NFDI4Cat

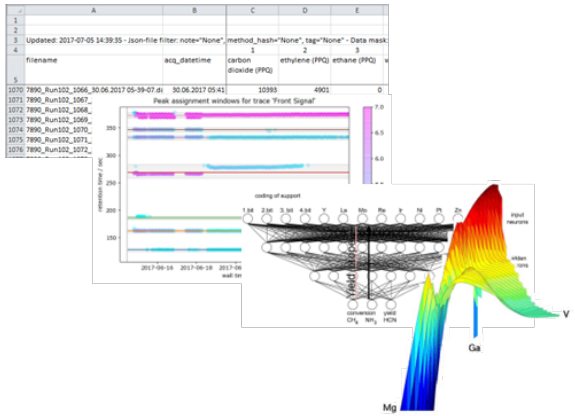
NFDI for Catalysis-Related Sciences

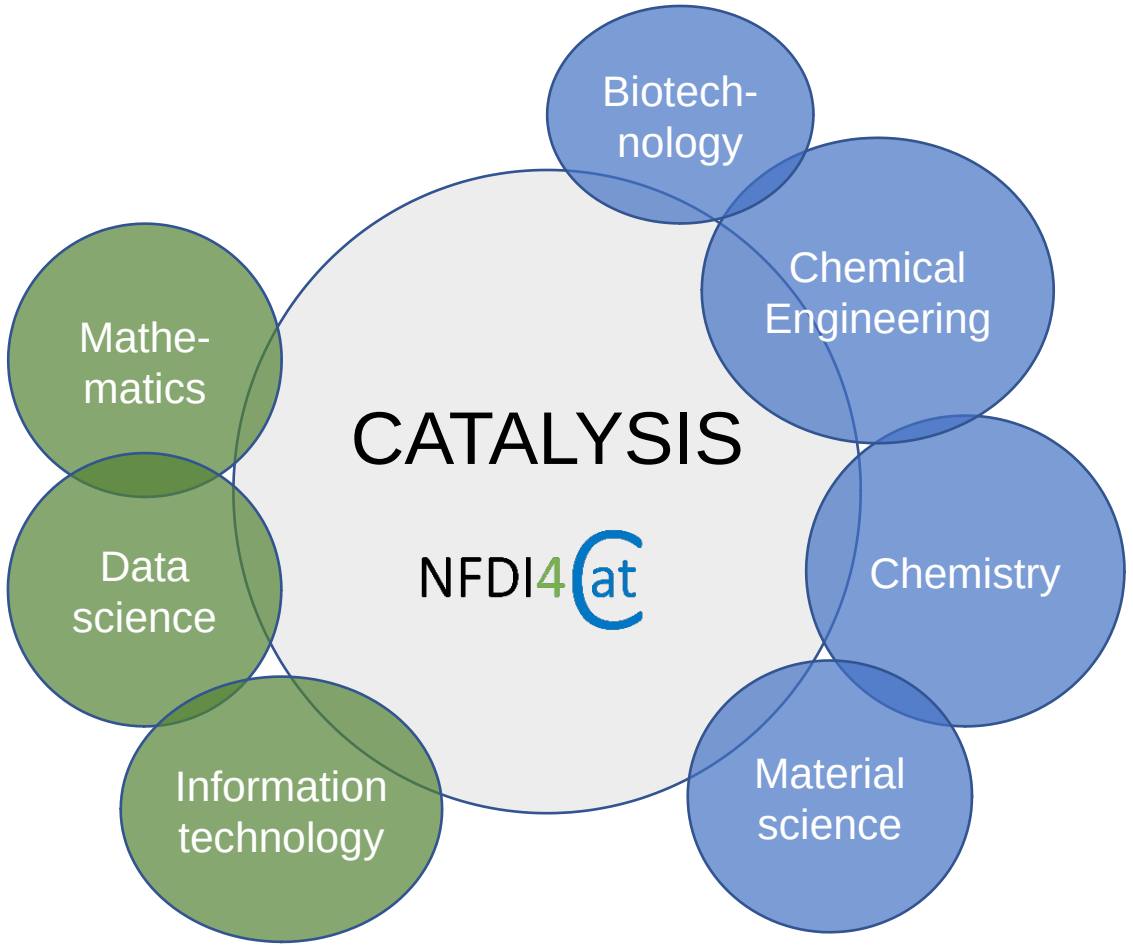
Some Important Aspects of the NFDI4Cat Ontology and Metadata Design

