



# DELIVERABLE REPORT D3.2

## Data Management System with extended search capabilities

GRANT AGREEMENT:	604134
ACRONYM:	eNanoMapper
NAME:	eNanoMapper - A Database and Ontology Framework for Nanomaterials Design and Safety Assessment
PROJECT COORDINATOR:	Douglas Connect GmbH
START DATE OF PROJECT; DURATION:	1 February 2014; 36 months
PARTNER(s) RESPONSIBLE FOR THIS DELIVERABLE:	IST, IDEA, UM
DATE:	07.03.2016
VERSION:	V.1.1



Call identifier	FP7-NMP-2013-SMALL-7
Document Type	Deliverable Report
WP/Task	WP3/T3.3 T3.4
Document ID	eNanoMapper D3.2
Status	Final

Partner Organisations	<ul style="list-style-type: none"> <li>• Douglas Connect, GmbH (DC)</li> <li>• National Technical University of Athens (NTUA)</li> <li>• In Silico Toxicology (IST)</li> <li>• Ideaconsult (IDEA)</li> <li>• Karolinska Institutet (KI)</li> <li>• European Bioinformatics Institute (EMBL-EBI)</li> <li>• Maastricht University (UM)</li> <li>• Misvik Biology (MB)</li> </ul>
Authors	Nina Jeliaskova (IDEA) Nikolay Kochev (IDEA) Vedrin Jeliaskov (IDEA) Micha Rautenberg (IST) Egon Willighagen (UM) Friederike Ehrhart (UM) Penny Nymark (MB) Final review by Barry Hardy (DC)
Purpose of the Document	[T3.3, T3.4]
Document History	1.First draft, 14/12/2015 2.Second draft 01/02/2016 3.Completed version, 20/04/2016

# TABLE OF CONTENTS

## Table of Contents

1. EXECUTIVE SUMMARY .....	5
2. INTRODUCTION .....	6
3. DATA MANAGEMENT SYSTEM WITH EXTENDED SEARCH CAPABILITIES.....	7
3.1 SCIENTIFIC QUESTIONS .....	7
3.2 EXTENDED SEARCH FUNCTIONALITIES.....	10
3.2.1 ENDPOINT SEARCH.....	10
3.2.2 CHEMICAL SEARCH.....	10
3.2.3 SEMANTIC SEARCH.....	11
3.3 ISA-TAB-NANO SUPPORT .....	12
3.3.1 MATERIAL SCHEMA.....	15
3.3.2 ISA-JSON EXPORT .....	17
3.4. CONFIGURABLE EXCEL PARSER.....	19
3.5 DATABASE CONTENT UPDATES.....	22
3.6 CLIENT SOFTWARE .....	25
3.6.1 JAVASCRIPT CLIENTS .....	25
3.6.2 R ENVIRONMENT .....	26
3.6.3. Ferret.ai.....	27
4. CONCLUSION.....	28
5. BIBLIOGRAPHY .....	29

## GLOSSARY

Abbreviation / acronym	Description
SPARQL	SPARQL Protocol is an RDF query language, that is, a semantic query language for databases
RDF	Resource Description Framework (RDF) <a href="https://www.w3.org/RDF/">https://www.w3.org/RDF/</a>
API	Application Programming Interface

# 1. EXECUTIVE SUMMARY

The eNanMapper prototype database, described in the eNanMapper report “Technical specification and initial implementation of the protocol and data management web services” (2015), was extended with new search functionality: chemically-aware search, several options for searching experimental data, semantic search including SPARQL endpoint free text search, supported by eNanMapper ontology and annotated database entries. The types of the new queries are informed by the list of scientific questions identified through the requirements analysis and during the eNanMapper ontology working meetings. Test and public instances of the prototype database are available at <http://apps.ideaconsult.net/enmtest> and <http://data.enanmapper.net> respectively. The data content of <http://data.enanmapper.net> has been updated. The configurable parser, enabling the import of data stored in spreadsheet templates, was enhanced with new capabilities. ISA-Tab import and export implementation is ongoing, based on a new version of the ISA-Tab specification with JSON schemas, published by the ISA team at the University of Oxford e-Research Centre. The schema and code for parsing and generation of ISA files are under active development and available on GitHub, including an ISA-Nano extension developed by eNanMapper work supporting database development. The database Application Programming Interface (API) and search functionality are used by visualization (JavaScript) and modelling (R) packages, described in this report. The database API and content is also being used by the modelling prototype, currently under development within eNanMapper.

## 2. INTRODUCTION

This report introduces new and extended functionalities of the eNanoMapper prototype database first described in the eNanoMapper report on “Technical specification and initial implementation of the protocol and data management web services” (D3.1, 2015). The database provides support for upload and search for nanomaterials and experimental data through a REST web services Application Programming Interface (API) and a demonstration web browser interface. In addition to the search API for experimental data and chemical structure search modules, new modules providing different types of queries are implemented. The types of the new queries are informed by the list of scientific questions identified through the requirement analysis and during the eNanoMapper ontology working meeting held in May 2015. Free text search extensions were developed, including free text search with query expansion based on the eNanoMapper ontology and annotated database entries with an indication of the relevance of the hits. All data in the database is targeted for annotation with relevant ontology entries from the composite eNanoMapper ontology. A SPARQL endpoint has been launched at <https://purl.enanomappper.org/test/> and initialized with eNanoMapper ontology entries.

Test and public instances of the prototype database have been deployed at <http://apps.ideaconsult.net/enmtest> and <http://data.enanomappper.net> respectively. The data content was updated and extended. The configurable parser, described in D3.1, enabling import of the data stored in the NSC spreadsheet templates, has been extended with new features. The prototype database is also being used by the modelling prototype, currently under development within the modelling workpackage (WP4).

The export formats are being extended with the new ISA JSON format, following the new ISA specification. During the ontology working meeting held at the EMBL-EBI in Hinxton in May 2015, we discussed in details the eNanoMapper requirements and the current developments of ISA-Tab and ISA-Tab-Nano with the core ISA-Tab developers from Oxford University. In October 2015 the ISA-Tab team published a new version of the ISA-Tab specification, based on JSON schemas<sup>1</sup>. The schema and code for parsing and generation of ISA files are under active development and available on GitHub. The new ISA specification was discussed during an eNanoMapper - ISA team teleconference (3 Nov 2015) and the first ISA Working group teleconference (9 Dec 2015). While the initial approach for eNanoMapper to provide ISA-Tab-Nano support was to extend the ISAcreeator tool<sup>2</sup>, we have revised the strategy towards using the new ISA JSON schema and developing an ISA-Nano extension. We have also been in close contact with the US NANO

<sup>1</sup> <https://github.com/ISA-tools/isa-api/tree/master/isatools/schemas>

<sup>2</sup> <https://github.com/enanomappper/ISAvailator-ISAconverter-BIImanager>

Working Group and provided comments to the challenges of NM data representation in ISA-Tab-Nano, identified within the FP7 NanoPUZZLES (<http://nanopuzzles.eu/>) project.

