eNanoMapper database, search tools and templates

Nina Jeliazkova, Nikolay Kochev

IdeaConsult Ltd.

Sofia, Bulgaria

www.ideaconsult.net

- eNanoMapper database: data model, technology;
  NANoREG data transfer examples
- Search tools: free text, chemistry, semantic; API access
- I/O support: ISA & Excel templates

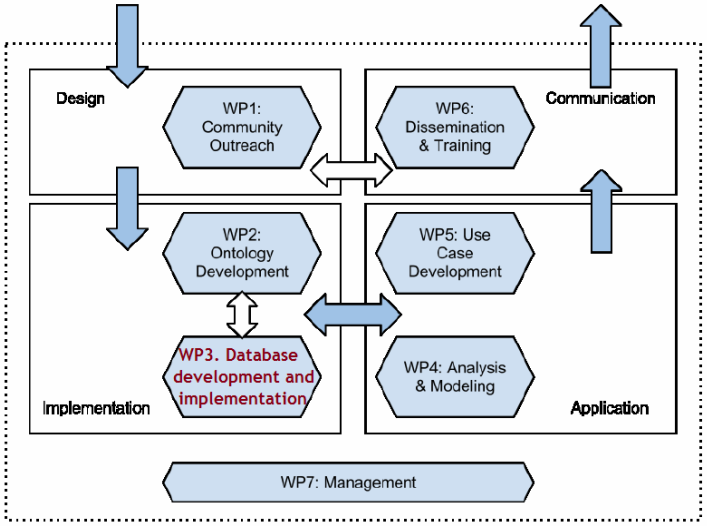
CONTENT

eNanoMapper summary

FP7 eNanoMapper - A Database and Ontology Framework for Nanomaterials Design and Safety Assessment

Grant Agreement: 604134,

1 Feb 2014 – 31 Jan 2017



Solutions available

Open source database and web application

- **Builds upon a Chemical structure database with support for substances.**
- **The data model supporting experimental data is capable of representing all endpoints of regulatory interests and other types of data.**

eNanoMapper ontology; developed by an experienced team at EBI. Existing ontologies are reused;

Tools to process and import data. Export in various formats

Searchable; Free text search based on ontology

Integration of data analysis tools via API

Flexible data hosting architecture

Organising the nanosafety data

- Challenges
  - Diverse data sources
  - Diverse data input formats
  - Different data organization
  - Diverse modelling tools
- Approach:
  - Enable mappings!
  - i.e. eNanoMapper

- **Physico-chemical identity**

Different analytic techniques, manufacturing conditions, batch effects, mixtures, impurities, size distribution, differences in the amount of surface modification, etc.

- **Biological identity**

Wide variety of measurements, toxicity pathways, effects of ENM coronas, modes-of-action, interactions (cell lines, assays).

- **Processes requiring information**

From raw data (science) to study summaries for regulatory purposes; linking with experimental protocols; risk assessment; grouping, safety-by-design

- **Support for data analysis**

Requires "spreadsheet" or matrix view of data. The experimental data in the public datasets is usually not in a form appropriate for modelling (merging multiple values, conditions, similar experiments into matrix form is a challenge).

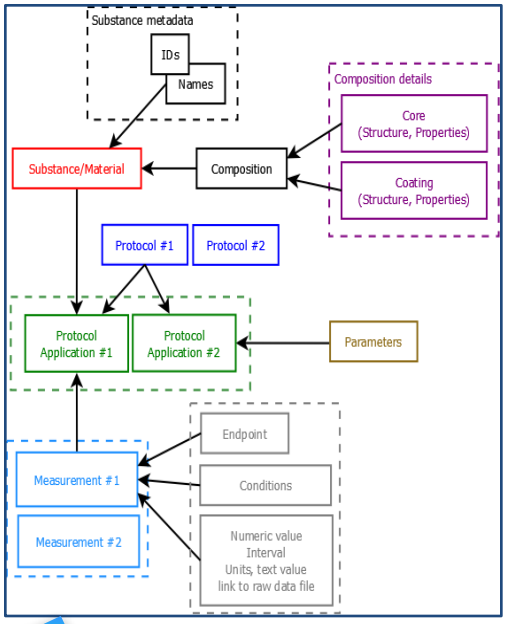
The common eNanoMapper data model : enables conversions

NSC Excel spreadsheets

SQL DB

OECD HT

RDF

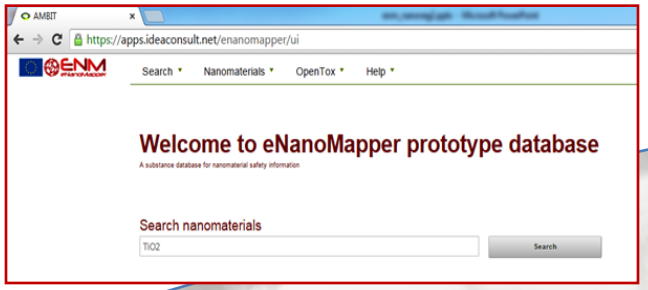


ISA-JSON → ISA-TAB

RDF → W3C SPARQL

JSON

data.enanomapper.net



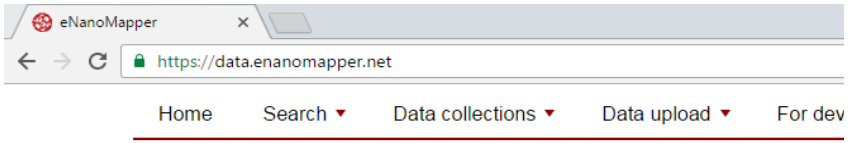
Yet another search service

WP5: Use Case Development

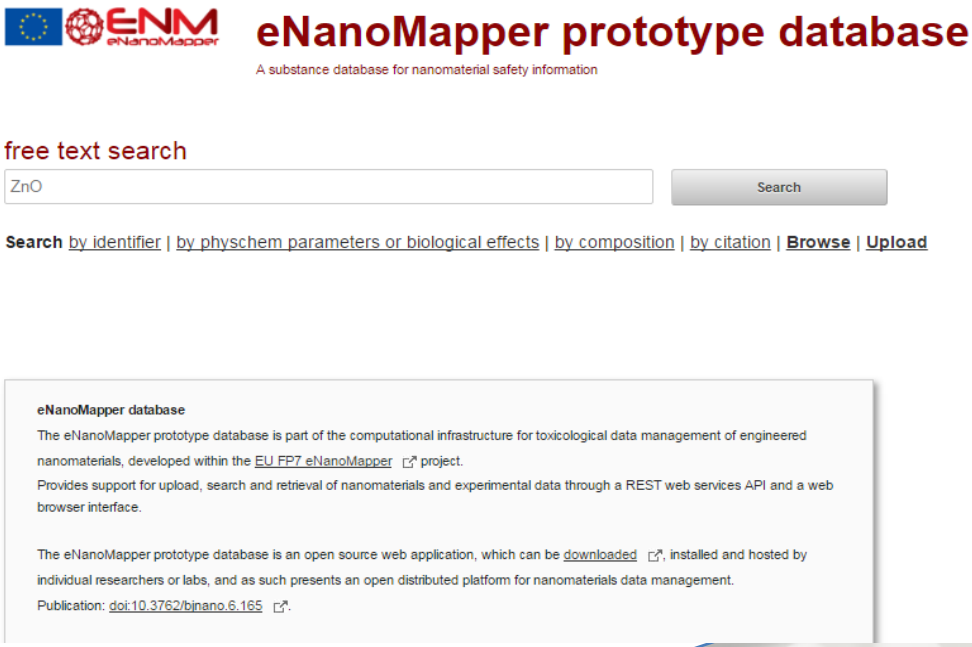
WP4: Analysis & Modeling

eNM eNanoMapper

data.enanomapper.net



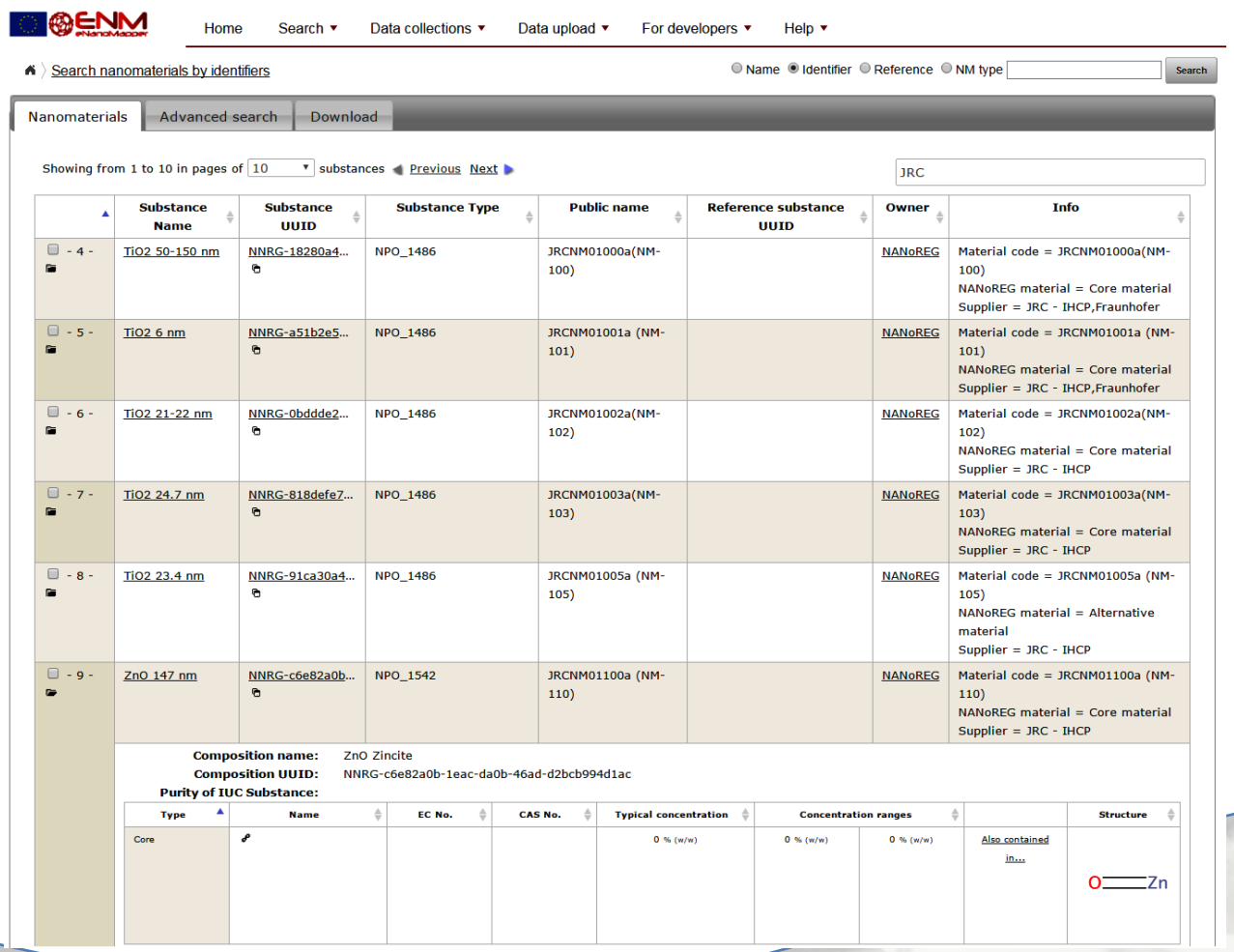
N. Jeliazkova, et al. **"The eNanoMapper database for nanomaterial safety information,"** *Beilstein J. Nanotechnol., vol. 6, pp. 1609–1634, Jul.* ***2015****.*



Implementation

- The database structure has two major concepts:
  - Substances, substance compositions, chemical structures
  - Experimental results (P-CHEM, ECOTOX, TOX, ENV-FATE)
- A generic description of any measurement. Does not specify what to record to describe particular experiment.
  - This information comes from NANoREG templates, IUCLID5 files, etc.
- The database software is based on an open source project http://ambit.sf.net
  - developed by eNanoMapper partner Ideaconsult since 2005, most recently : CEFIC LRI AMBIT tool for read across.
- The data model is capable of representing all endpoints of regulatory interests and other types of data.

