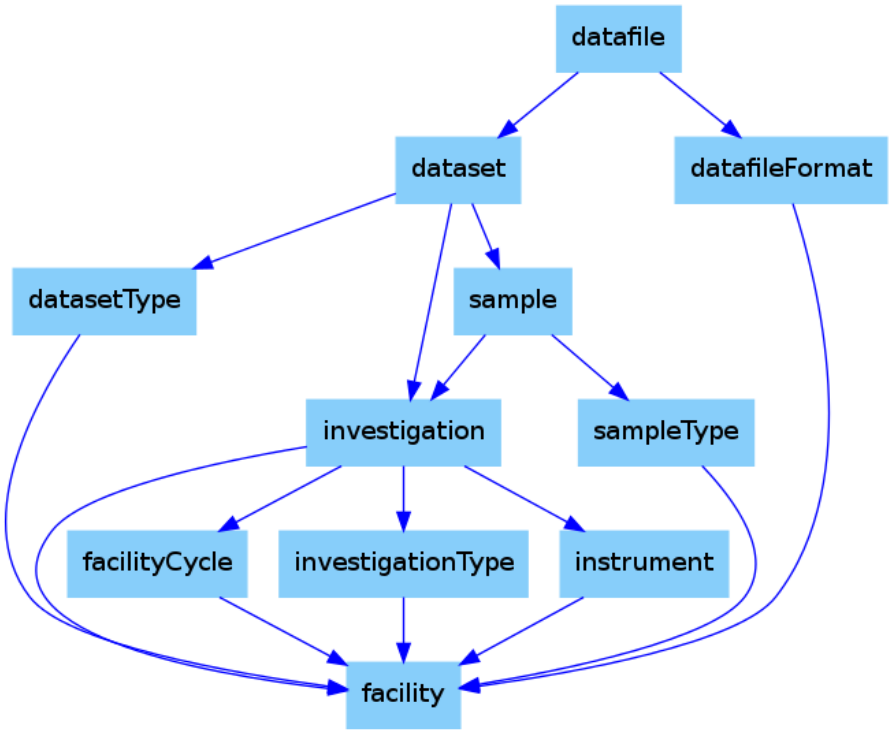
`ingestFacility(),...,igestDatafile()`

- Enhanced search results
- Identify optional elements
- Cross-version ICAT backup