Taras Petrenko

NFDI4Ing 26.10.2022

NFDI for Catalysis-Related Sciences





Project structure and management

TA1: Ontology Development and Metadata Standards

TA2: Data standards, Data Collection, Interfaces

TA3: Data Analysis and Re-Use, Quality Management

TA4: Linked Extensible Infrastructure and Access Management

TA5: Dissemination and Outreach/ Training

TA6: Networking with NFDIs, SFBs and International

TA7: IP and Confidentiality, Licences and Reward models

TA8: Management

Data & Metadata Standards

Data Science & Information Infrastructure Design

Community & User-related Aspects

Linked Data Principles

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful RDF information
4. Include RDF statements that link to other URIs so that they can discover related things

Tim Berners-Lee 2007
http://www.w3.org/DesignIssues/LinkedData.html

Linked Data Principles

Within NFDI URIs will be user for:

- Data
- Chemical substances
- Samples
- Devices
- Processes
- Domain-specific relations
- ....

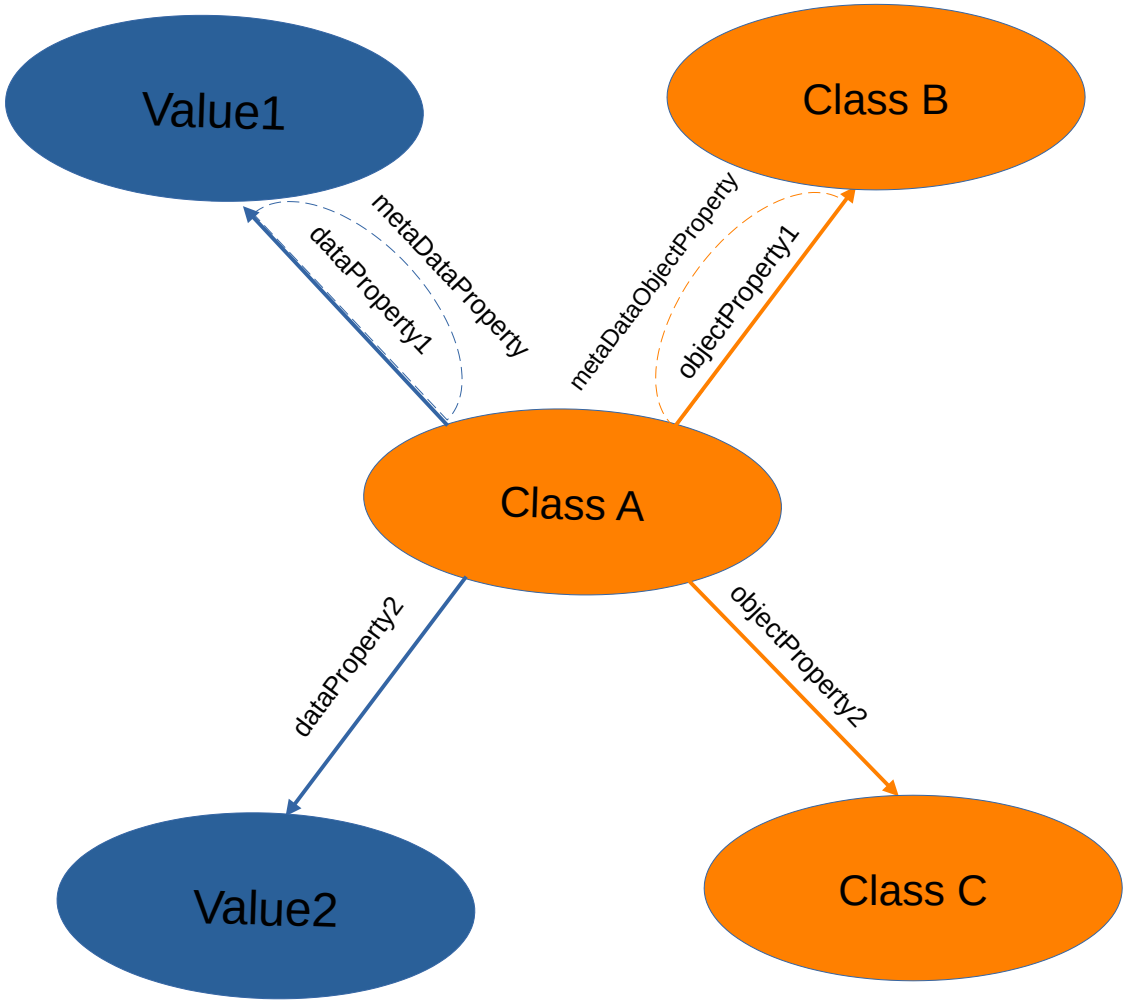
Challenges