NANoREG data transfer (ongoing)



http://www.nanoreg.eu

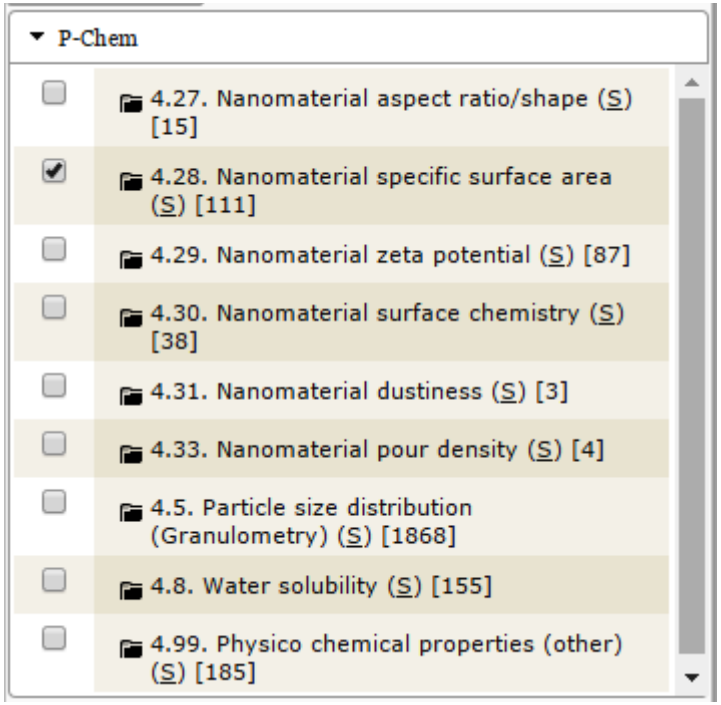
Installed:

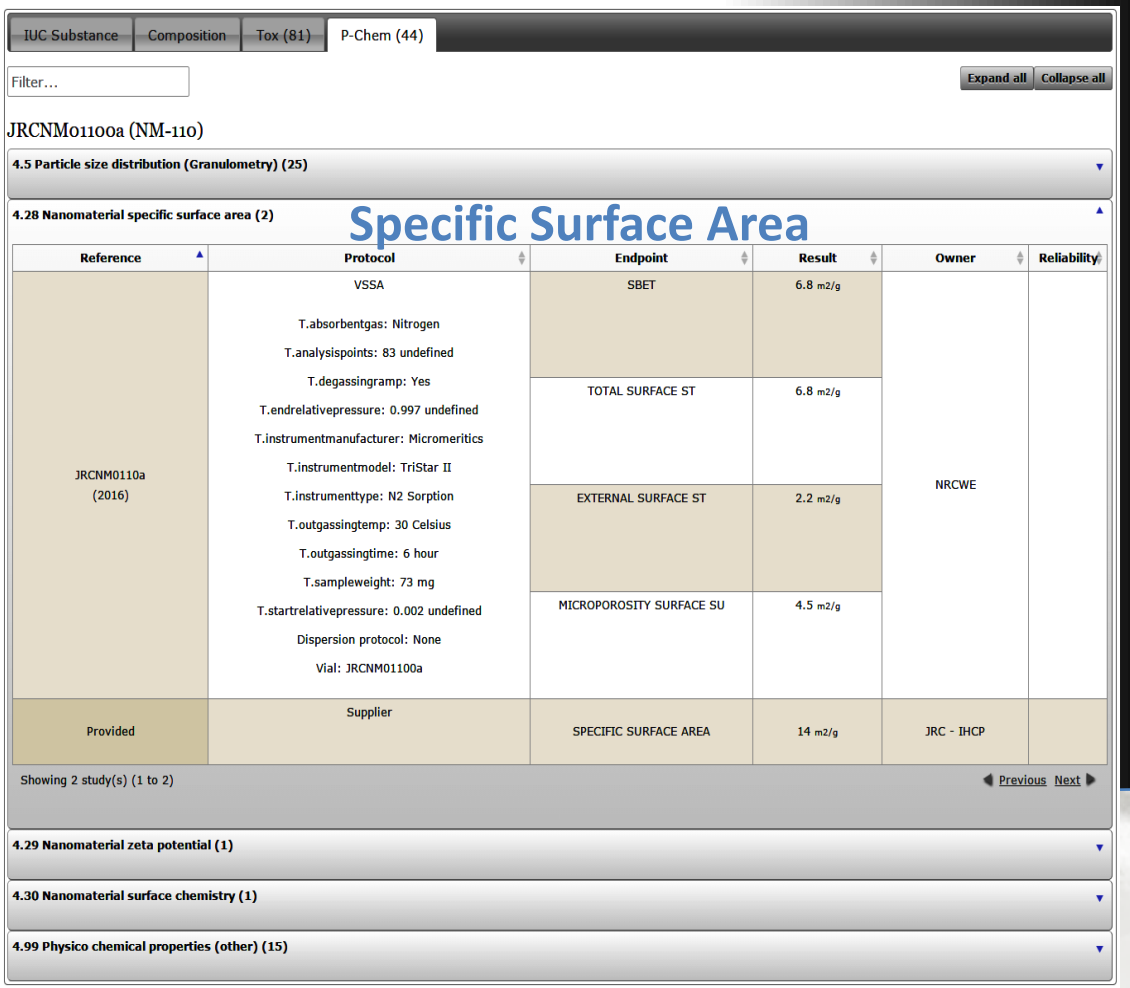
- A separate instance of the database
- Search application

Data sources:

- TNO SQL database *(converted into eNM SQL)*
- Excel files *(mapping ongoing)*

NANoREG example: phys-chem



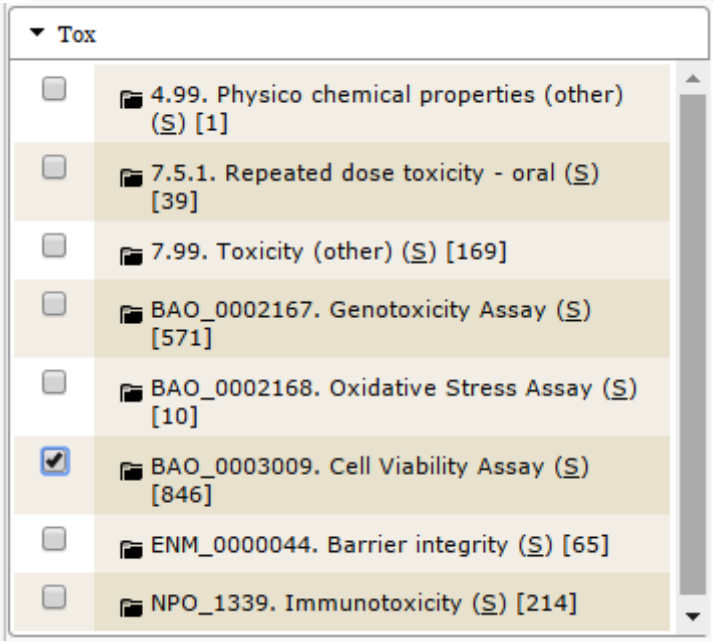


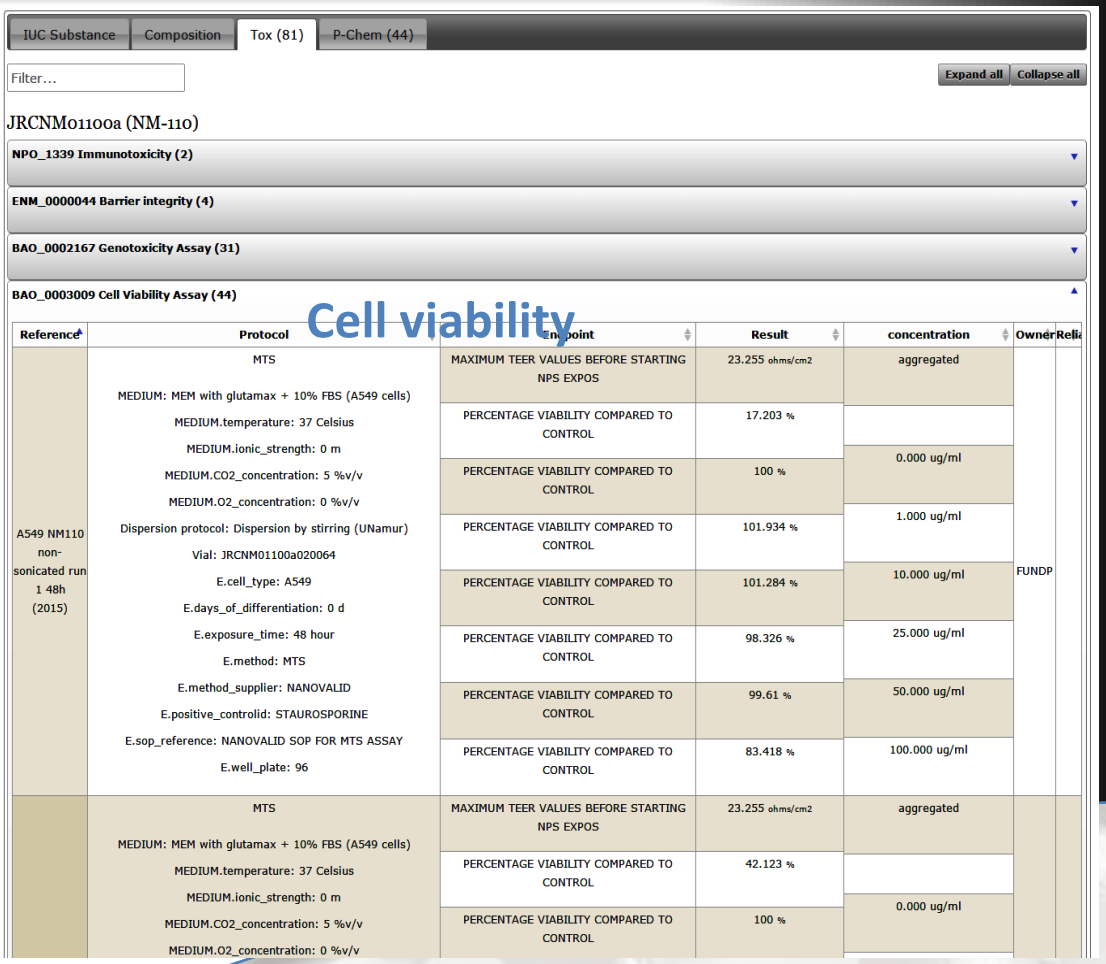
95 materials

~8475 data points

eNanoMapper

NANoREG example: bioassay

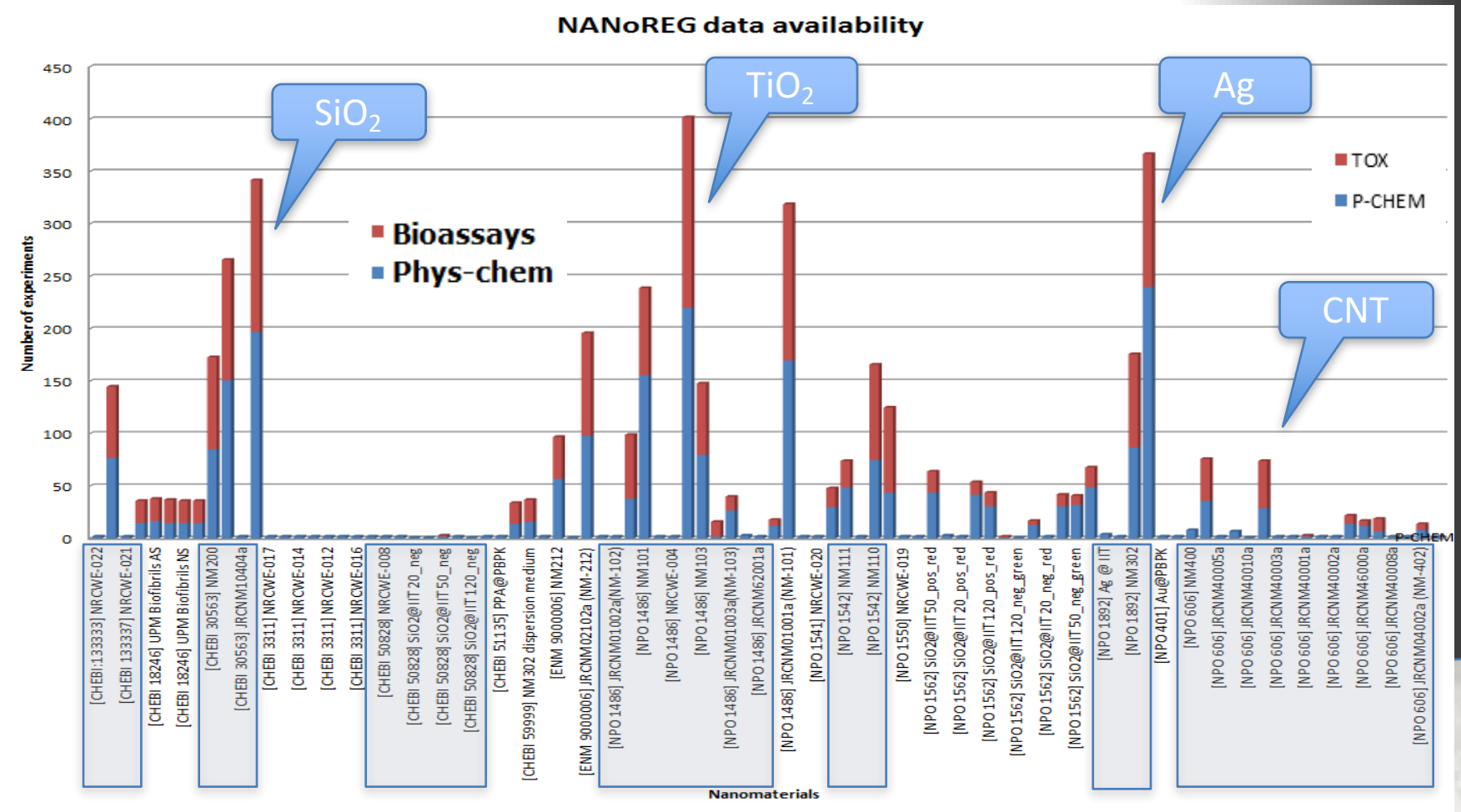




95 materials

~16876 data points

NANoREG data availability (as of Oct 2016)

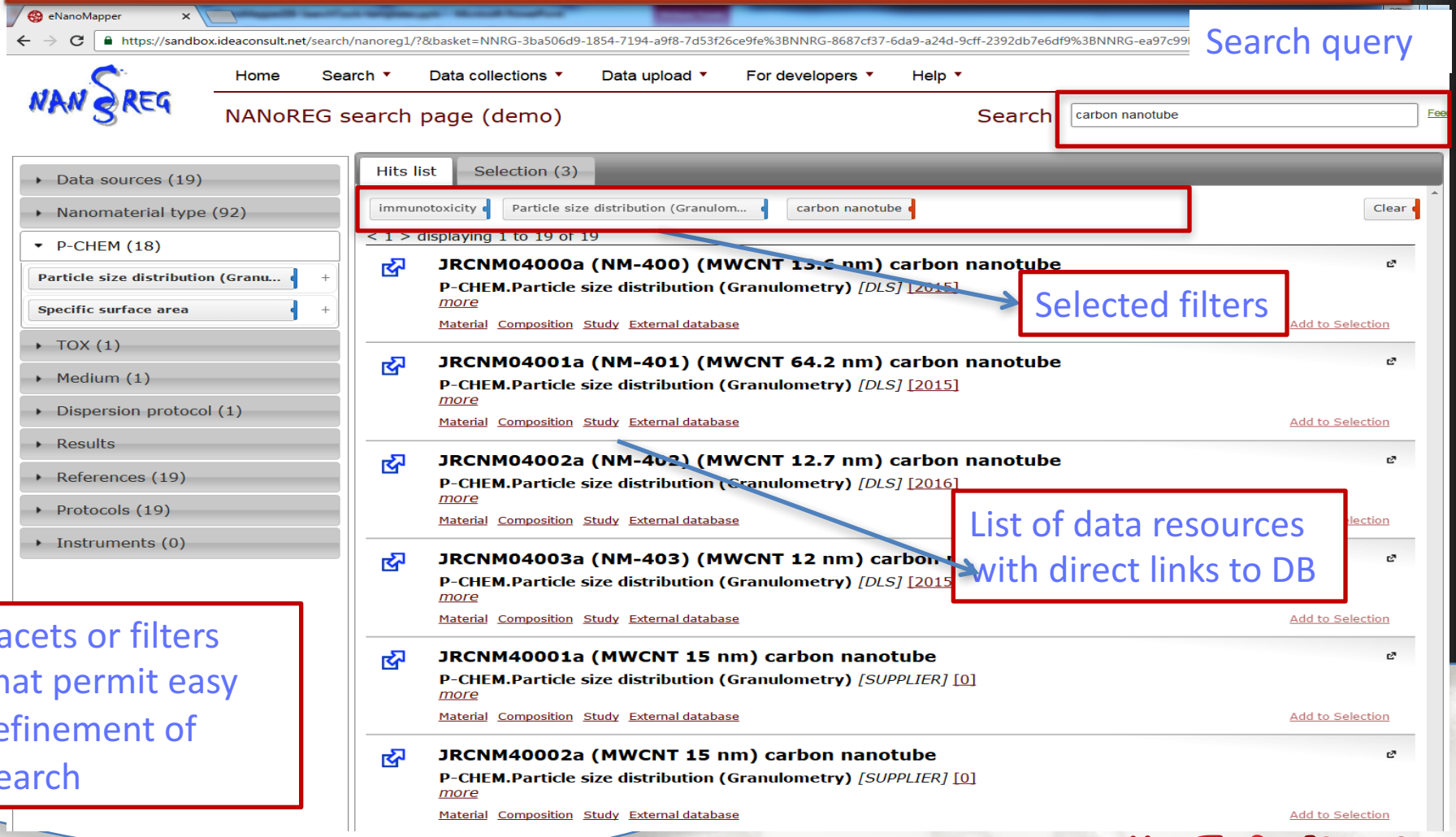


eNanoMapper

- Free text / faceted search
- Chemistry structure and similarity search
- Data access via API
- Semantic search
- Search integration