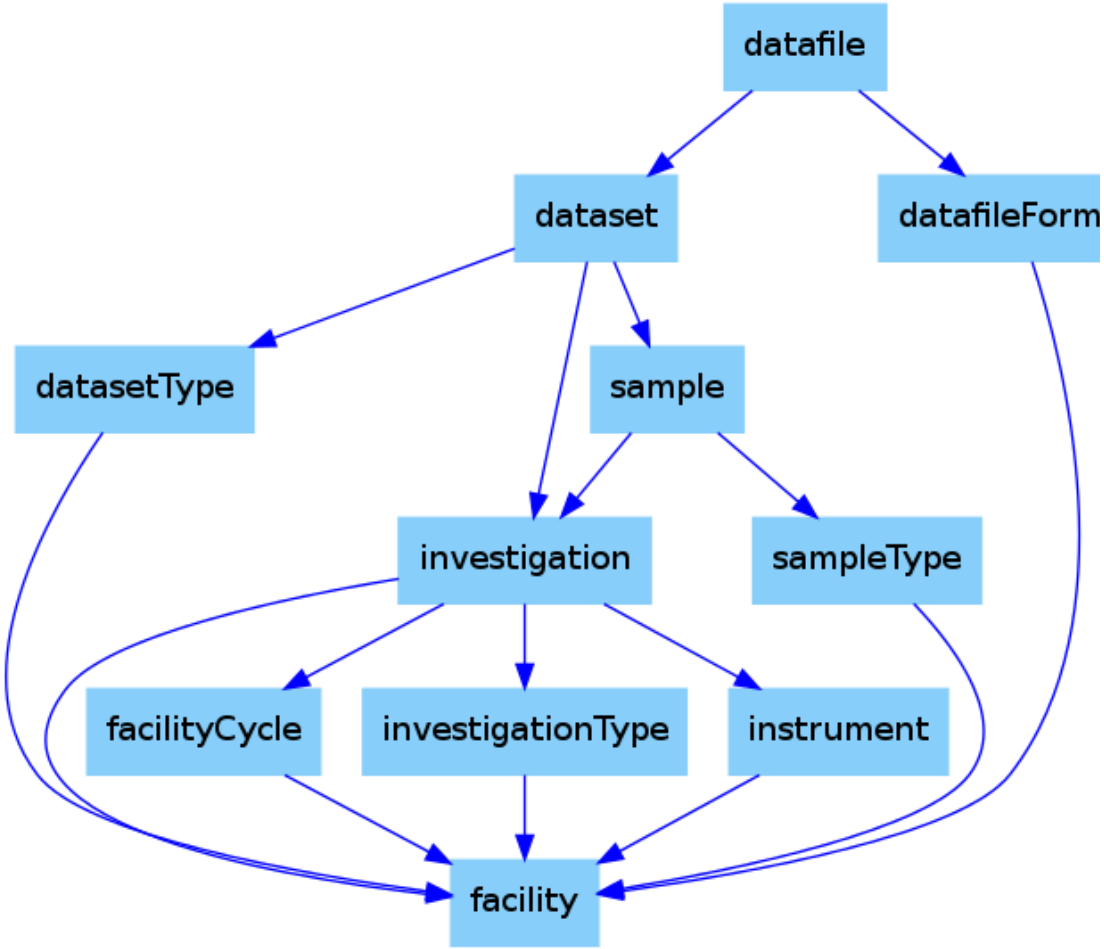
Elettra Sincrotrone Trieste

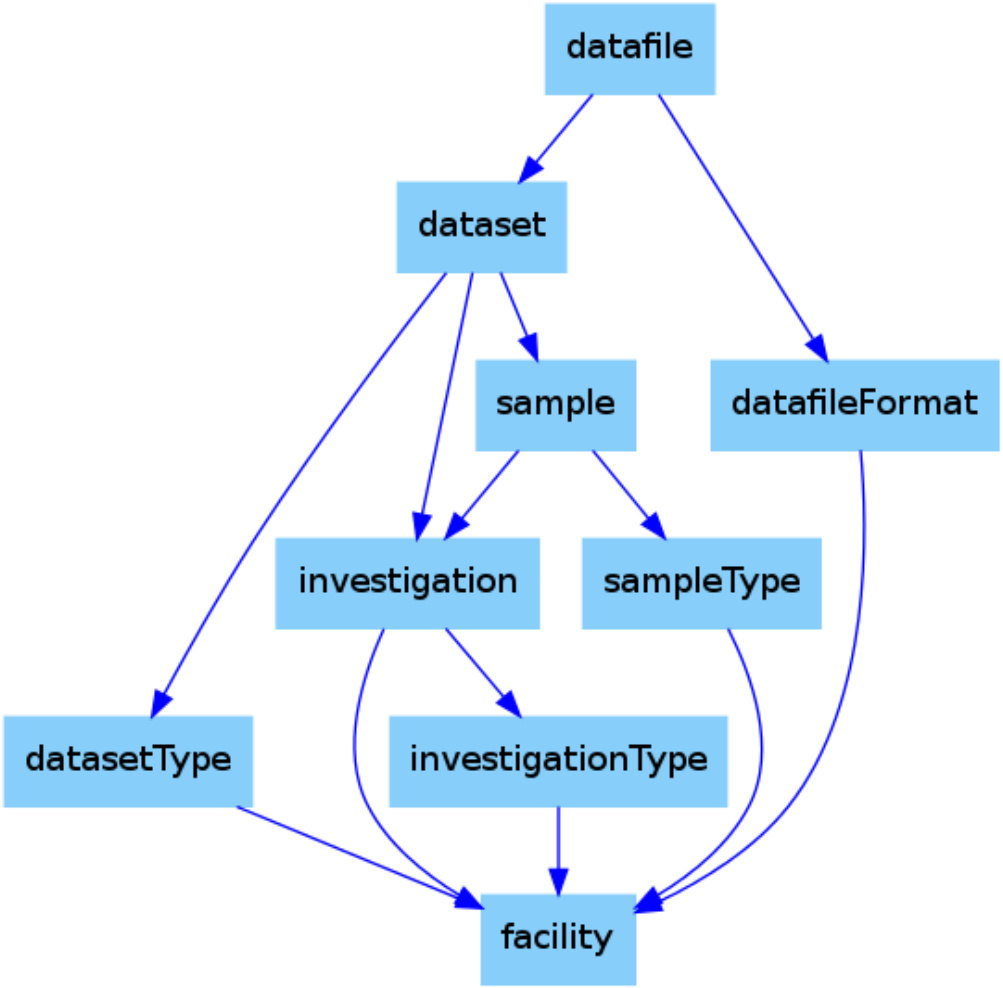


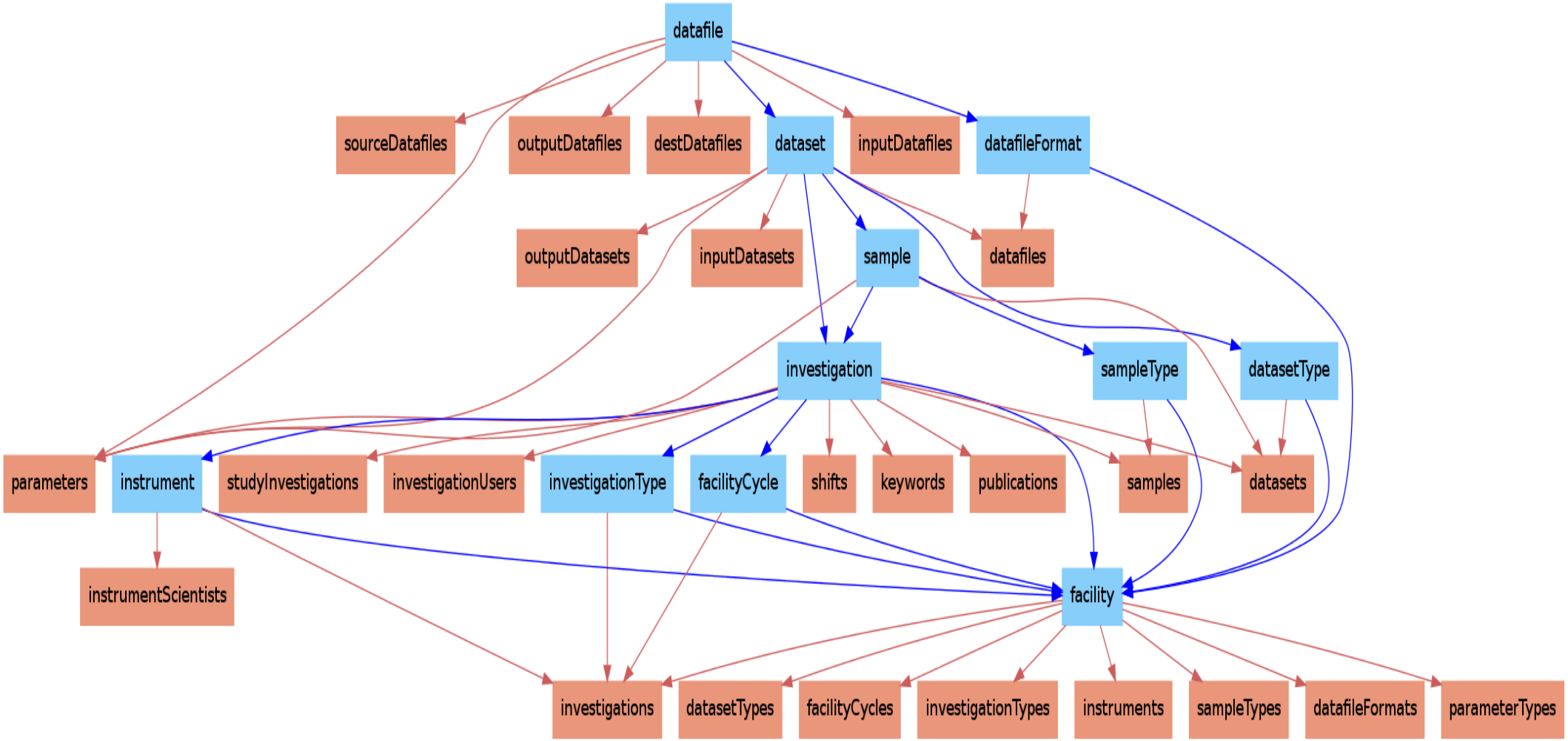

datafile
datafileFormat
dataset
facility
datasetType
investigation
sample
facility
investigationType
facility
instrument
facilityCycle
investigation
sampleType
facility
facility
facility
facility
instrument
facilityCycle
investigationType
facility
facility
facility
facility

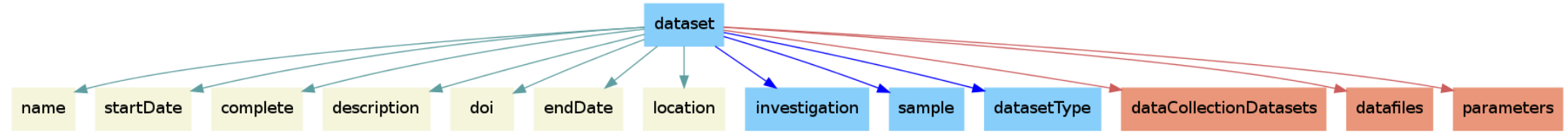

datafile