## 3. DATA MANAGEMENT SYSTEM WITH EXTENDED SEARCH CAPABILITIES

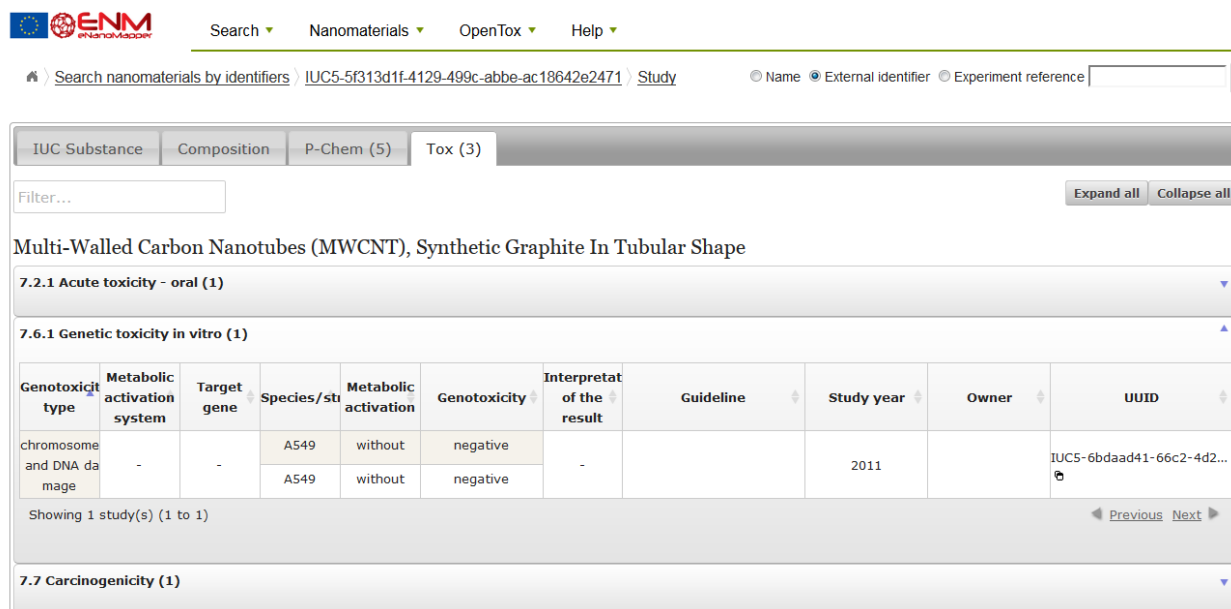
### 3.1 SCIENTIFIC QUESTIONS

The types of the new queries are informed by the list of scientific questions identified through the requirements analysis and during the ontology working meeting in May 2015. This list is currently used as repository for desired functions (to solve scientific questions) the eNanoMapper database should provide. A paper about the implementation of these functions into the database is currently in preparation. Some of the scientific questions and the implemented functions are given below. General questions for nanomaterial information filtered by e.g. size, material and available endpoint data have already been implemented. Pathway related functions, e.g. search for biological pathways affected by a certain nanomaterial, are currently under development. The questions involving computational modelling are subject of WP4 and are not described here.

**Examples for scientific questions concerning nanomaterial properties and how they are solved using the eNM database:**

#### 1. How genotoxic is a certain nanomaterial?

The eNM database offers the possibility to search for genotoxicity assays (e.g. *Search for Endpoint = Genotoxicity*) and by filtering the offered results for this nanomaterial (e.g. *name, UUID, substance type, publication/DOI, etc.*) the available results are presented (Figure 1).



The screenshot shows the eNMM database search interface. At the top, there are navigation links: Search, Nanomaterials, OpenTox, and Help. Below these, a search bar contains the identifier 'IUC5-5f313d1f-4129-499c-abbe-ac18642e2471' and a 'Study' button. The search results are displayed in a table with columns: IUC Substance, Composition, P-Chem (5), and Tox (3). The results are filtered by 'Multi-Walled Carbon Nanotubes (MWCNT), Synthetic Graphite In Tubular Shape'. The table shows two rows of data for 'Genotoxicity' and 'Metabolic activation system'.

Genotoxicity type	Metabolic activation system	Target gene	Species/st	Metabolic activation	Genotoxicity	Interpretation of the result	Guideline	Study year	Owner	UUID
chromosome and DNA damage	-	-	A549	without	negative	-		2011		IUC5-6bdaad41-66c2-4d2...
			A549	without	negative					

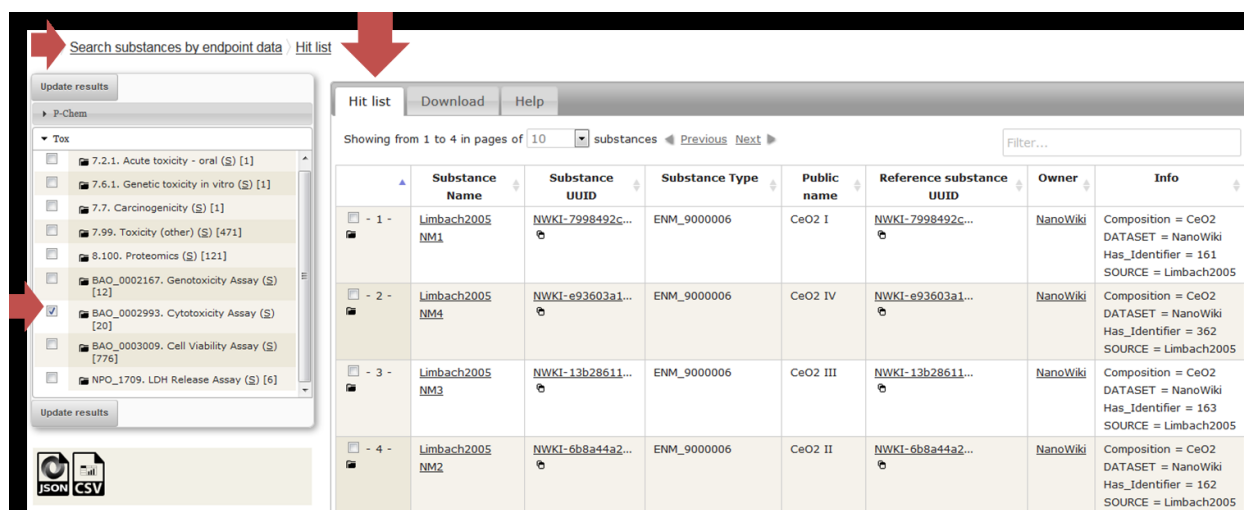
Showing 1 study(s) (1 to 1)

Figure 1. Search function example in the eNM database for genotoxicity of MWCNT.

## 2. How can we merge data from different assays/cell lines?

Merging several assays (including similar experiments) and showing their results on the same screen is available via endpoint search (see Figure 2). It is possible to filter the results for general terms like cell viability or cytotoxicity which include a variety of different assays - but with the same result. The list is dynamic and the number and the specificity of the assays in the list depend on how the imported data was annotated. For example the data from the FP7 MARINA project provided by KI was annotated as *LDH release assay*, which is one of the most abundant assays for cell viability/cytotoxicity assessment. With more (properly annotated) data being imported, the list will extend automatically. This function will give all available studies in the eNanoMapper database which investigate cell viability which was measured by LDH assay.





**Figure 2.** Endpoint search gives the possibility to search via phys/chem or toxicology endpoints. For cell viability assays the suitable search terms cytotoxicity, cell viability and LDH release assay are currently available as filter functions.

The filter functions in the hit list are currently limited to the textual information available in the table (i.e. substance names and study owners, external identifiers). Desirable would also be terms which allow filtering and ranking for quality measures like data size, data completeness, and date of data production (or publication). The derived results can be downloaded in the form of JSON or CSV files for further application. These additional types of search (publication date, different types of filtering) are now supported via free text search application, described in the eNanoMapper report on “User application for searching and downloading eNanoMapper data (D5.5).

### 3. Give us all nanomaterials of a certain size/core material/linkers

Query for size and material: “search by phys/chem properties” and select size to get all available studies which give size information. Different types of search, including size or range of endpoint results has been described in the publication (figure 14 in [1]). For identifying nanomaterials by core, shell or linker material, the function searching “nanomaterials by composition” can be used (figure 13 in [1]). The hits are labelled (e.g. core or shell), but the user interface doesn’t have an option to specify a filter at the time of the query. Such filtering will be added in future, as it is already supported by the API.

The scientific questions addressed by the database allow the creation of datasets that can serve as input to the eNanoMapper modelling infrastructure. The user can then apply the services and algorithms described in the eNanoMapper report on “nQSAR Modelling infrastructure” (D4.3) in order to analyze the data, identify key parameters, generate predictive models and gain insight on the relationships between nanoparticles and their adverse effects.

## 3.2 EXTENDED SEARCH FUNCTIONALITIES

The eNanoMapper data API offers access to a variety of searches by substance identifier, any combination of measurement endpoints, and/or chemical structure (Figure 6 in [1]). The overview of the eNanoMapper database architecture, supporting different format conversion and search technologies is shown on Figure 3.