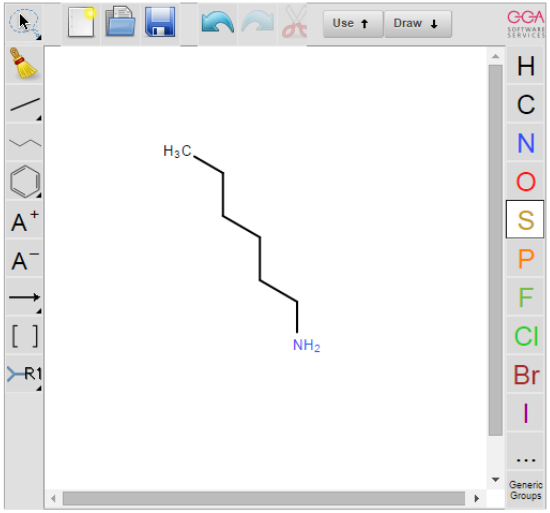
SEARCH TOOLS

NANoREG DB search application
https://sandbox.ideaconsult.net/search/nanoreg1

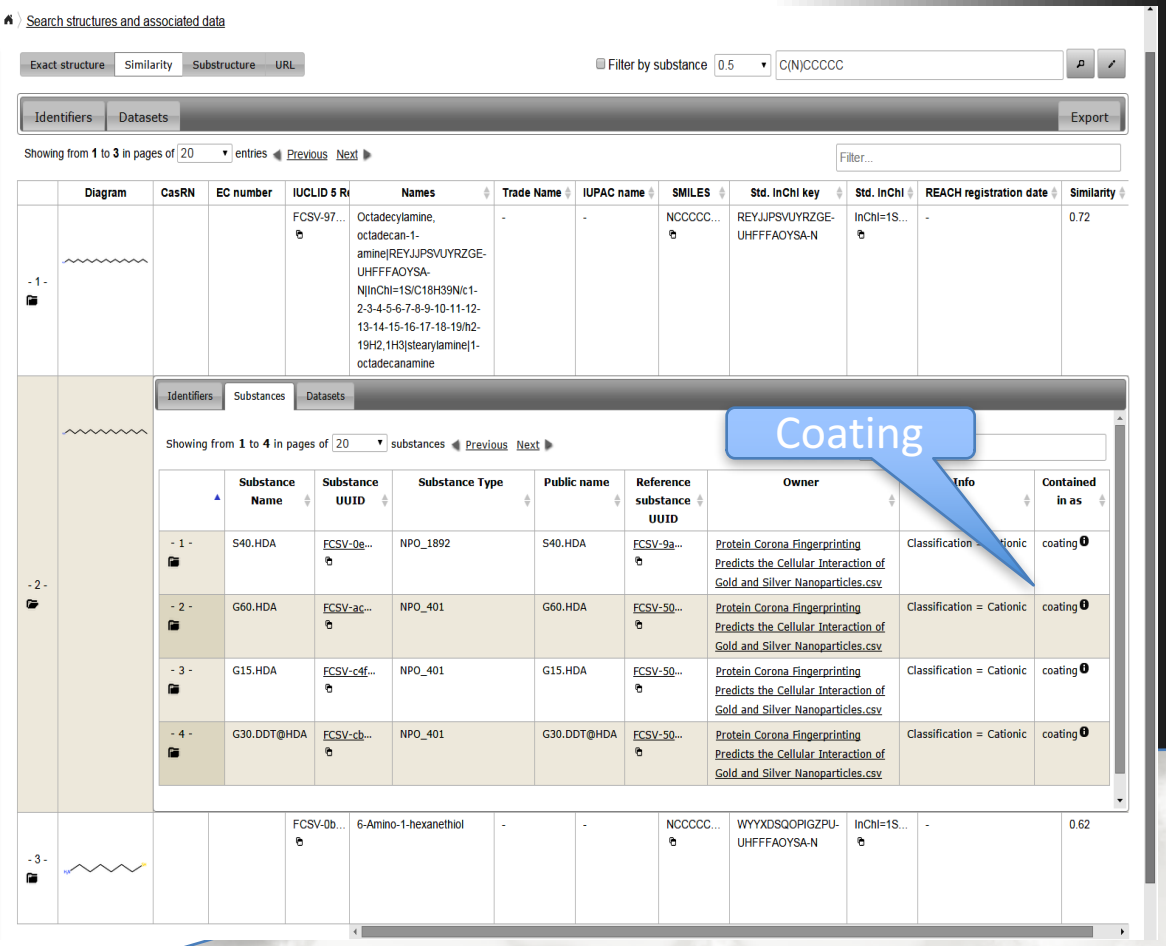


Search tools: chemistry

- Chemical structure search: exact, similarity, substructure

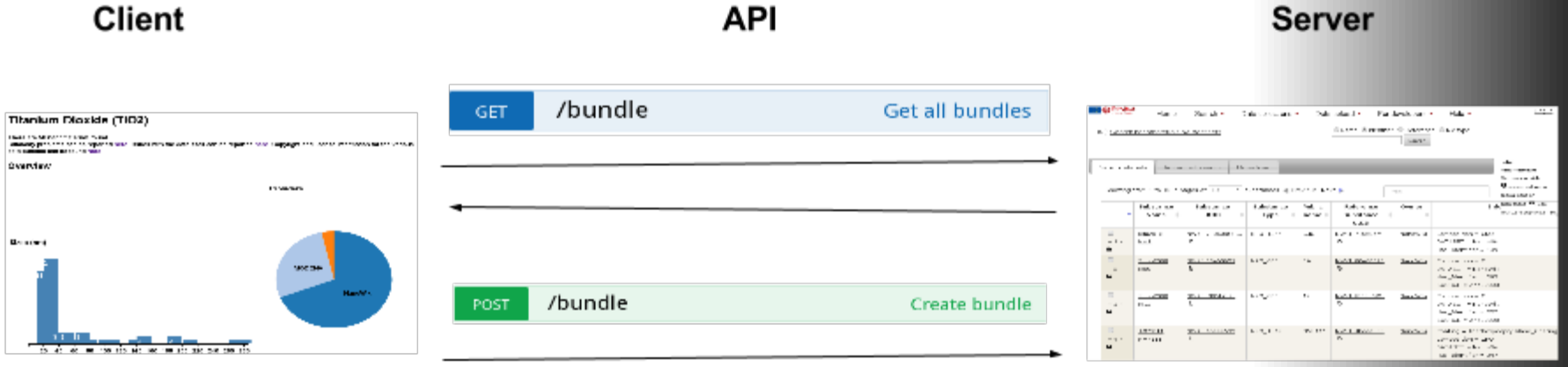


Chemical similarity is a pivotal concept in cheminformatics, encompassing a variety of computational methods quantifying the extent to which two chemical structures resemble each other.



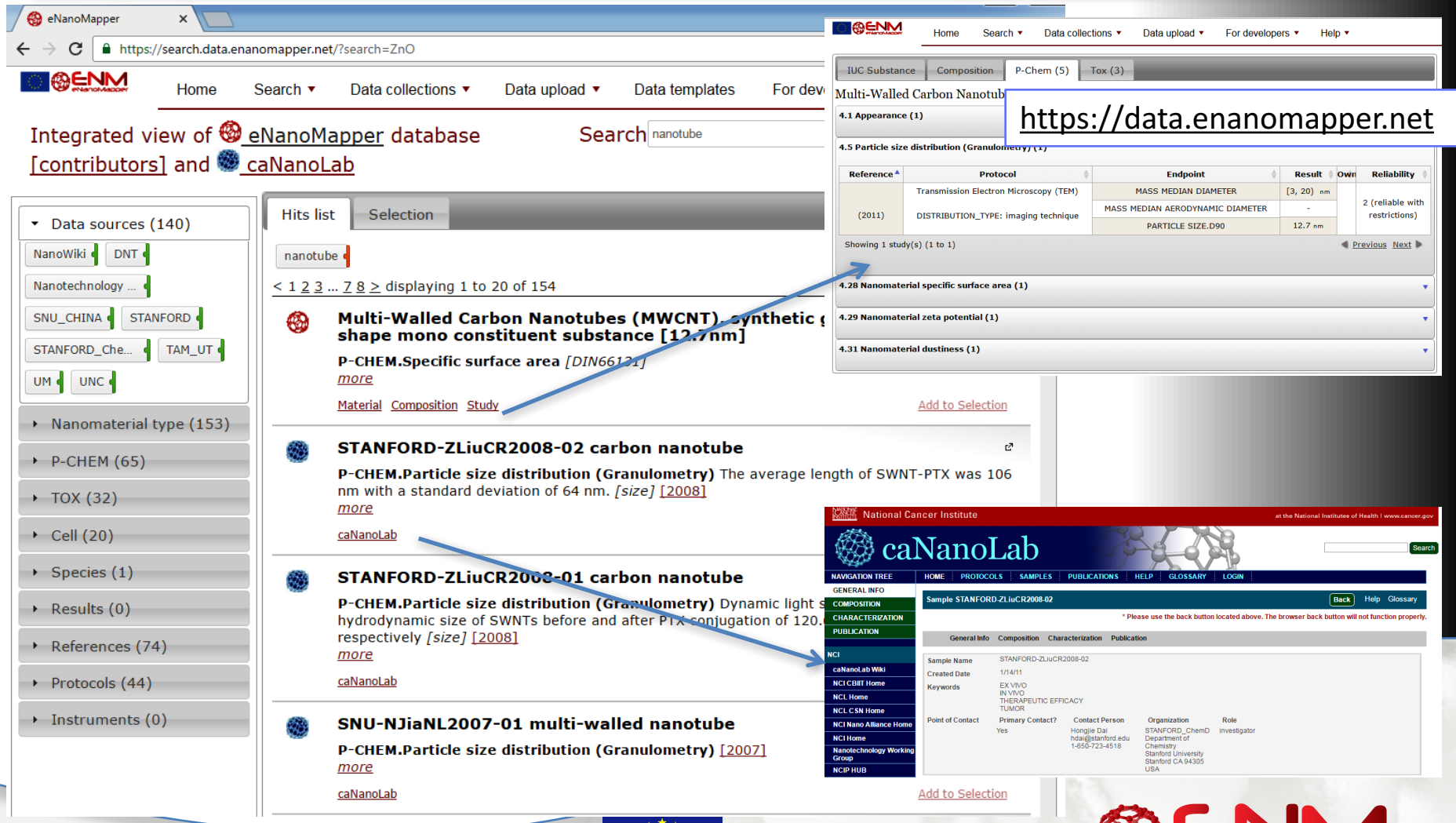
eNanoMapper

Data access: web browser, API http://enanomapper.github.io/API/



- *REST API: a way computer programs talk to one another. Can be understood in terms of how a programmer sends instructions between programs.*
- **Access the database via any programming language , Workflow systems , Data analysis tools** *(R, JavaScript, Java, Ruby used by eNanomapper partners)*
- **eNanoMapper Tutorials:**
  - **http://www.enanomapper.net/enm-tutorials**
  - **https://github.com/enanomapper/tutorials**

Search data integration: https://search.data.enanomapper.net



- ISA-TAB, ISA-TAB-NANO
- ISA-JSON
- Excel spreadsheets
- Export formats

I/O SUPPORT: ISA & EXCEL TEMPLATES

ISA-TAB/ISA-JSON

Version 1 – ISA-TAB (Nov 2008)

Data is described in 3 layers

Tab delimited format (*.txt)