- Overall structure of metadata
- Structure and assignment of URIs
- Representation of complex and diverse workflows
- Tedious collection of metadata
- User-invarience of collected metadata
- Designing subsumption relations in ontology
- Determining whether to design an item as a class or property
- Designing complex relations

Ideal NFDI4Cat Ontology

- metadata-oriented
- suitable for a clear representation of the research workflow and all resources involved
- the classes and relations include all relevant semantic search terms
- suitable for semantic query according to workflow patterns
- uniquely shape an interactive metadata inquiry by appropriate software
- allows the metadata inquiry and collection software to be only loosely coupled with the context of metadata

Metadata Model

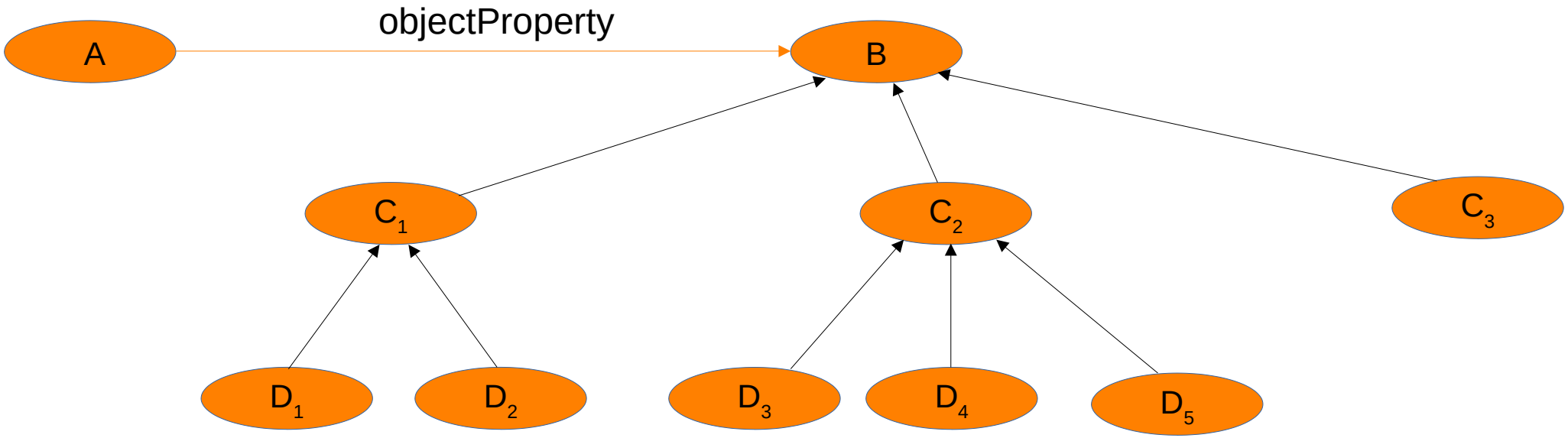


The metadata related to an instance of class A are given by its properties. This would imply that in the case of an object property relating classes A and B, there is also a need for providing the metadata for the instance of class B, and so forth.