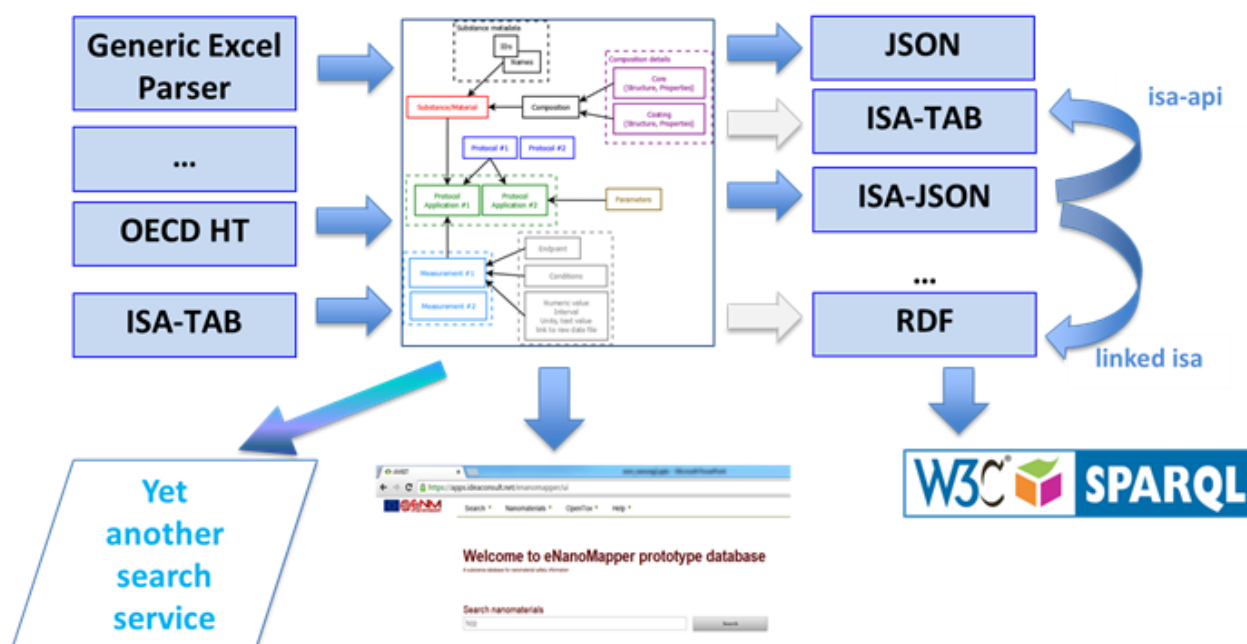


Figure 3. Architecture Overview.

### 3.2.1 ENDPOINT SEARCH

The search by endpoints is illustrated in the Scientific Questions section above (3.1) and screenshots are also shown in Figure 1 and Figure 2 in the eNanoMapper database publication [1].

### 3.2.2 CHEMICAL SEARCH

The chemical similarity search in the eNanoMapper prototype database enables querying by a chemical structure of a NM component and highlighting the results as a core, coating or functionalisation component (Figure 13 in [1]). The API is tightly integrated with a chemical structure and chemical similarity search, as described in [1]. The reason for the wide adoption of the similarity approach in cheminformatics is the assumption of the “similarity property principle” or “neighbourhood behaviour”, namely that “similar compounds should have similar properties”. This principle puts the chemical similarity at the core of methods and tools supporting property prediction, structure–activity relationship, chemical database screening,

virtual screening in drug design, and diversity selection. The similarity assessment based on structure analogy is the basis of read across and chemical grouping for chemical substances. The pros and cons of the approach have been reviewed by multiple researchers; we give a brief summary in the “chemical search” section of the publication [1]. In the context of nanosafety assessment there is not yet a standardized approach for NM similarity; however a number of attempts for NM grouping and read across have been published recently. As chemical search is typically provided by chemical structure databases, but not specialized nanomaterial databases, we are not aware of another publicly available database solution supporting both nanomaterial composition and chemical aware search.

### 3.2.3 SEMANTIC SEARCH

#### 3.2.3.1. SPARQL ENDPOINT

A SPARQL endpoint deployed at <https://purl.enanomapper.net/sparql/> has been established and can be accessed by SPARQL queries using the standard SPARQL HTTP query protocol. The endpoint is based on the RDF database software 4store<sup>3</sup>. 4store is open source software available under the GNU General Public License, version 3. Access control to the endpoint is implemented with the Apache webserver LDAP module. The webserver can directly use the eNanoMapper user-database to authenticate users and allow access to project members. SPARQL queries can be requested by a graphical interface at <https://purl.enanomapper.net/test/> or by REST calls (e.g.: via the command line tool *curl*) at <https://purl.enanomapper.net/sparql/>. Dependent on the SPARQL query form, different output formats are available. SELECT queries can return XML, JSON and n-triples. CONSTRUCT queries return RDF and turtle format.

The full eNanoMapper ontology data is imported into the RDF store. A method to synchronize RDF store content with new eNanoMapper ontology development, held at <https://github.com/enanomapper/ontologies>, is also introduced. As a first testing data, NanoWiki substance and compounds data were successfully imported into the SPARQL endpoint. A simple SPARQL query example carried out with the *curl* command line tool is:

```
curl -Lk -u user:password -H 'accept:application/rdf+xml' -d 'query=CONSTRUCT {?s ?p ?o.} FROM <http://purl.enanomapper.org/onto/internal/endpoints.owl> WHERE {?s ?p ?o. FILTER isURI(?s).}' https://purl.enanomapper.org/sparql/
```

#### 3.2.3.1. ONTOLOGY SUPPORTED FREE TEXT SEARCH

Initially, a feasibility setup of a free text search with query expansion based on the eNanoMapper ontology v.1 and annotated database entries was implemented via MySQL free text search engine and illustrated in Figure 1 in [1]. Subsequently, the same approach of using the ontology (version 2 released Sep 2015) as a list of synonyms and query expansion was implemented using

<sup>3</sup> 4store RDF database <http://4store.org>  
eNanoMapper 604134

Apache Solr<sup>4</sup> as a search engine. The user application is described in the eNanoMapper report “User application for searching and downloading eNanoMapper data “ D5.5.

### 3.3 ISA-TAB-NANO SUPPORT

ISA (Investigation-Study-Assay) is a general purpose multi-layer platform for description of complex meta-data needed for the interpretation of experiments developed by S. Sansone’s group at the University of Oxford e-Research Centre [2]. ISA includes three major layers: *Investigation*, *Study* and *Assay* where the actual experimental readouts are stored in an additional data layer. Since the ISA version 1 published in 2008, the ISA platform implementation relies on tab delimited text files and the file format representation is known as ISA-Tab.

The ISA-Tab-Nano project is an effort of the National Cancer Institute (NCI), National Cancer Informatics Program (NCIP) and Nanotechnology Informatics Working Group (US Nano WG). The ISA-Tab-Nano extension [3], includes an additional material file within the study layer, which is used for the description of the NM composition and nominal NM characteristics (size, shape, as reported by the manufacturer).

Support for ISA-Tab import and export is an ongoing activity in WP3, in close collaboration with the ISA-Tab and ISA-Tab-Nano developers at the University of Oxford e-Research Centre and the US Nano WG. We have also provided comments to the challenges of NM data representation in ISA-Tab-Nano, identified within the NanoPUZZLES project [4].

During the eNanoMapper hackathon at EBI (May 2015) the ISA-Tab team (Philippe Rocca-Serra, Alejandra Gonzalez-Beltran) shared their plans to migrate to a new ISA-JSON format as well as the planned development of a new ISA version 2. The ISA v2 is based solely on JSON, where the backward compatibility with the ISA v1 and ISA-TAB format will be preserved via ISA-tools project utilities (Figure 4). The new specification was extensively discussed and worked on during the *ISA hack\_the\_spec-athon*, an ELIXIR-UK Node<sup>5</sup> event (July 20 , 2015), organised and hosted at the University of Oxford, with remote eNanoMapper participation (E. Willighagen, N. Jeliaskova). The eNanoMapper WP3 Database team was intensively involved in the consultations for the new developments of the ISA platform.