Only meta data is stored

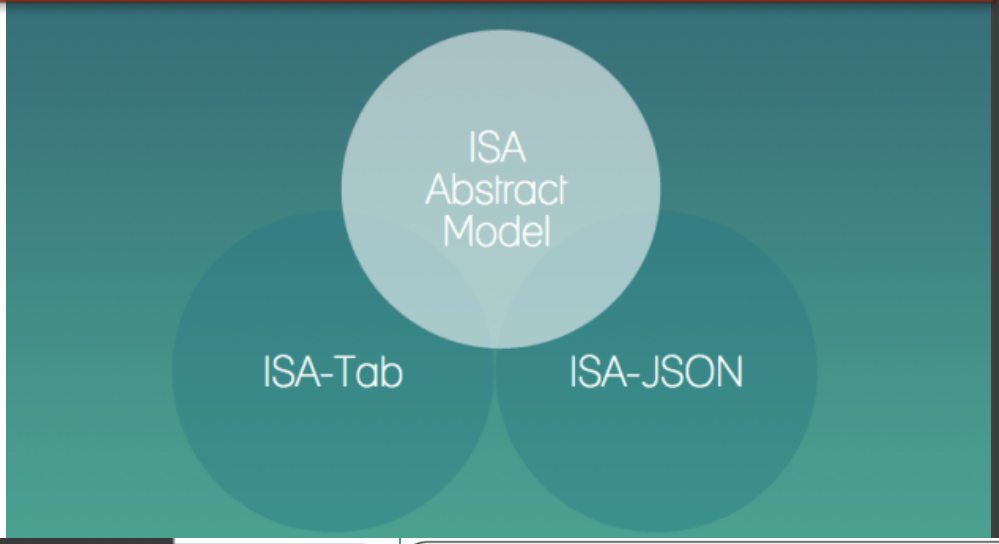
Pointers to the real data

Ontology references

Additional configurations

ISA-JSON version 1 (officially released 2016)

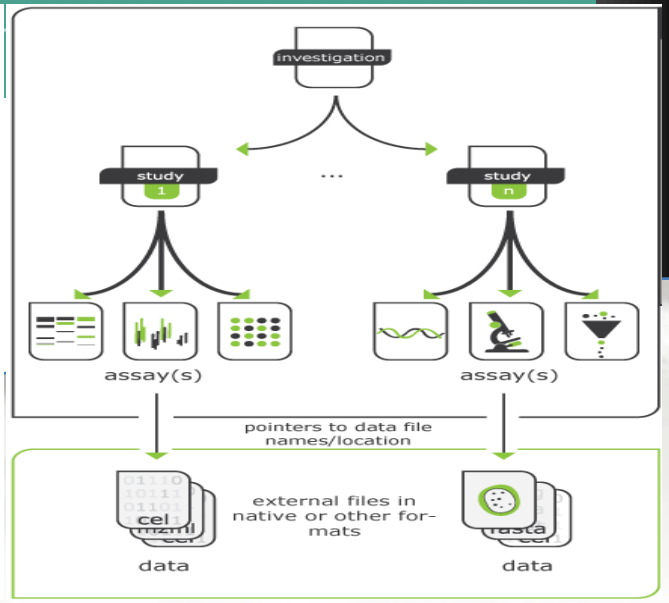
ISA-JSON Version 2 (under development)



investigation
high level concept to link related studies

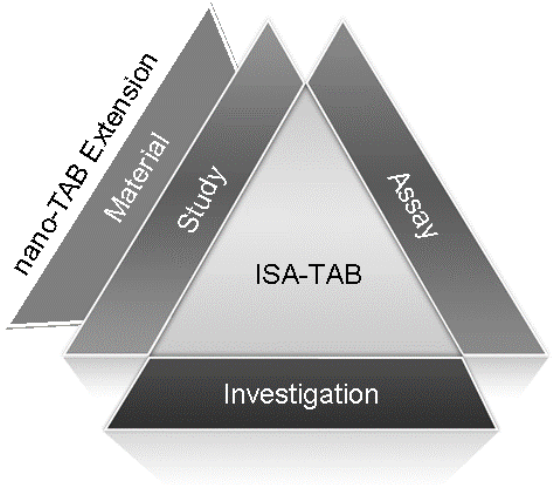
study
the central unit, containing information on the subject under study, its characteristics and any treatments applied.
a study has associated ***assays***

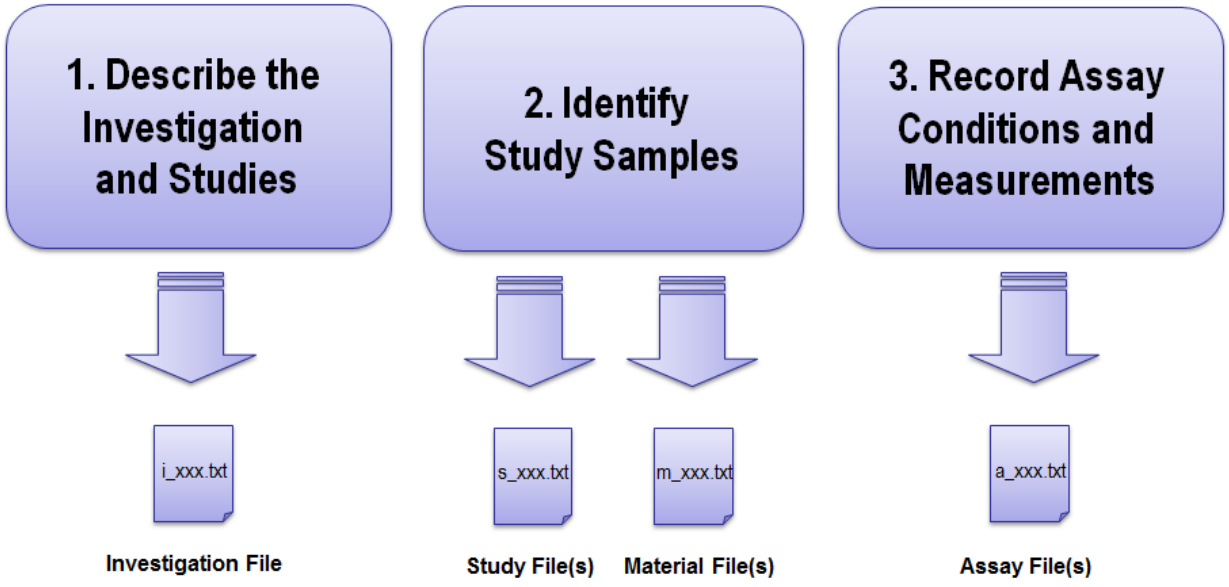
assay
test performed either on material taken from the subject or on the whole initial subject, which produce qualitative or quantitative measurements (data)



ISA-TAB-Nano

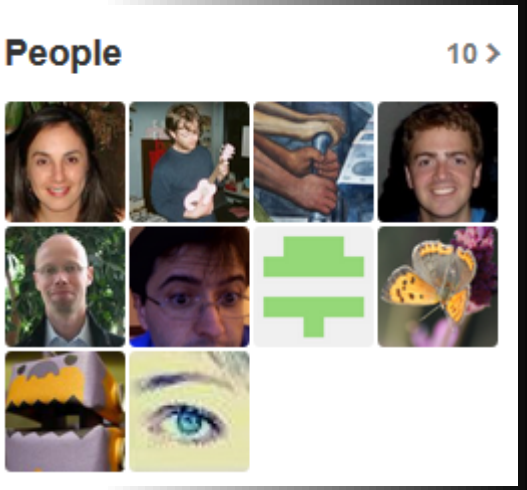
ISA-TAB-Nano uses the three primary files of ISA-TAB investigation file, study file, and assay file; and introduces a fourth file called the **material file**.





ISA-JSON project

https://github.com/ISA-tools/isa-api

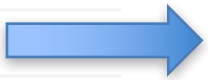


- Developed by S.Sansone group (University of Oxford) and collaborators
- Python based **ISA** API library
- New data format based on JSON describes the ISA experimental graph
- Full support of ISA-TAB (released ISA-JSON v.1)
- More efficient data storage than the TAB delimited
- New extended ISA v.2 (under development)

ISA-JSON schemas

https://github.com/ISA-tools/isa-api/tree/master/isatools/schemas/isa_model_version_1_0_schemas/core

- assay_schema.json
- comment_schema.json
- data_schema.json
- factor_schema.json
- factor_value_schema.json
- material_attribute_schema.json
- material_schema.json
- ontology_annotation_schema.json
- ontology_source_reference_schema.json
- organization_schema.json
- parameter_schema.json
- person_schema.json
- process_schema.json
- protocol_schema.json



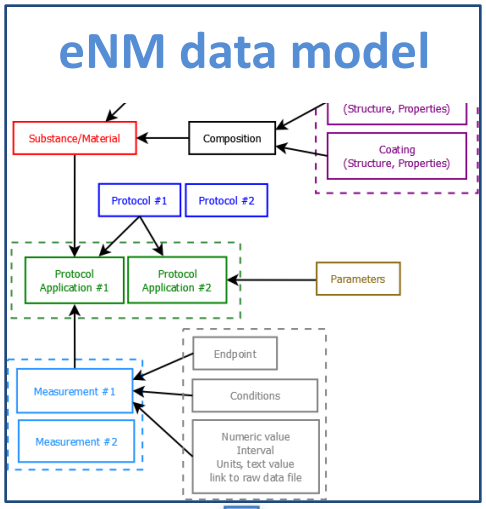
```
{
    "$schema": "http://json-schema.org/draft-04/schema",
    "title" : "ISA investigation schema",
    "description" : "JSON-schema representing an investigation in the ISA model",
    "type" : "object",
    "properties" : {
        "identifier" : { "type" : "string" },
        "title" : { "type" : "string"},
        "description" : { "type" : "string"},
        "submissionDate" : { "type" : "string", "format" : "date-time"},
        "publicReleaseDate" : { "type" : "string", "format" : "date-time"},
        "commentCreatedWithConfiguration" : {
            "$ref": "comment_schema.json#",
            "name": "Created With Configuration"
        },
        "commentLastOpenedWithConfiguration" : {
            "$ref": "comment_schema.json#",
            "name": "Last Opened With Configuration"
        },
        "ontologySourceReferences" : {
            "type" : "array",
            "items" : {
                "$ref": "ontology_source_reference_schema.json#"
            }
        },
```

eNanoMapper ISA-JSON export

ISA JSON schemas

- assay_schema.json
- comment_schema.json
- data_schema.json
- factor_schema.json
- factor_value_schema.json
- investigation_schema.json
- material_attribute_schema.json
- material_schema.json
- ontology_annotation_schema.json
- ontology_source_reference_schema.json
- organization_schema.json
- parameter_schema.json
- person_schema.json
- process_schema.json
- protocol_schema.json
- publication_schema.json
- sample_schema.json
- source_schema.json
- study_schema.json