dataset
datafileFormat
datasetType
sample
investigation
sampleType
facilityCycle
investigationType
instrument
facility

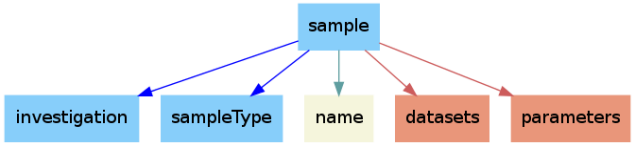
Elettra Sincrotrone Trieste

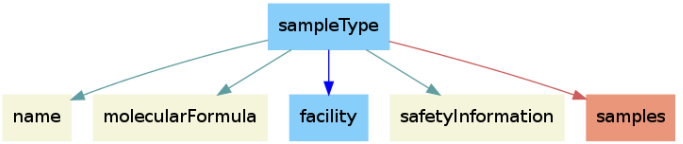


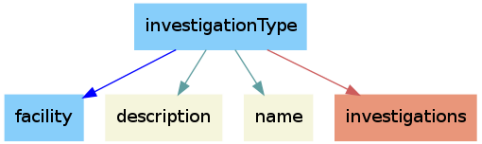


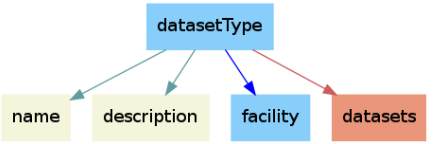

datafile
sourceDatafiles
outputDatafiles
destDatafiles
dataset
inputDatafiles
datafileFormat
outputDatasets
inputDatasets
sample
datafiles
investigation
sampleType
datasetType
parameters
instrument
studyInvestigations
investigationUsers
investigationType
facilityCycle
shifts
keywords
publications
samples
datasets
instrumentScientists
facility
investigations
datasetTypes
facilityCycles
investigationTypes
instruments
sampleTypes
datafileFormats
parameterTypes

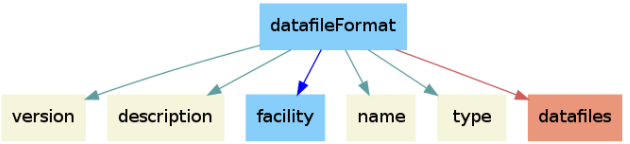

dataset
name
startDate
complete
description
doi
endDate
location
investigation
sample
datasetType
dataCollectionDatasets
datafiles
parameters