- Inheritance
- Modularity: all expansions are independent

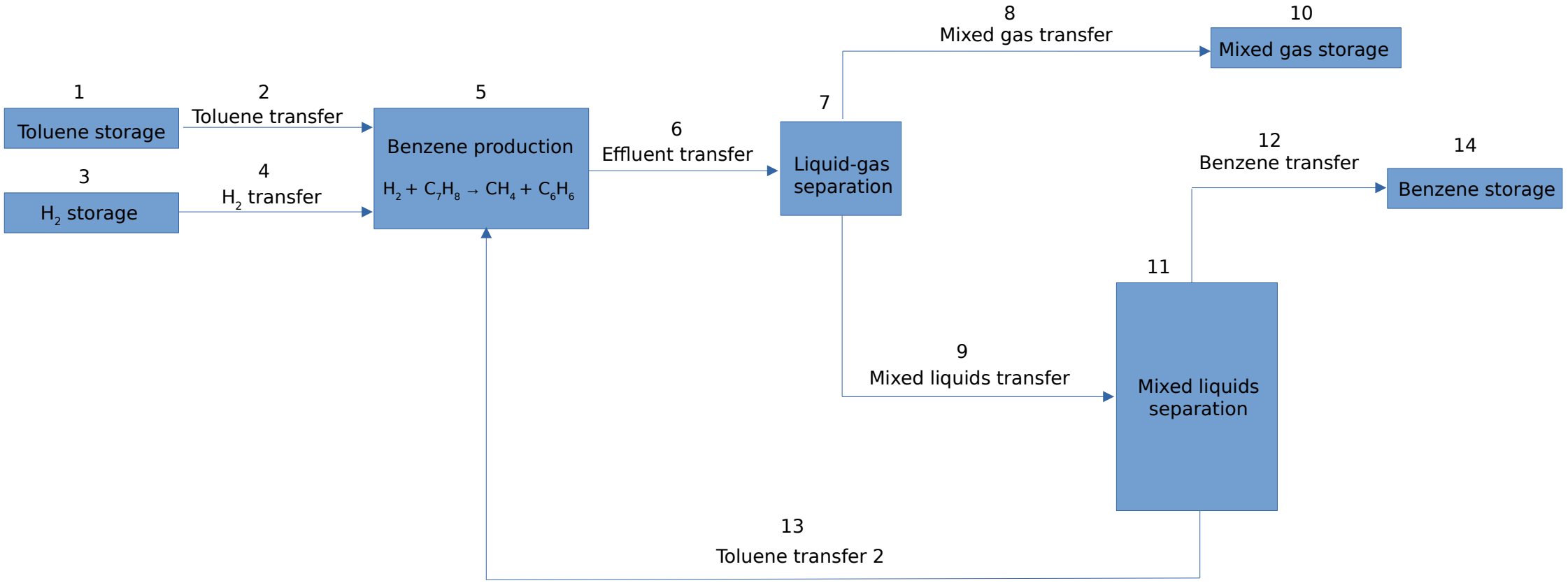
Polymorphism

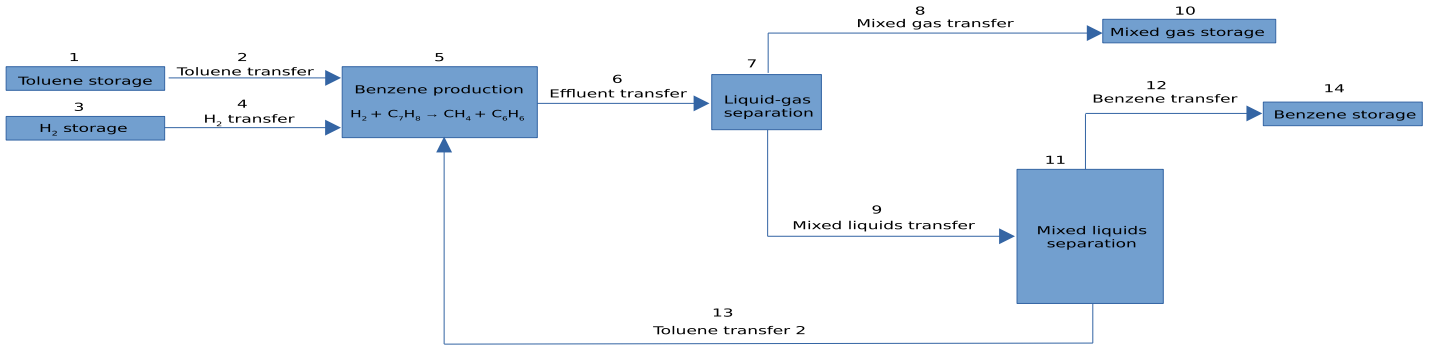
A general pathway for a semiautomatic construction of metadata



The range of properties representing the metadata within a domain-specific model should be represented using the most general class, while only the most specific classes should be used to instantiate metadata.