eNM data model



additional data files

Single ISA-JSON file

export

isa-api

ISA-TAB files

Conversion to pojo

Material JSON schema

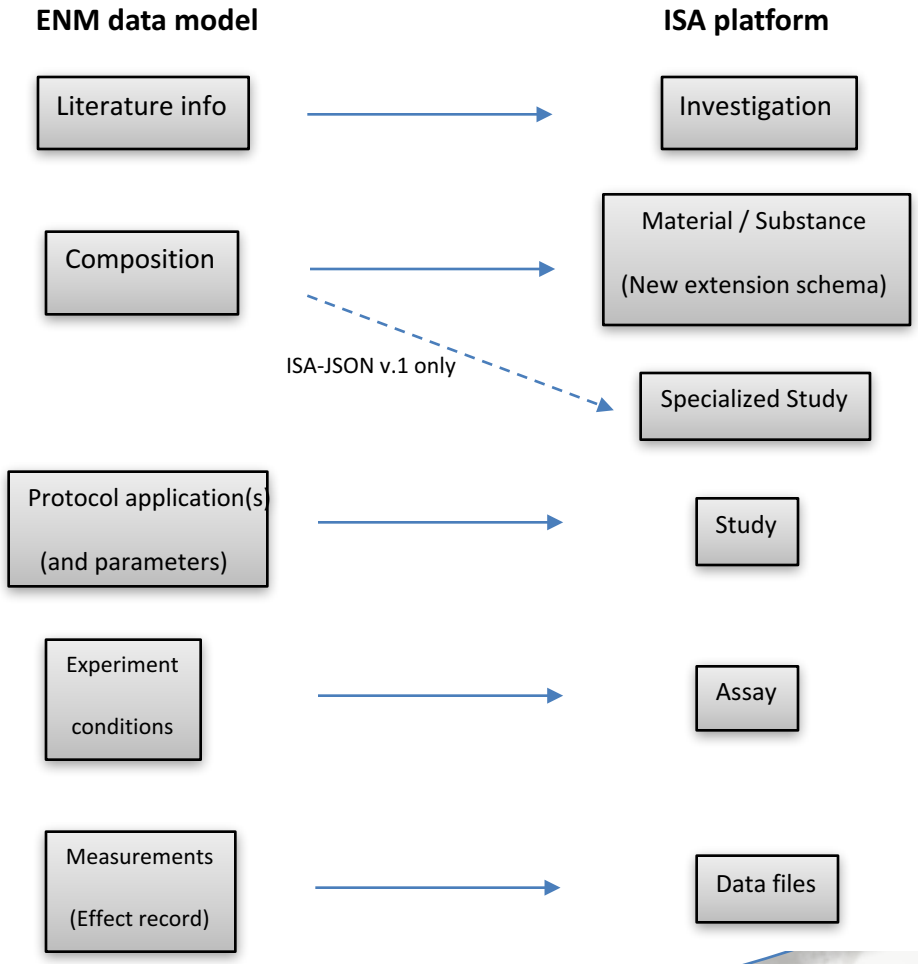
Store eNM substances as objects corresponding to the ISA schemas

- Assay.java
- Comment.java
- Component.java
- Data.java
- Factor.java
- FactorValue.java
- Investigation.java
- Material.java
- MaterialAttribute.java
- MeasurementType.java
- OntologyAnnotation.java
- OntologySourceReference.java

Export configuration

eNM eNanoMapper

eNanoMapper – ISA mapping



Workflow

- The ISA JSON schema is used to generate Java classes *(next slide)*
- The ISA Java classes correspondence to the eNanoMapper data model *(this slide)*
- The data is loaded into eNanoMapper data model
- Converted into ISA model
- Exported into ISA-JSON
- ISA-JSON can be converted to ISA-TAB

ISA (v.1) Java classes

- Assay.java
- Comment.java
- Component.java
- Data.java
- Factor.java
- FactorValue.java
- Investigation.java
- Material.java
- MaterialAttribute.java
- MeasurementType.java
- OntologyAnnotation.java
- OntologySourceReference.java
- Organization.java
- Parameter.java
- Person.java
- Process.java
- Protocol.java
- Publication.java
- Role.java
- Sample.java
- Source.java
- Study.java
- TechnologyType.java
- Value.java

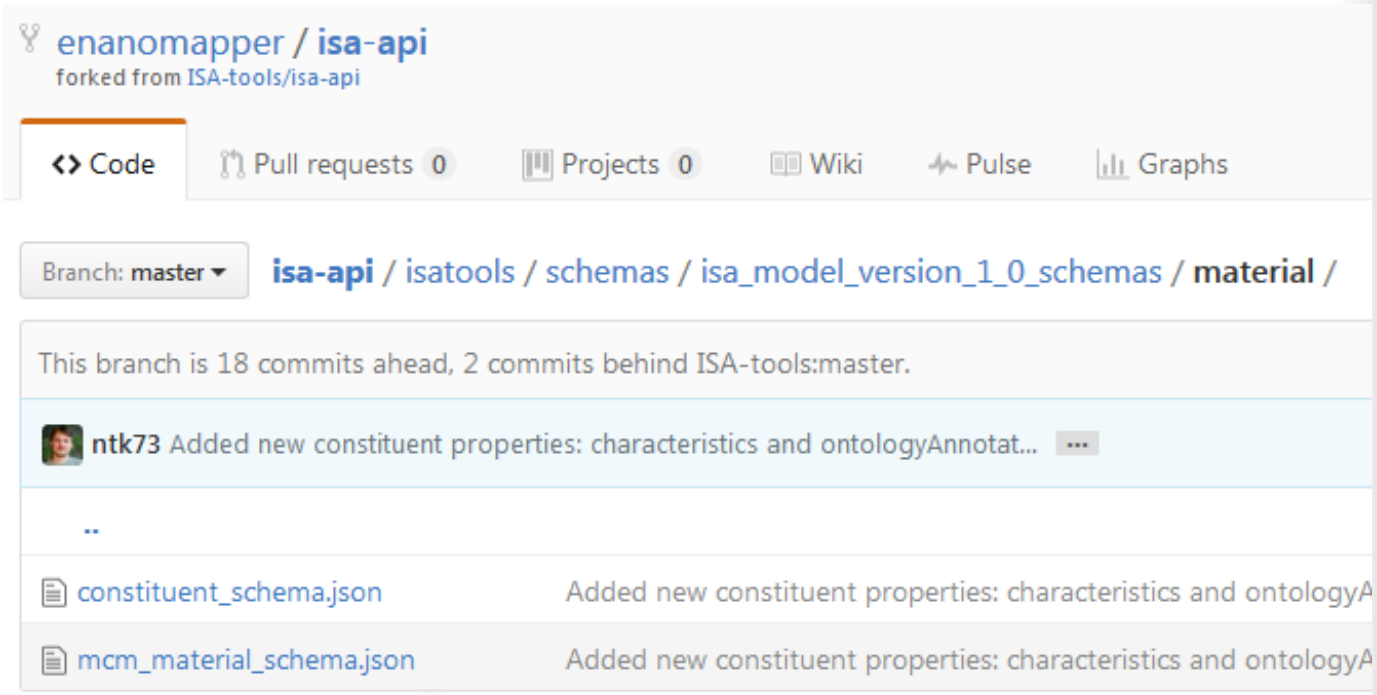
Investigation.java
Investigation
- commentCreatedWithConfiguration
- commentLastOpenedWithConfiguration
- description
- identifier
- ontologySourceReferences
- people
- publications
- publicReleaseDate
- studies
- submissionDate
- title

```
@JsonProperty("identifier")
public String identifier;
@JsonProperty("title")
public String title;
@JsonProperty("description")
public String description;
@JsonProperty("submissionDate")
public Date submissionDate;
@JsonProperty("publicReleaseDate")
public Date publicReleaseDate;

@JsonProperty("commentCreatedWithConfiguration")
public Comment commentCreatedWithConfiguration;
@JsonProperty("commentLastOpenedWithConfiguration")
public Comment commentLastOpenedWithConfiguration;
@JsonProperty("ontologySourceReferences")
public List<OntologySourceReference> ontologySourceReferences = new A
@JsonProperty("publications")
public List<Publication> publications = new ArrayList<Publication>();
@JsonProperty("people")
public List<Person> people = new ArrayList<Person>();
@JsonProperty("studies")
public List<Study> studies = new ArrayList<Study>();

}
```

Study.java
Study
- assays
- description
- identifier
- people
- processSequence
- protocols
- publications
- publicReleaseDate
- samples
- sources
- studyDesignDescriptors
- submissionDate
- title

```
public class Study {

    @JsonProperty("identifier")
    public String identifier;
    @JsonProperty("title")
    public String title;
    @JsonProperty("description")
    public String description;
    @JsonProperty("submissionDate")
    public Date submissionDate;
    @JsonProperty("publicReleaseDate")
    public Date publicReleaseDate;
    @JsonProperty("publications")
    public List<Publication> publications = new ArrayList<Publication>();
    @JsonProperty("people")
    public List<Person> people = new ArrayList<Person>();
    @JsonProperty("studyDesignDescriptors")
    public List<OntologyAnnotation> studyDesignDescriptors = new ArrayList<Ontolog
    @JsonProperty("protocols")
    public List<Protocol> protocols = new ArrayList<Protocol>();
```

ISA-JSON material extension



Contributing new extension to isa-api

JSON schema corresponding to ISA-TAB-Nano material file

eNanoMapper

ISA-JSON material schema

$schema : http://json-schema.org/draft-04/schema#

title : Material

description : Definition of Material (or Substance)