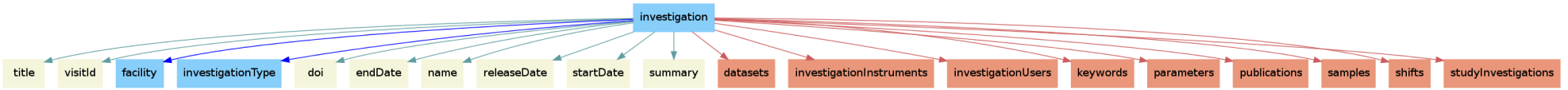

sample
investigation
sampleType
name
datasets
parameters

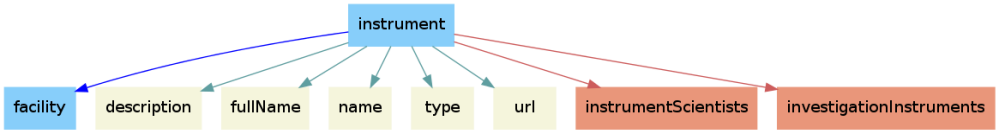

sampleType
name
molecularFormula
facility
safetyInformation
samples

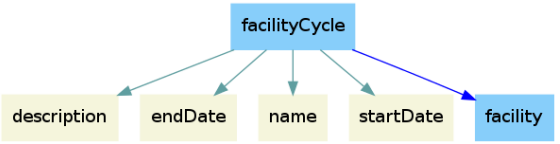

investigationType
facility
description
name
investigations

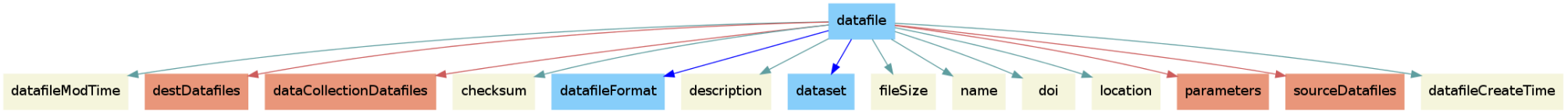

datasetType
name
description
facility
datasets

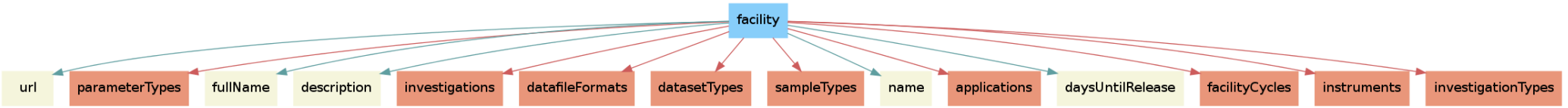

datafileFormat
version
description
facility
name
type
datafiles


investigation
title
visitId
facility
investigationType
doi
endDate
name
releaseDate
startDate
summary
datasets
investigationInstruments
investigationUsers
keywords
parameters
publications
samples
shifts
studyInvestigations


instrument
facility
description
fullName
name
type
url
instrumentScientists
investigationInstruments


facilityCycle
description
endDate
name
startDate
facility


datafile
datafileModTime
destDatafiles
dataCollectionDatafiles
checksum
datafileFormat
description
dataset
fileSize
name
doi
location
parameters
sourceDatafiles
datafileCreateTime


facility
url
parameterTypes
fullName
description
investigations
datafileFormats
datasetTypes
sampleTypes
name
applications
daysUntilRelease
facilityCycles
instruments
investigationTypes

Ingestion (again...)

Elettra Sincrotrone Trieste

- Check **what** you care to catalogue
- Check if ICAT **supports it**
- Find out the ICAT prerequisites
  - Check if they are present in your data

then python-icat, iGest, suds, xmlingest, pyIcat, etc