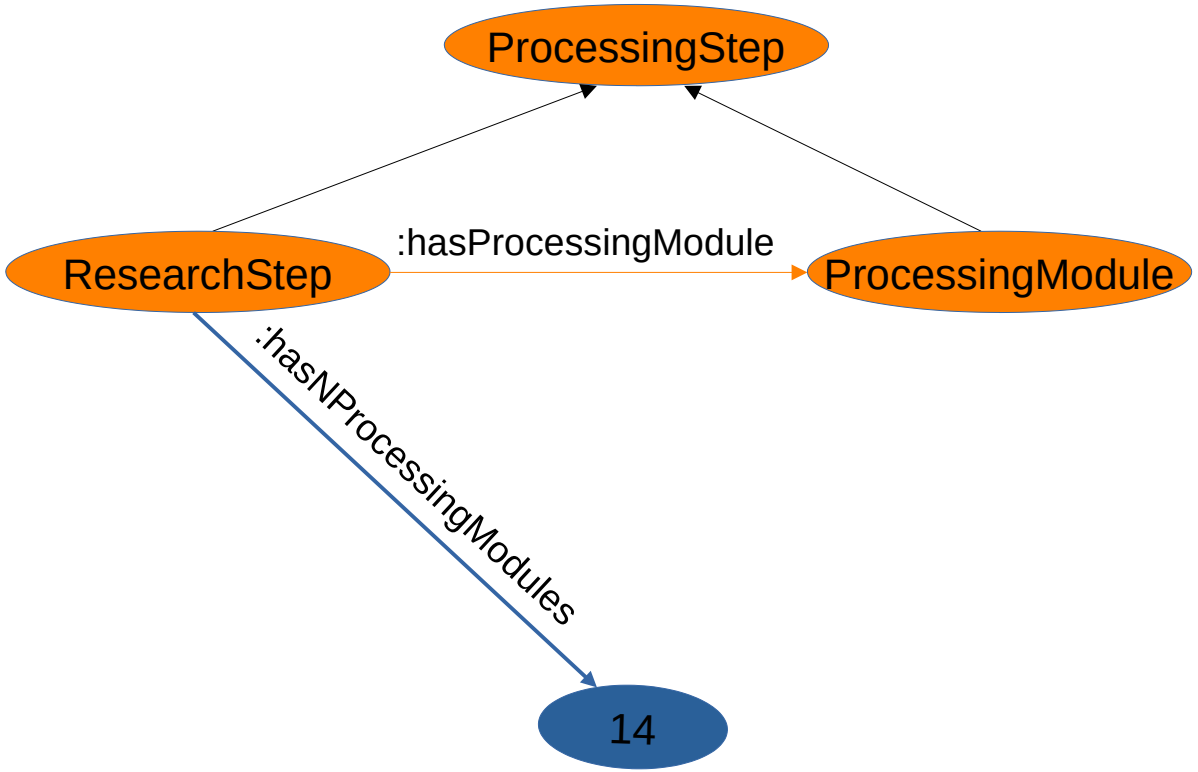
Toluene Hydrodealkylation Process



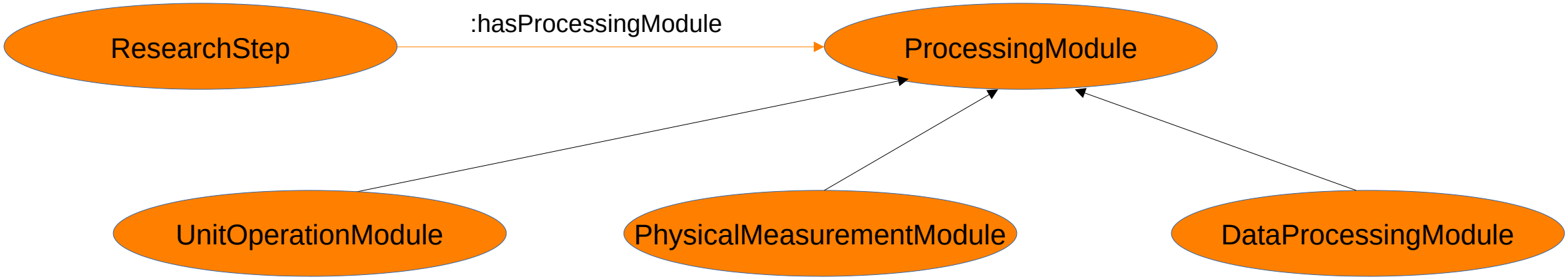


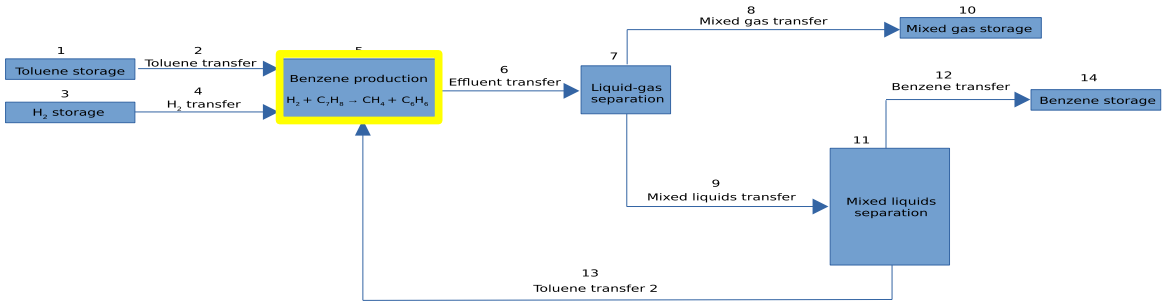

```
#
# Metadata prototype for toluene hydrodealkylation process
#
@prefix ms:      <http://www.nfdi4cat.org/metaset_000001#> .
@prefix sm:      <http://www.nfdi4cat.org/substance_manufacturers#> .
@prefix dev:     <http://www.nfdi4cat.org/tud/devices#> .
@prefix o4cat:   <http://www.nfdi.org/nfdi4cat/ontochem#> .
@prefix rdfs:    <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl:     <http://www.w3.org/2002/07/owl#> .

# Specification of the current research step
ms:ResearchStep_1       a                               o4cat:ResearchStep.
ms:ResearchStep_1       o4cat:hasNProcessingModules     14 .
ms:ResearchStep_1       o4cat:hasProcessingModule       ms:ProcessingModule_1 ,
                                                        ms:ProcessingModule_2 ,
                                                        ms:ProcessingModule_3 ,
                                                        ……..
                                                        ms:ProcessingModule_14 .
```