<sup>4</sup> <http://lucene.apache.org/solr/>

<sup>5</sup> ELIXIR-UK Node (<http://elixir-uk.org/node-events/2015isa-as-a-fair-research-object2015d-hack-the-spec-event-1> ).

ISA is also an interoperability component from the UK Node contributing to ELIXIR-EXCELERATE platform for integration of data and services (<http://elixir-uk.org/interoperability-infrastructure> )

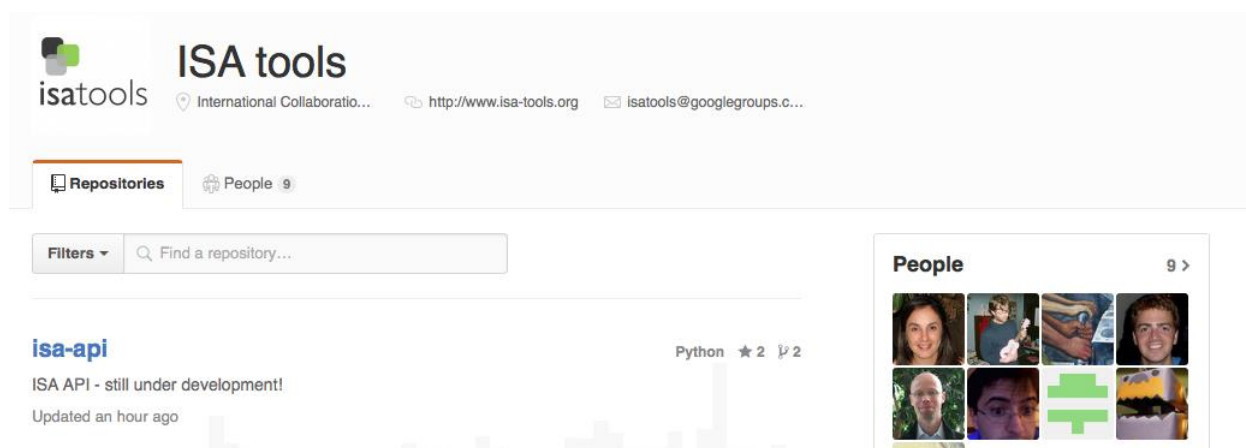


Figure 4. ISA tools GitHub repository

While the new ISA specification is still not officially released at the time of writing this report, the ISA team shared with eNanoMapper (21 Oct 2015) all the important documents (*ISA model version 1.0 - ISA JSON Schemas and JSON representation, ISA 2.0 specifications normative documentation*), schemas and code (Figure 4), and these new developments were discussed in a dedicated eNanoMapper - ISA team meeting in November. Both teams are in close contact since and the communication is documented via the ISA-tools issue tracker (Figure 4).

The new ISA format is labelled *ISA-JSON version 1* and uses JSON instead of tab delimited files as a more flexible and efficient way to store the experimental meta-data. The ISA-JSON specification is based on a set of JSON schemas used to validate the syntactical and logical correctness of the ISA-JSON information. It distinguishes between core components and extensions; for example, the support for nanomaterials is considered an extension.

Every JSON schema (Figure 5, Figure 6) describes an object, representing a specific component of the ISA platform (i.e. investigation, study, factor, etc.). The schemas are available at [https://github.com/ISA-tools/isa-api/tree/master/isatools/schemas/isa\\_model\\_version\\_1\\_0\\_schemas/core](https://github.com/ISA-tools/isa-api/tree/master/isatools/schemas/isa_model_version_1_0_schemas/core).

The ISA-JSON version 2 is not complete and is under active development. The latest core schemas are available at [https://github.com/ISA-tools/isa-api/tree/master/isatools/schemas/isa\\_model\\_version\\_2\\_0\\_schemas](https://github.com/ISA-tools/isa-api/tree/master/isatools/schemas/isa_model_version_2_0_schemas).

ISA-tools / isa-api

Watch 5 Star 2 Fork 2

Code Issues 31 Pull requests 0 Wiki Pulse Graphs

Branch: master isa-api / isatools / schemas / isa\_model\_version\_1\_0\_schemas / core /

New file Find file History

**djcomlab** Add comments to data schema Latest commit b9d9195 3 days ago

..		
assay_schema.json	Additions to schemas	3 days ago
comment_schema.json	Adding ids to schemas and conversion of protocols, fixed test dataset #...	3 days ago
data_schema.json	Add comments to data schema	3 days ago
factor_schema.json	Adding ids to schemas and conversion of protocols, fixed test dataset #...	3 days ago
factor_value_schema.json	Adding ids to schemas and conversion of protocols, fixed test dataset #...	3 days ago
investigation_schema.json	Adding ids to schemas and conversion of protocols, fixed test dataset #...	3 days ago
material_attribute_schema.j...	Adding ids to schemas and conversion of protocols, fixed test dataset #...	3 days ago
material_attribute_value_sc...	Adding ids to schemas and conversion of protocols, fixed test dataset #...	3 days ago
material_schema.json	Adding ids to schemas and conversion of protocols, fixed test dataset #...	3 days ago
ontology_annotation_sche...	Add comments to assay, factor, ontology, person, process, protocol an...	4 days ago
ontology_source_reference...	Add comments to assay, factor, ontology, person, process, protocol an...	4 days ago
organization_schema.json	Add additionalProperties as false constraint	2 months ago
person_schema.json	Adding ids to schemas and conversion of protocols, fixed test dataset #...	3 days ago
process_parameter_value_...	Fix to schema #50	4 days ago

Figure 5. ISA JSON v1 schema



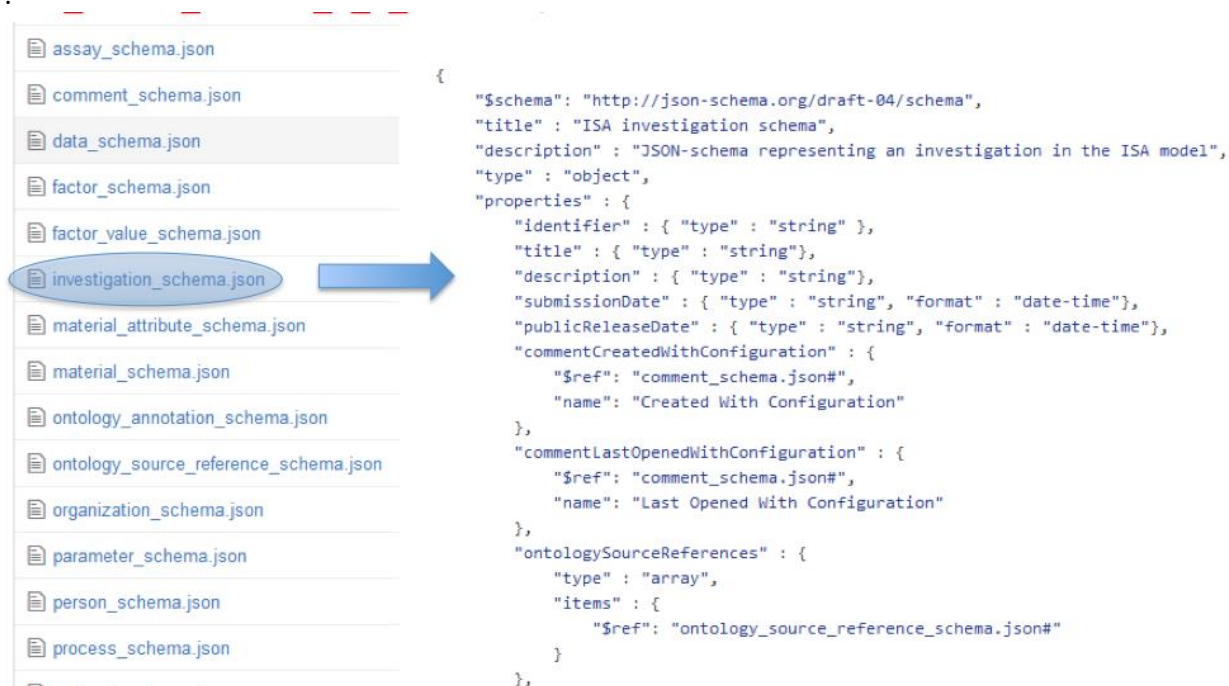


Figure 6. JSON schema for Investigation (corresponds to the ISA-TAB Investigation file)

We developed a working prototype of an export utility from the eNanoMapper database to ISA-JSON format version 1 (see section 3.3.2). The export utility is designed to allow at least two options - export to ISA-JSON version 1 only (with composition information stored as experimental data), and ISA-JSON v1 with nanomaterial extension (analogous to ISA-TAB-NANO format). In case that export to the old ISA-TAB format is needed, ISAtools can be used to convert from ISA-JSON to ISA-TAB and back.