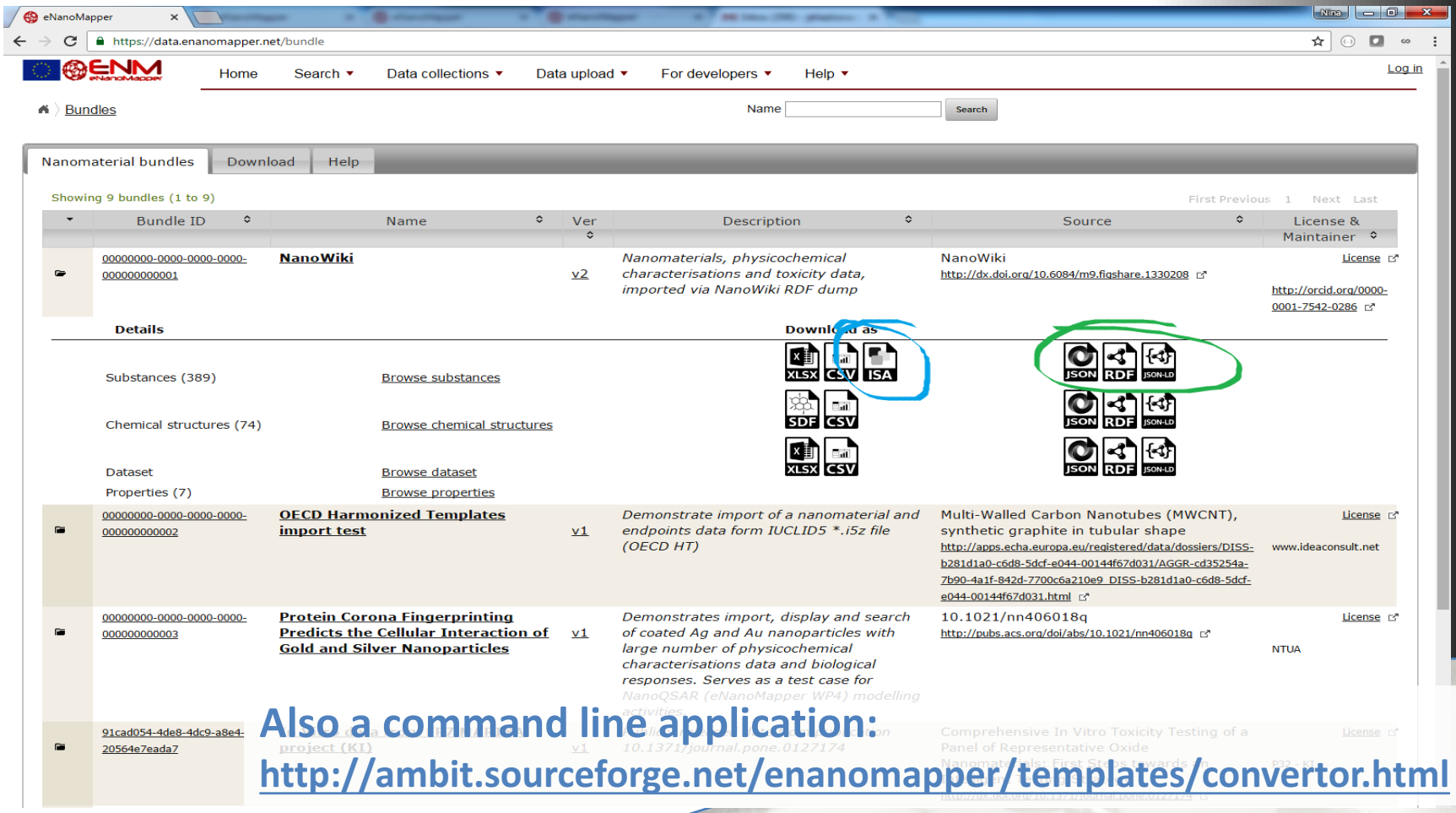
type : object

▼ properties {14}

- ▶ @id {2}
- ▶ lotIdentifier {1}
- ▶ name {2}
- ▶ sourceName {1}
- ▶ root {2}
- ▶ {2}
- ▶ description {1}
- ▶ synthesis {1}
- ▶ designRationale {1}
- ▶ intendedApplication {2}
- ▶ characteristics {2}
- ▶ mcmType {2}
- ▶ chemicalName {2}
- ▶ dataFiles {2}