Metadata4Ing



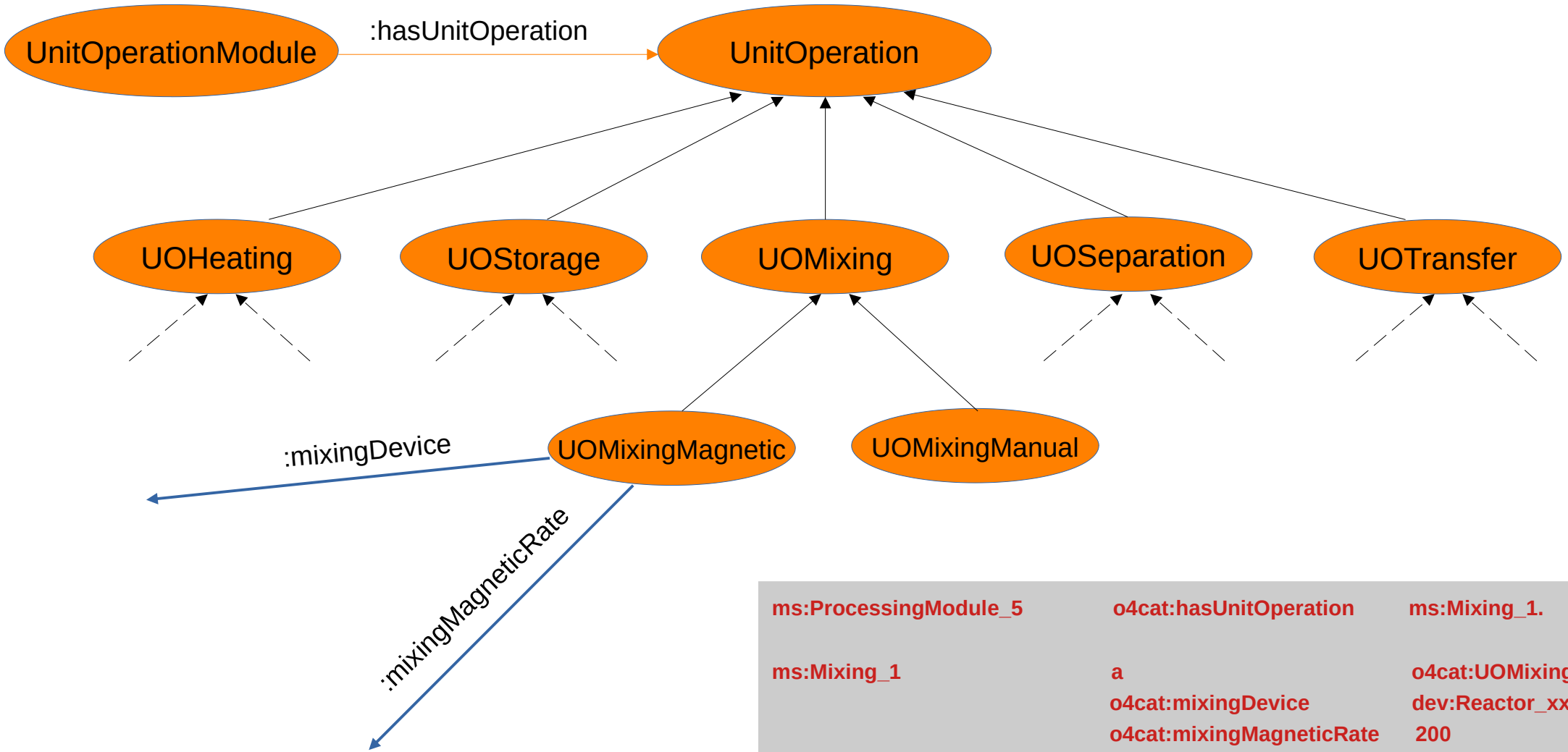
Scheme for Metadata Inquiry






```
ms:ResearchStep_1       a o4cat:ResearchStep.
ms:ResearchStep_1       o4cat:hasNProcessingModules  14 .
……..
ms:ResearchStep_1       o4cat:hasProcessingModule       ms:ProcessingModule_5 .
ms:ProcessingModule_5                                   a       o4cat:UnitOperationModule
```