### 3.3.1 MATERIAL SCHEMA

We are currently working on a new *Substance/Material* JSON schema (Figure 7), which will become a NanoMaterial extension of ISA-JSON v1, the counterpart of the ISA-Tab-Nano format (see section 3.3.1). The schema is available at the enanomapper/isa-api GitHub<sup>6</sup> repository which is a fork of the isa-api project, part of the ISA-tools. The schema is developed within close collaboration and consultations with the core ISA-Tab team and the US Nano WG and will be merged with the original isa-api repository when finalised.

<sup>6</sup> <https://github.com/enanomapper/isa-api/tree/master/tests/data/nano>

```

"material": [{
  "sourcename": "MSKCC-CU-UA-CVillaNL2008-02",
  "name": "xxx",
  "root": "SWNT-ODNFAM-NH2_carbon_nanotube",
  "constituents": [{
    "name": "SWNT_ODNFAM-NH2-carbon_nanotube",
    "type": [
      "nanoparticle sample",
      "carbon nanotube"
    ],
    "lot identifier": "",
    "description": "",
    "synthesis": "",
    "linkage": [{
      "name": "SWNT-NHBo-COOH_core",
      "type": "XXX"
    }
  ], {
    "name": "ODNFAM-NH2_modifier",
    "type": "XXX"
  }, {
    "name": "EDC_coat",
    "type": "coating"
  }, {
    "name": "TFA_coat",
    "type": "XXX"
  ]
}], {
  "name": "SWNT-COOH_core",
  "type": [
    "core"
  ],
  "linkageType": "covalent",
  "linkage": [{
    "name": "linker_coat",
    "type": "XXX"
  }, {
    "name": "paraformaldehyde_coat",
    "type": "XXX"
  ]
}], {

```



Figure 7. Material schema excerpt (corresponds to ISA-Tab-Nano material file)

### 3.3.2 ISA-JSON EXPORT

The export to ISA-JSON requires mapping between the internal eNanoMapper data model and the ISA data model (Figure 8). The eNanoMapper data model is exemplified by its Java implementation and JSON serialization, described previously. The ISA data model is defined by the ISA-JSON v1 schemas. In order to provide a convenient means for working with the ISA data model, Java classes are generated from the ISA-JSON schema using <http://www.jsonschema2pojo.org/> and custom refactoring tools. The generated classes are available as an open source *ambit-export* package<sup>7,8</sup>.

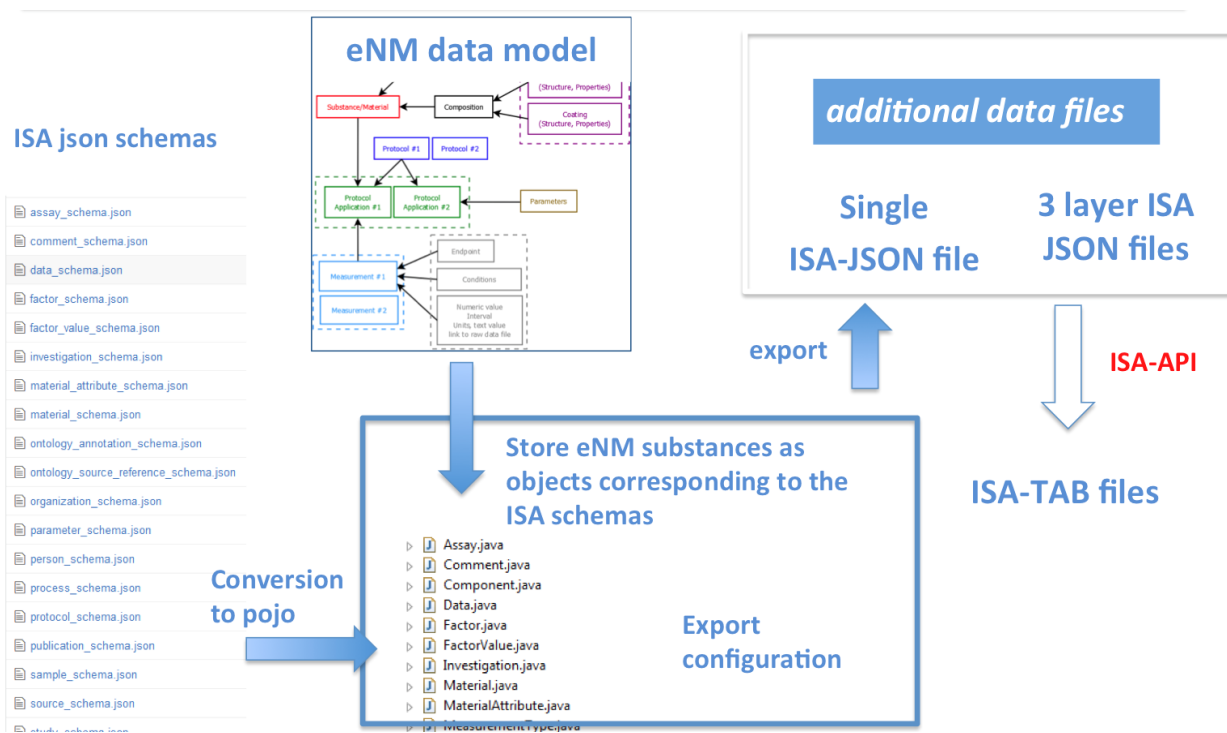


Figure 8. Overview of the format conversion workflow

<sup>7</sup> <https://svn.code.sf.net/p/ambit/code/trunk/ambit2-all/ambit2-apps/ambit2-export/> (code)

<sup>8</sup> <http://ambit.uni-plovdiv.bg:8083/nexus/#nexus-search;quick~ambit2%20export> (Maven artifact)

The main class used to export eNM data into ISA-JSON version 1 is class `ISAJsonExporter1_0` (Figure 9). An example usage is illustrated as follows:

```
ISAJsonExporter1_0 exporter = new ISAJsonExporter1_0(records, outputDir,
    exportConfig, endpointBundle);
exporter.export();
System.out.println(exporter.getResultAsJson());
```

The exporter is configurable since both the eNanoMapper data model and the ISA platform are complex and various mappings are possible. A default configuration is provided.