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "title": "Constituent",
  "description": "Definition of a constituent of a material or another
    constituent",
  "type": "object",
  "properties":{
    "@id": { "type": "string", "format": "uri" },
    "name": {
      "type": "string",
      "description":"Constituent name"
    },
    "role": { "type": "string" },
    "description": { "type": "string" },
    "synthesis": { "type": "string" },
    "linkages":{
      "type": "array",
      "items": {
        "type": "object",
        "properties": {
          "constituent": {"type": "string", "format": "uri" },
          "linkageType": {"type": "string" }
        }
      }
    },
    "characteristics" : {
      "type" : "array",
      "items" :  {
        "$ref": "material_attribute_value_schema.json#"
      }
    },
    "ontologyAnnotation" : {
      "$ref": "ontology_annotation_schema.json#"
    }
  }
}
```

Data export: ISA-JSON, RDF, etc.



Also a command line application:
http://ambit.sourceforge.net/enanomapper/templates/convertor.html

Data Import: EU NanoSafety Cluster Excel templates

Two types of Excel templates:

1) ISA-TAB Logic templates (NANoREG)

Not strictly following the ISA-TAB and ISA-TAB-Nano formats, designed around ISA-Tab-logic, i.e. structuring the data in investigation-study-assay related groups.

One sheet: many materials, one assay, both metadata and data; CC BY-SA 4.0 license

http://www.nanoreg.eu/media-and-downloads/templates/269-templates-for-experimental-data-logging

2) One material, one assay;

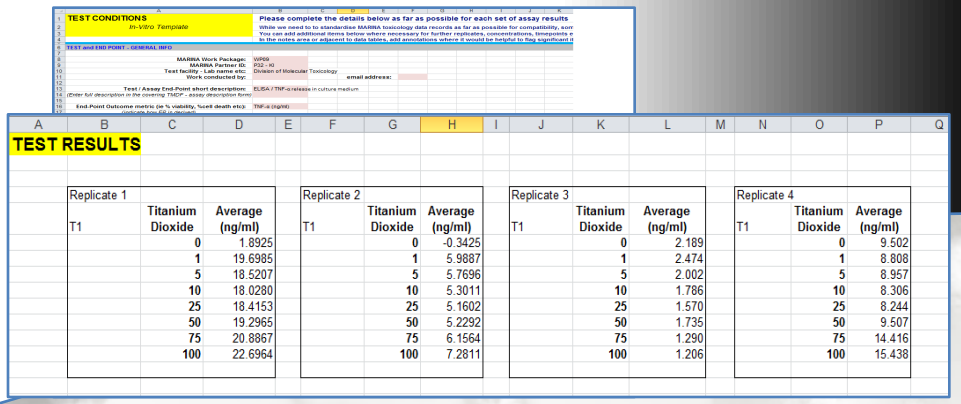
first sheet: metadata; next sheets: raw and processed data (used by several EU projects; many variations, not publicly available)

Solution: A configurable Excel Parser for custom spreadsheets

JSON configuration mapping the Excel layout into the eNanoMapper data model

(next slide)

https://github.com/enanomapper/nmdataparser



TEST CONDITIONS

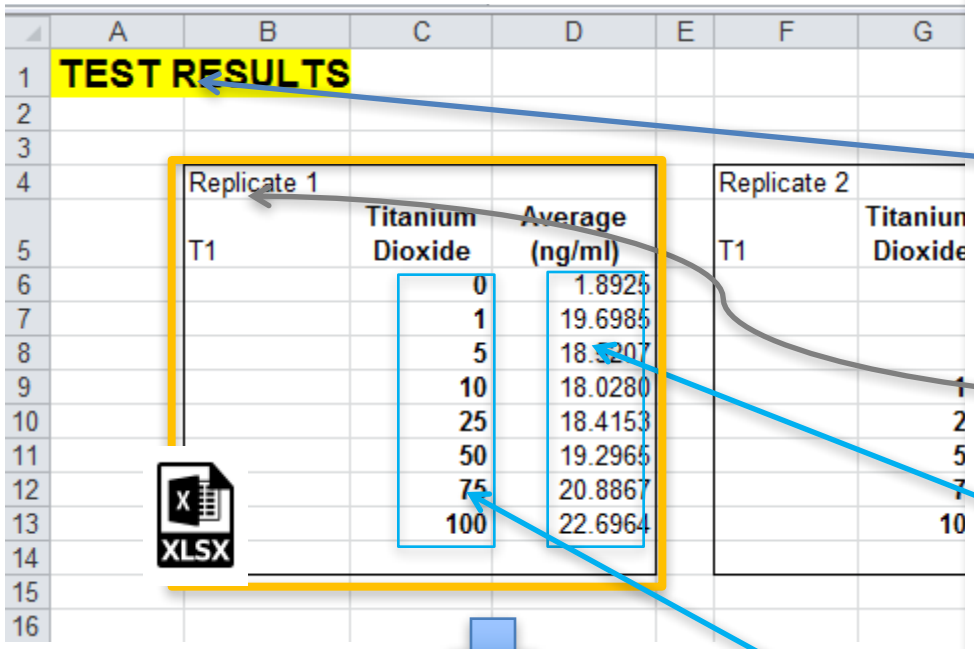
In-Vitro Template

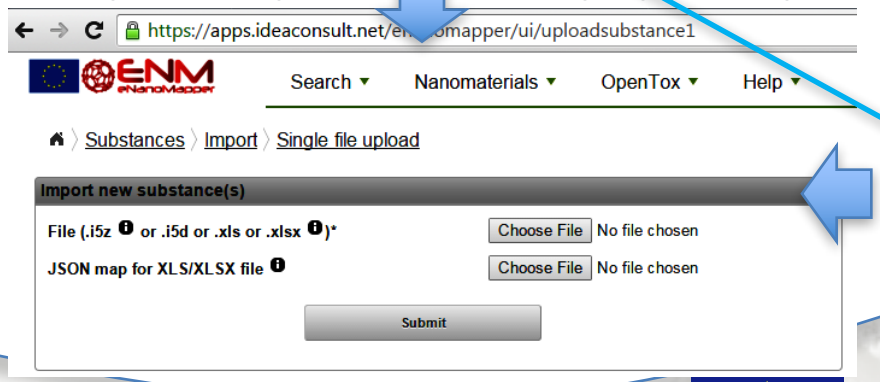
Please complete the details below as far as possible for each set of assay results

TEST RESULTS

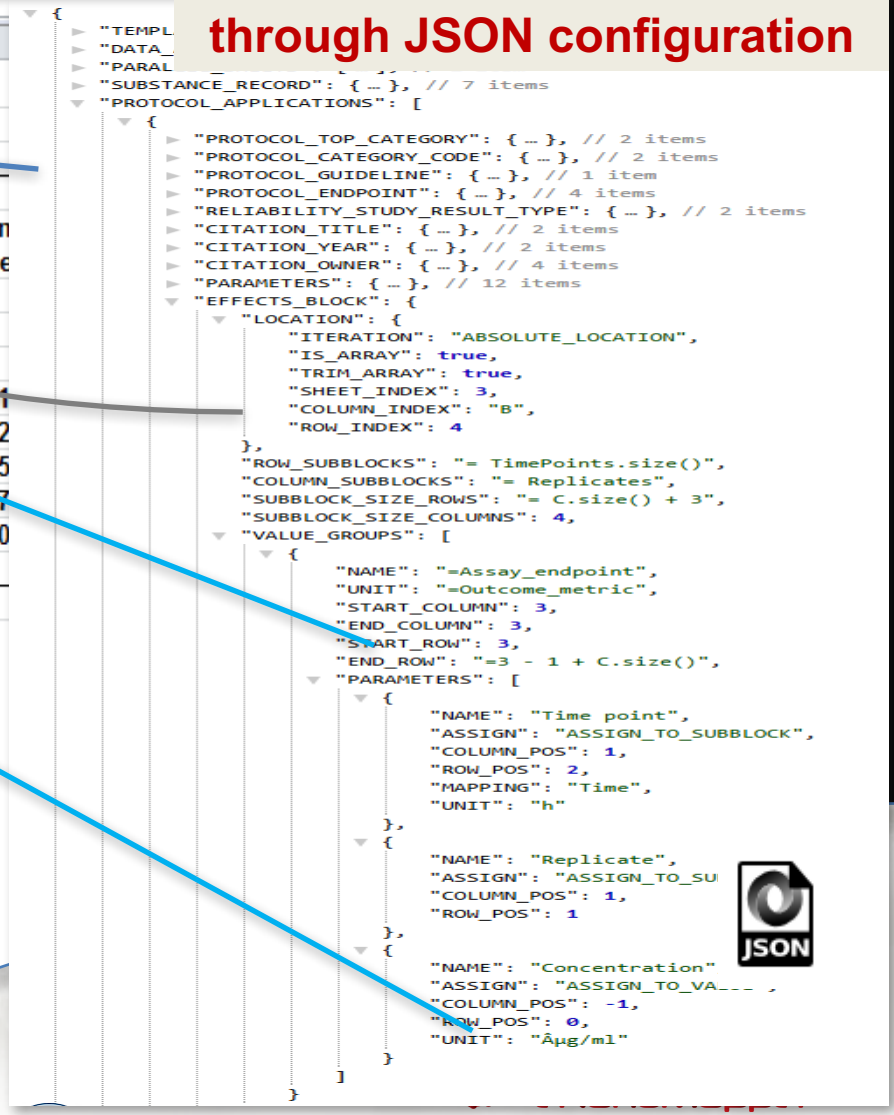
Replicate 1			Replicate 2			Replicate 3			Replicate 4		
T1	Titanium Dioxide	Average (ng/ml)	T1	Titanium Dioxide	Average (ng/ml)	T1	Titanium Dioxide	Average (ng/ml)	T1	Titanium Dioxide	Average (ng/ml)
	0	1.8925		0	-0.3425		0	2.189		0	9.502
	1	19.6985		1	5.9887		1	2.474		1	8.808
	5	18.5207		5	5.7696		5	2.002		5	8.957
	10	18.0280		10	5.3011		10	1.786		10	8.306
	25	18.4153		25	5.1602		25	1.570		25	8.244
	50	19.2965		50	5.2292		50	1.735		50	9.507
	75	20.8867		75	6.1564		75	1.290		75	14.416
	100	22.6964		100	7.2811		100	1.206		100	15.438

Mapping the spreadsheet content into the data model

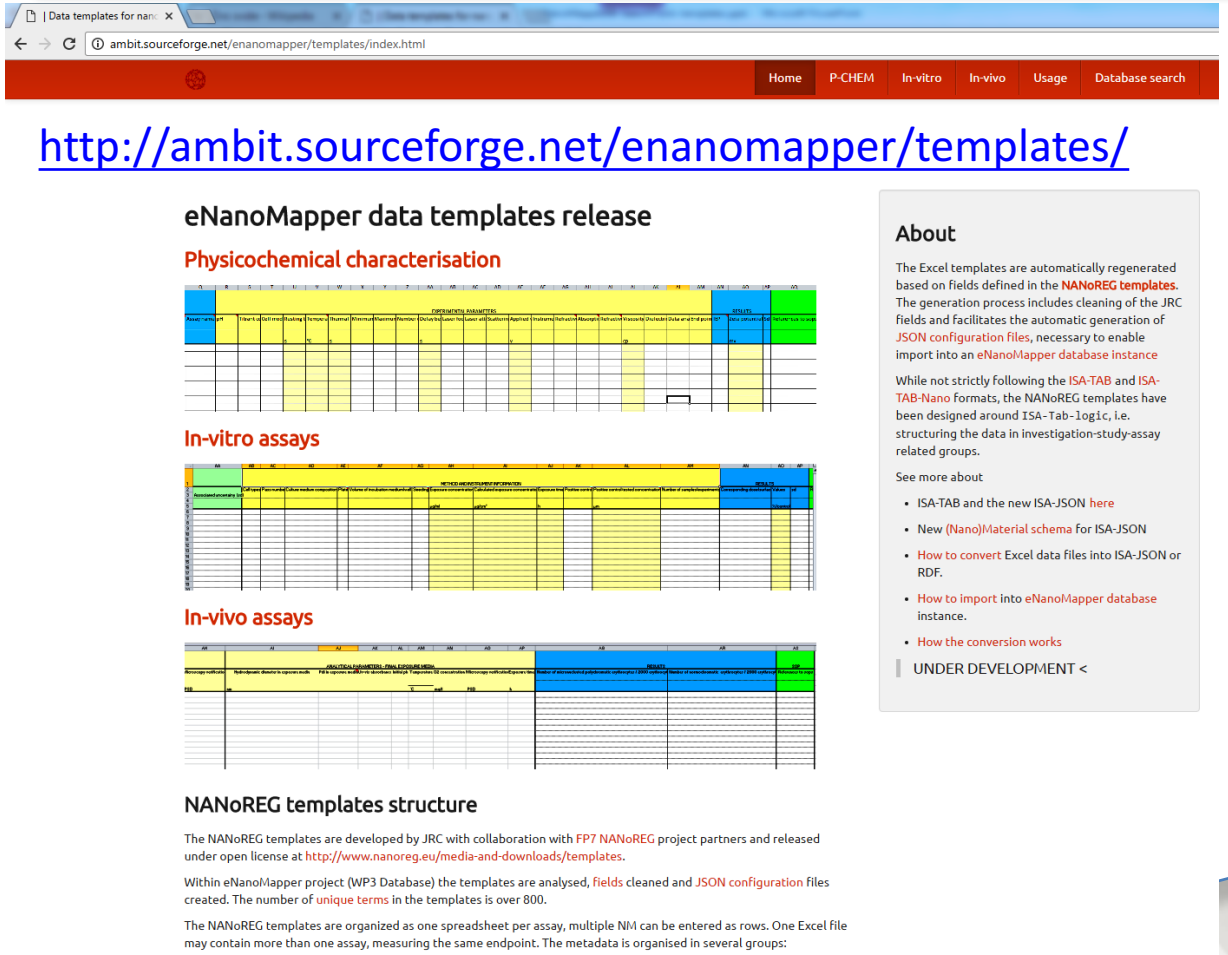




through JSON configuration

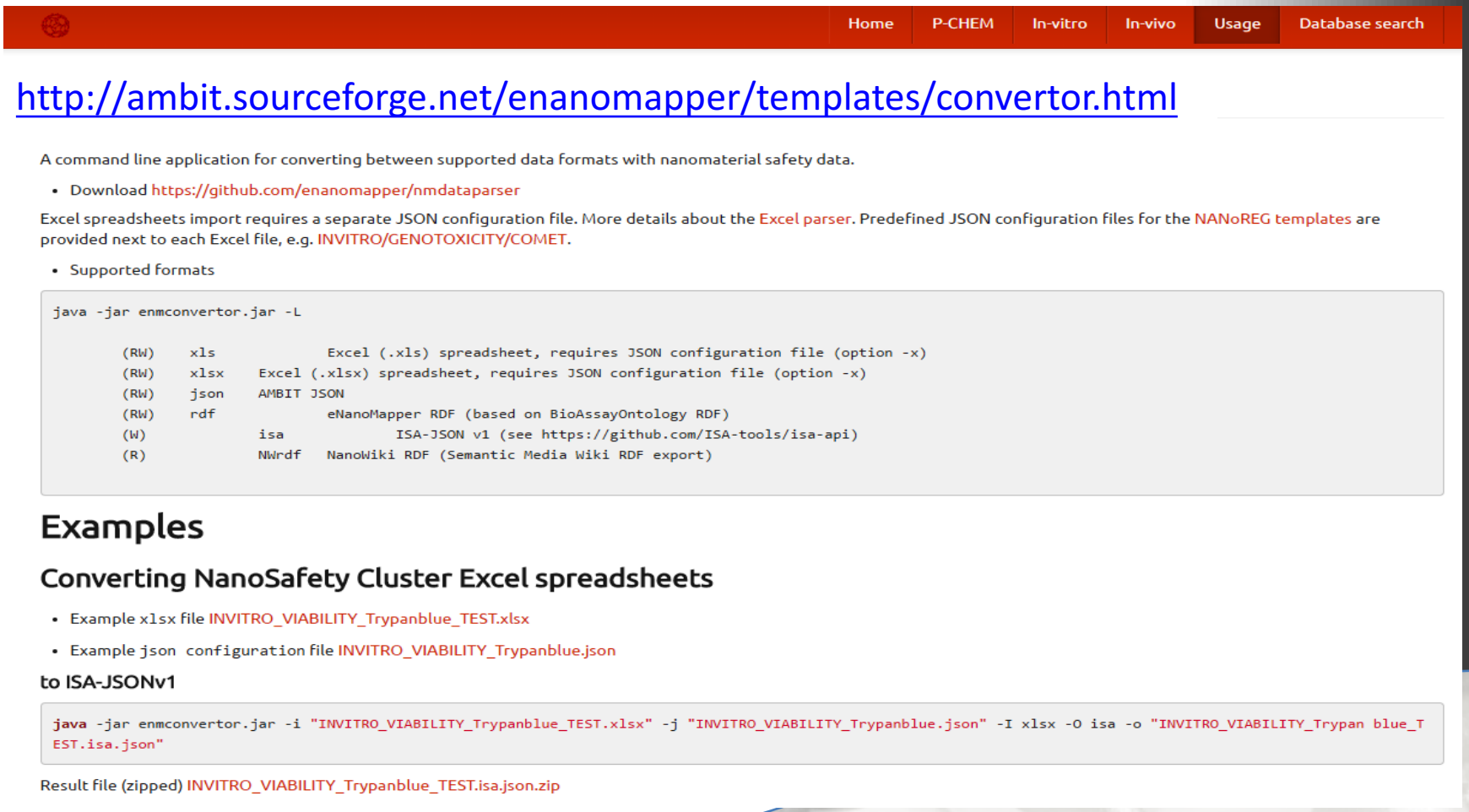


Automating the JSON configuration (under development)



- Extract all fields from NANoREG templates;
- Cleanup (typos, units), sync between different templates;
- Annotation;
- Generate the templates based on cleaned fields and JSON configurations;
- **One-assay Excel template + JSON, ready for upload;**
- Next step – dynamic generation

Finally – a bonus: command line XLSX- ISA-JSON/RDF convertor



Summary

- Open source database and web application
- Demo at https://data.enanomapper.net
- Import: Excel templates, RDF, OECD OHT, SQL
- Export: ISA-JSON, RDF, XLSX
- Enables distributed setup: many databases; search integration https://search.data.enanomapper.net
- Integration with data analysis tools
- Search tools: free text, chemistry, semantic
- More on ontology: NanoWG, Dec 8, by Maastricht U.

Karolinska Institutet

VTT

misvik biology

Maastricht University

Leading in Learning!

EMBL-EBI

Douglas Connect

Working communities

gmbh

in silico toxicology

IOFA consult

Questions?

THANK YOU!