Scheme for Metadata Inquiry



ms:ProcessingModule_5	o4cat:hasUnitOperation	ms:Mixing_1.
ms:Mixing_1	a	o4cat:UOMixingMagnetic
	o4cat:mixingDevice	dev:Reactor_xxx
	o4cat:mixingMagneticRate	200

Metadata

```
……...
# Processing module #5
ms:ProcessingModule_5      a                          o4cat:PhysChemProcessingModule;
                           o4cat:hasProcessingTask    o4cat:CHEMICAL_REACTION;
                           o4cat:hasChemicalReaction  ms:Reaction_1;
                           o4cat:hasProcessingDevice  dev:Reactor_xxx;
                           o4cat:hasUnitOperation     ms:Mixing_1;
                           o4cat:hasUnitOperation     ms:Heating_1;
                           o4cat:referenceLabel       "Benzene production" ;
                           rdfs:label                 "C7H8 + H2 -> C6H6 + CH4 conversion" ;
                           o4cat:hasChemicalMaterialInput   ms:Substance_1, ms:Substance_2;
                           o4cat:hasChemicalMaterialOutput  ms:Mixture_1.
ms:Reaction_1              a                          o4cat:ChemicalReaction;
                           o4cat:chemicalReactionReagent    ms:Substance_1,  ms:Substance_2;
                           o4cat:chemicalReactionProduct    ms:Substance_3,  ms:Substance_4;
                           o4cat:hasReactionMixture   ms:Mixture_1;
                           o4cat:isReactionProved     1 .
ms:Mixture_1               a                          o4cat:ChemicalMixture;
                           o4cat:hasChemicalComponent ms:Substance_1,  ms:Substance_2, ms:Substance_3,  ms:Substance_4 .
ms:Mixing_1                a                          o4cat:UOMixingMagnetic;
                           o4cat:mixingDevice         dev:Reactor_xxx ;
                           o4cat:mixingMagneticRate   200.
ms:Heating_1               a                          o4cat:UOHeating;
                           o4cat:heatingMethod        o4cat:HEATING_THERMOBATH;
                           o4cat:heatingTemperature   300.  # in kelvin units
ms:Substance_3             a                          o4cat:ChemicalCompound;
                           o4cat:referenceLabel       "Benzene";
                           o4cat:inChIKey             "YXFVVABEGXRONW-UHFFFAOYSA-N". #  IUPAC Standard InChIKey
……...
```

Disclaimer.
No human brain is harmed by collecting RDF metadata. All tedious manipulations are performed programmatically in a *semi-automatic fashion* by applying an informal metadata inquiry shaped by the appropriate ontology.

Metasets

- Metaset is a collection of the RDF graph statements representing the metadata for a given (publishable) reseach step
- Each metaset has its own URI
- It can be serialized into a file(s) or triple store
- Principal composition of the URI of a metaset:

http://www.institute*XXX*/group*YYY*/metaset*NNN*

URIs

- URI of a metaset:

  http://www.institute*XXX*/group*YYY*/metaset*NNN*

- URI of a resource first provided in the associated study :

  http://www.instituteXXX/groupYYY/metaset*NNN*#ClassName_M = ms:ClassName_M

- Straightforward automatic URI scheme for a detailed documentation of a resource's provenance
- Efficient indexing
- Straightforward navigation between metasets by referring e.g. a resource $\mathbf{ms_1}$**:name** in metaset $\mathbf{ms_2}$

THANK YOU!