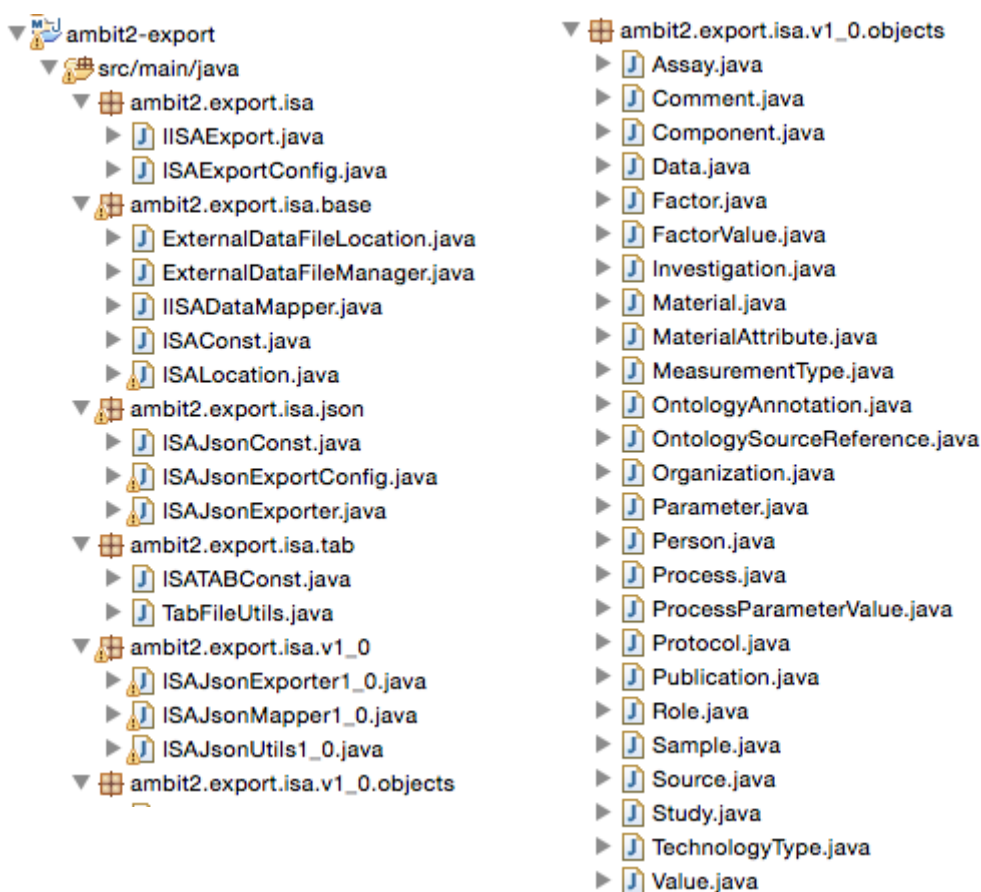


Figure 9. Java classes for representing ISA object and enabling export to ISA

The java code is based on a set of Java classes generated from ISA-JSON version 1 schema (Figure 9, right). The Java classes allow to conveniently represent the ISA in an object-oriented fashion where each JSON-schema has a corresponding java object (Figure 10).



Figure 10. Java classes representing ISA objects

### 3.4. CONFIGURABLE EXCEL PARSER

We have previously reported (D3.1, [1]) the development a configurable Excel parser (<https://github.com/enanmapper/nmdataparser>), enabling import of NanoSafety Cluster Excel templates. The configuration metadata for the parser is defined in a separate file, mapping the custom spreadsheet structure into the internal eNanoMapper storage components (Figure 11). The NMDDataParser uses a JSON configuration with a syntax that includes several basic JSON sections: DATA\_ACCESS, PARALLEL\_SHEETS, SUBSTANCE\_RECORD, PROTOCOL\_APPLICATIONS, EFFECTS, REPOSITORY.

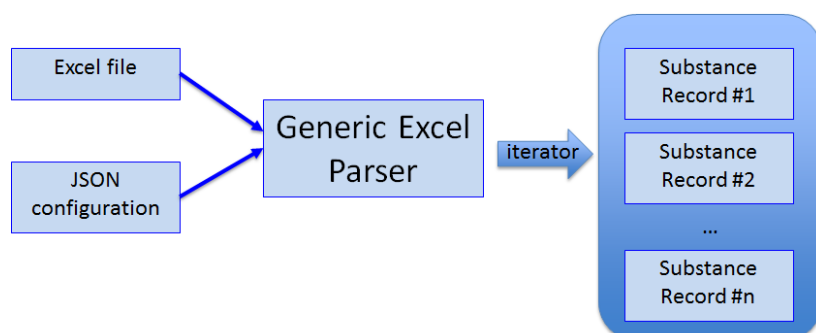


Figure 11. Overview of the Excel Parser supporting NanoSafety Cluster templates

The NMDDataParser functionality was extended to support block of effects. The typical study consists of measurements with varying experimental factors, as in the dose response data recorded in the FP7 MARINA project Excel templates and illustrated at Figure 12. The TNF- $\alpha$  (Tumor necrosis factor alpha [ng/ml]) assay is performed using 3 replicates and 8 concentrations of titanium dioxide.

	A	B	C	D	E	F	G	H	I	J	K	L
1	<b>TEST RESULTS</b>											
2												
3												
4		Replicate 1				Replicate 2				Replicate 3		
5		T1	Titanium Dioxide	Average (ng/ml)		T1	Titanium Dioxide	Average (ng/ml)		T1	Titanium Dioxide	Average (ng/ml)
6			0	1.8925			0	-0.3425			0	2.189
7			1	19.6985			1	5.9887			1	2.474
8			5	18.5207			5	5.7696			5	2.002
9			10	18.0280			10	5.3011			10	1.786
10			25	18.4153			25	5.1602			25	1.570
11			50	19.2965			50	5.2292			50	1.735
12			75	20.8867			75	6.1564			75	1.290
13			100	22.6964			100	7.2811			100	1.206
14												

Figure 12. Excel data templates with measurements grouped in blocks of values and corresponding conditions being varied.

The extension allows to describe the entire block and to avoid enumerating each combination of the condition values. A new JSON section EFFECT\_BLOCKS is introduced, which can be used to configure the simultaneous reading of many effects, grouped in blocks of measurement values according to the variations of the experimental factors.

```
"EFFECTS_BLOCK":
{
  "LOCATION":
  {
    "ITERATION": "ABSOLUTE_LOCATION",
    "IS_ARRAY": true,
```

```

        "TRIM_ARRAY": true,
        "SHEET_INDEX": 3,
        "COLUMN_INDEX": "B",
        "ROW_INDEX": 4
    },
    "ROW_SUBBLOCKS": "= TimePoints.size()",
    "COLUMN_SUBBLOCKS": "= Replicates",
    "SUBBLOCK_SIZE_ROWS": "= C.size() + 3",
    "SUBBLOCK_SIZE_COLUMNS": 4,
    "VALUE_GROUPS": [
    {
        "NAME": "=Assay_endpoint",
        "UNIT": "=Outcome_metric",
        "START_COLUMN": 3,
        "END_COLUMN": 3,
        "START_ROW": 3,
        "END_ROW": "=3 - 1 + C.size()",
        "PARAMETERS":
        [
            {
                "NAME": "Time point",
                "ASSIGN": "ASSIGN_TO_SUBBLOCK",
                "COLUMN_POS": 1,
                "ROW_POS": 2,
                "MAPPING": "Time",
                "UNIT": "h"
            },
            {
                "NAME": "Replicate",
                "ASSIGN": "ASSIGN_TO_SUBBLOCK",
                "COLUMN_POS": 1,
                "ROW_POS": 1
            },
            {
                "NAME": "Concentration",
                "ASSIGN": "ASSIGN_TO_VALUE",
                "COLUMN_POS": -1,
                "ROW_POS": 0,
                "UNIT": "µg/ml"
            }
        ]
    }
    ]
}

```

The LOCATION section defines the upper left corner of the effect block. The effect block is divided to a number of sub-blocks according to the keywords ROW\_SUBBLOCKS and COLUMN\_SUBBLOCKS. Taking into account the number of sub-blocks along both axes and the sub-block sizes, the size of the entire effect block is obtained. Within each sub-block, one or more groups of values are defined by means of section VALUE\_GROUPS. Accordingly each value group

spans several effect measurements where for each effect one or more conditions are varied via keyword PARAMETERS.

### 3.5 DATABASE CONTENT UPDATES

The initial eNanoMapper database content as published in D3.1 was further updated with *in vitro* data, generated by partner KI within the FP7 MARINA project data, as described in D5.4. The ModNanoTox data described in the eNanoMapper report “User application for importing NanoSafety Cluster data” (D5.4) is still not online due to the request by the former ModNanoTox team to delay the public access to this dataset until the data is published in the peer reviewed literature.

Since July 2015 the database content was updated with a new version of NanoWiki and the modelling dataset compiled by the FP7 MODENA project, which was publicly released after the end of the MODENA project. The version of the MODENA dataset, properly formatted and its JSON configuration is also available at the *nmdataparser* Github repository<sup>9</sup>. A subset of the MODENA dataset was used for the data uploading exercise used during the eNanoMapper hands on workshop held in Basel in Feb 2016.

NanoWiki [6] has been updated to release 2 [7]. This release adds some Linked Data features, the concept of bundles, materials from other data sets (Sigma-Aldrich, JRC Representational Materials, Crystallography Open Database, ArrayExpress). The data now describes 400 nanomaterials with over 800 measurements (physchem and bioassay data).

Datasets	Materials	Chemical structures
NanoWiki	389	74
OECD Harmonized Templates import test	1	1
Protein Corona dataset	121	65
In Vitro data from FP7 MARINA project (KI)	6	3
Sigma-Aldrich Products	7	1
JRC Representative Nanomaterials	51	7
Crystallography Open Database	6	6
ArrayExpress	9	3
MODENA	61	27

Table 1. Datasets (bundles) with statistics of data.enanomapper.net content

<sup>9</sup>

<https://github.com/enanomapper/nmdataparser/tree/master/src/test/resources/net/enanomapper/parser/xlsx/modena>

Links to omics data in ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>) have been added to the database, through the addition of metadata on a dataset on multi-walled carbon nanotubes [8]. In addition, physicochemical and cytotoxicity data from another related publication [9] will be added to the eNM database.

A new data transfer between FP7 NANoREG and eNanoMapper (WP3) was completed Jan 22, 2016. The data was provided by the Prof. Bengt Fadeel lab (Karolinska Institutet, NANoREG WP5), containing cytotoxicity assessment of the entire panel of NANoREG nanomaterials (19 materials from the JRC nanomaterial repository). The data transfer follows the current practice of starting a discussion on the data content (over email or teleconference) between eNanoMapper WP3 and the data provider. The purpose of this interaction is to exchange information on all metadata, which is not explicitly provided in the files transferred (i.e. protocol metadata). After the discussion, a JSON configuration accompanying the Excel input file is created and a test import is performed. The data provider is invited to comment. This step could be iterative, refining the JSON configuration file and possibly receiving new information by the data provider. In the case of the data provided by Prof. Bengt Fadeel, it will be shortly uploaded into the database, but will not be publicly accessible until permission is obtained by the data providers.



ENM eNanoMapper Search ▾ Nanomaterials ▾ OpenTox ▾ Help ▾

Bundles




























Title ▾	Version ▾	Code ▾	Status ▾	Owner ▾	Bundle ID ▾	Export
Sigma-Aldrich Products	1	NanoWiki	published	enanomapper	13e147c3-ccb5-11e5-946b-80ee7350bfa7	  
Protein Corona Fingerprinting Predicts the Cellular Interaction of Gold and Silver Nanoparticles	1	10.1021/nn406018q	published	enanomapper	00000000-0000-0000-0000-000000000003	  
OECD Harmonized Templates import test	1	Multi-Walled Carbon Nanotubes (MWCNT), synthetic graphite in tubular shape	published	enanomapper	00000000-0000-0000-0000-000000000002	  
NanoWiki	1	NanoWiki	published	enanomapper	00000000-0000-0000-0000-000000000001	  
MODENA	1	MODENA	published	enanomapper	62b0a66b-cd01-11e5-a67e-80ee7350bfa7	  
JRC Representative Nanomaterials	1	NanoWiki	published	enanomapper	646ab3a4-ccb5-11e5-946b-80ee7350bfa7	  
In Vitro data from FP7 MARINA project (KI)	1	Comprehensive In Vitro Toxicity Testing of a Panel of Representative Oxide Nanomaterials: First Steps towards an Intelligent Testing Strategy	published	enanomapper	91cad054-4de8-4dc9-a8e4-20564e7eada7	  
Crystallography Open Database	1	NanoWiki	published	enanomapper	804608b3-ccb5-11e5-946b-80ee7350bfa7	  
ArrayExpress	1	NanoWiki	published	enanomapper	ee0cdb4c-cce1-11e5-946b-80ee7350bfa7	  

Figure 13. Bundles overview

Figure 13 shows a screenshot of the data bundles currently online at <http://data.enanomapper.net>.



## 3.6 CLIENT SOFTWARE

### 3.6.1 JAVASCRIPT CLIENTS

To further demonstrate the use of the eNanoMapper API for query and visualisation we have developed a series of example web pages (HTML) using the JavaScript d3.js library<sup>10</sup>. This library has been used for a wide variety of visualisations (as can be seen on their website), and here used to summarize some of the data in the database. To simplify the interaction with the eNanoMapper API, a JavaScript client library, ambit.js, was written to allow asynchronous calls to the web service [10][10.5281/zenodo.16517]. However, because the d3.js methods require the data to be provided in a specific JavaScript object, the JSON returned by the API has to be converted to a structure understood by the d3.js library. The sources of the examples presented here are available from the ambit.js project page at <http://github.com/enanomapper/ambit.js/>. The source code and documentation of the ambit.js library are available at the same location.

The ambit.js JavaScript client was described in the [1] publication along with a number of visualization examples. A new example was created that combines one of the earlier examples with look up of ontology term labels separately as shown in Figure 14.

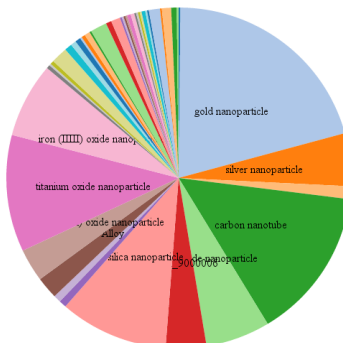


Figure 14. Output of one of the new ambit.js code examples , <http://enanomapper.github.io/ambit.js/example11.html>

The JavaScript visualization was further used. For the eNanoMapper workshop in Basel (Feb 2016), web pages were created to demonstrate the functionality. These pages give an overview of the amount of data available in the data.enanomapper.net instance. An example for titanium dioxide is given in Figure 15, which can be accessed at <http://enanomapper.github.io/enmSummaries/titaniumoxide.html>.

<sup>10</sup> <https://d3js.org/>

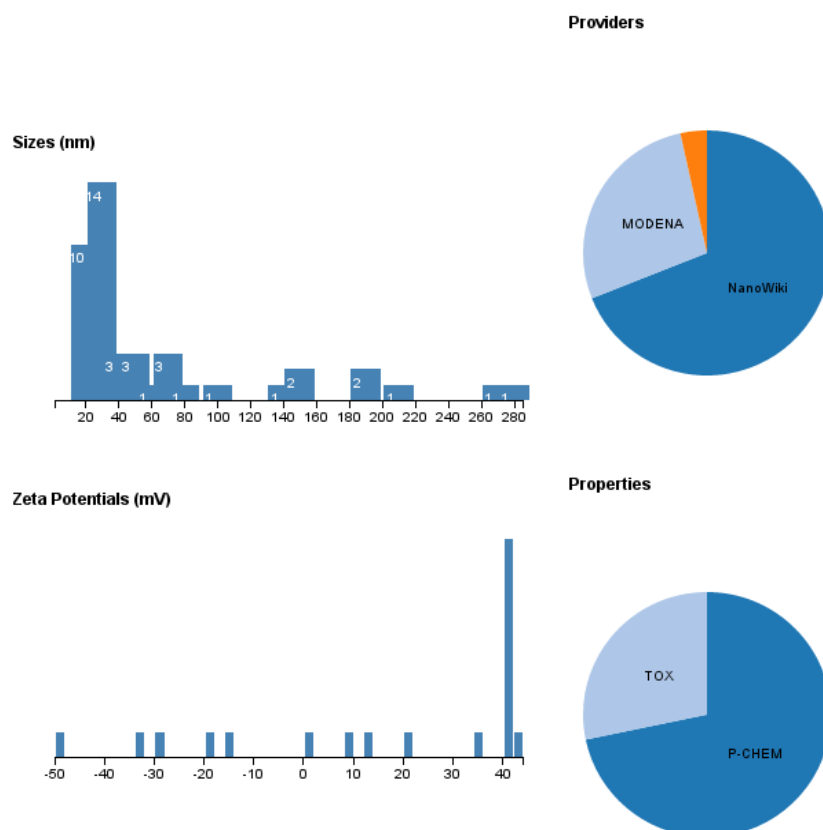


Figure 15. Webpage summarizing data in data.enanomapper.net for TiO<sub>2</sub> using the ambit.js client library and d3.js

### 3.6.2 R ENVIRONMENT

An R package has been developed, available from GitHub at <https://github.com/enanomapper/renm>, that allows querying an eNanoMapper data server, for example <http://data.enanomapper.net/> and downloading data on nanomaterials found in this data server. An example bit of R code looks like:

```
substances = listSubstances("http://data.enanomapper.net/")
info = substanceInfo(
  paste("http://data.enanomapper.net/substance/",
        "NWKI-71060af4-1613-35cf-95ee-2a039be0388a", sep="")
)
experiments = info$protocol
info$effects[[1]]
```

### 3.6.3. FERRET.AI

Ferret.ai is an independently developed Chrome browser add-on that uses the eNanoMapper API to query <http://data.enanomap.net/> for data from a particular publication, using the DOI of that publication. In the browser sidebar created by the add-on, queries like this are used to link out to databases that have data from that paper. Version 0.5.7 is the first to support the eNanoMapper API.



The screenshot shows the Beilstein Journal on Nanotechnology website. The article title is "The protein corona protects against size- and dose-dependent toxicity of amorphous silica nanoparticles" by Dominic Docter<sup>1</sup>, Christoph Bantz<sup>2</sup>, Dana Westmeier<sup>1</sup>, Hajo J. Galla<sup>3</sup>, Qiangbin Wang<sup>4</sup>, James C. Kirkpatrick<sup>5</sup>, Peter Nielsen<sup>6</sup>, Michael Maskos<sup>2</sup> and Roland H. Stauber<sup>1</sup>. The article is part of the Thematic Series "Biological responses to NPs". The FERRET add-on sidebar on the right shows the article's impact (2,326), authors (Roland H. Stauber), and a knowledge graph linking to eNanoMapper. The sidebar also lists cited works, including "Toxicity, genotoxicity and proinflammatory effects of amorphous nanosilica in the human intestinal Caco-2 cell line" and "No king without a crown--impact of the nanomaterial-protein corona on nanobiomedicine".

Figure 16. Webpage of a nano paper in the Beilstein Journal on Nanotechnology. The add-on creates a side bar on the right side, showing a link out to the eNanoMapper database, indicating seven of the materials from this paper are found in the database.

## 4. CONCLUSIONS

The eNanoMapper prototype database is publicly available at <http://data.enanomapper.net>. We have demonstrated the integration of data from multiple sources, using the common data model and API defined previously in the eNanoMapper report “Technical specification and initial implementation of the protocol and data management web services” (D3.1). New search functionalities have been implemented, including semantic search and free text search, supported by ontology annotated database entries. The database web application software is an extension of an existing open source project and available for download at <http://ambit.sourceforge.net/enanomapper.html>. The demonstration data provided by eNanoMapper partners illustrates the capability of the associated REST API to support a variety of tests and endpoints, recommended by the OECD Working Party of Manufactured Nanomaterials. We continue to maintain and extend the eNanoMapper database functionality and content, taking into consideration the feedback and new sets of data received. The data import supports a variety of NanoSafety cluster Excel templates via a configurable parser. ISA-Tab import and export is an ongoing activity, now based on a new version of the ISA specification, developed and extended within active collaboration with ISA-TAB and ISA-TAB-Nano users and developers. The next WP3 deliverables are: D3.3 (Modules and services for linking and integration with third party databases) and D3.4 (ISA-Tab templates for common bioselected set of data), due in M33 (end of October 2016). The work on both tasks is ongoing. The sets of templates will be based on existing Nano Safety cluster templates, selected with collaboration with respective template developers, and annotated with ontology terms by eNanoMapper and automatically converted to ISA-JSON.

## 5. BIBLIOGRAPHY

- [1] N. Jeliaskova, C. Chomenidis, P. Doganis, B. Fadeel, R. Grafström, B. Hardy, J. Hastings, M. Hegi, V. Jeliaskov, N. Kochev, P. Kohonen, C. R. Munteanu, H. Sarimveis, B. Smeets, P. Sopasakis, G. Tsiliki, D. Vorgrimmler, and E. Willighagen, "The eNanoMapper database for nanomaterial safety information," *Beilstein J. Nanotechnol.*, vol. 6, pp. 1609–1634, Jul. 2015.
- [2] S.-A. Sansone, P. Rocca-Serra, D. Field, E. Maguire, C. Taylor, O. Hofmann, H. Fang, S. Neumann, W. Tong, L. Amaral-Zettler, K. Begley, T. Booth, L. Bougueleret, G. Burns, B. Chapman, T. Clark, L.-A. Coleman, J. Copeland, S. Das, A. de Daruvar, P. de Matos, I. Dix, S. Edmunds, C. T. Evelo, M. J. Forster, P. Gaudet, J. Gilbert, C. Goble, J. L. Griffin, D. Jacob, J. Kleinjans, L. Harland, K. Haug, H. Hermjakob, S. J. H. Sui, A. Laederach, S. Liang, S. Marshall, A. McGrath, E. Merrill, D. Reilly, M. Roux, C. E. Shamu, C. A. Shang, C. Steinbeck, A. Trefethen, B. Williams-Jones, K. Wolstencroft, I. Xenarios, and W. Hide, "Toward interoperable bioscience data," *Nat. Genet.*, vol. 44, no. 2, pp. 121–126, Jan. 2012.
- [3] D. G. Thomas, S. Gaheen, S. L. Harper, M. Fritts, F. Klaessig, E. Hahn-Dantona, D. Paik, S. Pan, G. A. Stafford, E. T. Freund, J. D. Klemm, and N. A. Baker, "ISA-TAB-Nano: A Specification for Sharing Nanomaterial Research Data in Spreadsheet-based Format," *BMC Biotechnol.*, vol. 13, no. 1, p. 2, 2013.
- [4] R. L. Marchese Robinson, M. T. D. Cronin, A.-N. Richarz, and R. Rallo, "An ISA-TAB-Nano based data collection framework to support data-driven modelling of nanotoxicology," *Beilstein J. Nanotechnol.*, vol. 6, pp. 1978–1999, Oct. 2015.
- [5] N. Jeliaskova, P. Doganis, B. Fadeel, R. Grafstrom, J. Hastings, V. Jeliaskov, P. Kohonen, C. R. Munteanu, H. Sarimveis, B. Smeets, G. Tsiliki, D. Vorgrimmler, and E. Willighagen, "The first eNanoMapper prototype: A substance database to support safe-by-design," in *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2014, pp. 1–9.
- [6] E. Willighagen, "NanoWiki (release 1)," *figshare*, 2015. [Online]. Available: [https://figshare.com/articles/NanoWiki\\_release\\_1\\_/1330208](https://figshare.com/articles/NanoWiki_release_1_/1330208).
- [7] E. Willighagen, "NanoWiki (release 2)," *figshare*, 2016. [Online]. Available: [https://figshare.com/articles/NanoWiki\\_release\\_2\\_/2075347/1](https://figshare.com/articles/NanoWiki_release_2_/2075347/1).
- [8] P. Nymark, P. Wijshoff, R. Cavill, M. van Herwijnen, M. L. J. Coonen, S. Claessen, J. Catalán, H. Norppa, J. C. S. Kleinjans, and J. J. Briedé, "Extensive temporal transcriptome and microRNA analyses identify molecular mechanisms underlying mitochondrial dysfunction induced by multi-walled carbon nanotubes in human lung cells," *Nanotoxicology*, vol. 9, no. 5, pp. 624–635, Jul. 2015.
- [9] P. Nymark, J. Alstrup, S. Suhonen, Y. Kembouche, M. Vippola, J. Kleinjans, J. Catalán, H. Norppa, J. van Delft, and J. Briedé, "Free radical scavenging and formation by multi-walled carbon nanotubes in cell free conditions and in human bronchial epithelial cells," *Part. Fibre Toxicol.*, vol. 11, no. 1, p. 4, 2014.
- [10] E. Willighagen, "ambit.js," 2015. [Online]. Available: <https://zenodo.org/record/16517?ln=en#.Vtrmxvl97Dc>.