



DELIVERABLE REPORT D3.4

ISA-Tab templates for selected set of common bioassays

Deliverable D3.4

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:	IDEA



DATE:	21.11.2016
VERSION:	V.1.0.

Call identifier	FP7-NMP-2013-SMALL-7
------------------------	-----------------------------

Document Type	Deliverable Report
WP/Task	WP3/T3.5
Document ID	eNanoMapper D3.4
Status	Final

Partner Organisations	<ul style="list-style-type: none"> ● Douglas Connect, GmbH (DC) ● National Technical University of Athens (NTUA) ● In Silico Toxicology (IST) ● Ideaconsult (IDEA) ● Karolinska Institutet (KI) ● European Bioinformatics Institute (EMBL-EBI) ● Maastricht University (UM) ● Misvik Biology (MB)
Authors	<p>Nina Jeliaskova (IDEA) Nikolay Kochev (IDEA) Vedrin Jeliaskov (IDEA) Penny Nymark (MB) Pekka Kohonen (MB)</p> <p>Reviewed by Barry Hardy (DC)</p>
Purpose of the Document	To report on the Task 3.5

Document History

1. Table of Contents, 10/10/2016
2. First draft, 31/10/2016
3. Second draft, 14/11/2016
4. Final draft, 21/11/2016
5. Completed version, 01/12/2016

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY	7
2. INTRODUCTION	8
3. INVESTIGATION-STUDY-ASSAY	10
3.1 ISA-JSON NANO EXTENSION (NANO MATERIAL JSON SCHEMA)	11
3.2 ISA-JSON EXPORT	13
3.2.1 ISA-JSON implementation	14
4. SPREADSHEET DATA TEMPLATES	17
4.1 "ISA-TAB LOGIC" TEMPLATES	17
4.1.1 Configuration files for the NANoREG templates	20
4.2. IOM TEMPLATES	23
5. NANOREG DATA TRANSFER	28
6. CONCLUSION	31
7. BIBLIOGRAPHY	32
ANNEX 1. NANoREG EXCEL FILES OVERVIEW	33

TABLE OF FIGURES

Figure 1. The common eNanoMapper data model enables conversions between different formats.	9
Figure 2. eNanoMapper isa-api fork at GitHub repository. The new JSON-schema for nanomaterial description.	11
Figure 3. The JSON schema for Material/Substance consists of the basic nominal nanomaterial characteristic and array of material components.....	12
Figure 4. JSON schema for material component. The component is represented by basic description and an array of linkages to other components.....	13
Figure 5. eNanoMapper web page with download links in different formats	14
Figure 6. ISA export workflow: (i) ISA-JSON schemas and Material schema are converted to Java classes, (ii) eNanoMapper data records are mapped into ISA model; (iii) data is serialized as a ISA-JSON file and a data file.	14
Figure 7 Mapping between eNanoMapper and ISA data models	16
Figure 8. The NANOREG template for endpoint cell viability , MTS assay	21
Figure 9 The data templates documentation menu at https://search.data.enanmapper.net	22
Figure 10 Documentation site providing templates and configuration files download	22
Figure 11 IOM template: Test conditions sheet	23
Figure 12 IOM template: Raw data sheet	25
Figure 13 IOM template: Test results sheet	26
Figure 14 IOM template: Test summary sheet	26
Figure 15 NANOREG database application front page	28
Figure 16 Search application with NANOREG data	29
Figure 17 NANOREG data availability	30

GLOSSARY

Abbreviation / acronym	Description
SPARQL	SPARQL Protocol and RDF Query Language) is an RDF query language, that is, a semantic query language for databases
RDF	Resource Description Framework (RDF) https://www.w3.org/RDF/ https://www.w3.org/RDF/
ISA-TAB	Built around the 'Investigation' (the project context), 'Study' (a unit of research) and 'Assay' (analytical measurement) data model and serializations (tabular, JSON and RDF). http://isa-tools.org/
ISA-JSON	http://isa-tools.org/2016/10/release-of-the-isa-specs/
ISA-TAB-NANO	ISA-TAB-Nano extends ISA-TAB https://wiki.nci.nih.gov/display/icr/isa-tab-nano
JSON	Javascript Object Notation http://www.json.org/
NANoREG	FP7 NANoREG – A common European approach to the regulatory testing of Manufactured Nanomaterials, grant agreement 310584
NANoREG templates	http://www.nanoreg.eu/media-and-downloads/templates
NSC, NanoSafety Cluster	http://www.nanosafetycluster.eu/

1. EXECUTIVE SUMMARY

Data preparation is an activity pertinent to most of the NanoSafety Cluster (NSC) projects. The importance of standardized means for data logging and data sharing is acknowledged by both data providers and data users, but the current practice consists mostly of disparate efforts, resulting in incompatible files. The ISA-TAB (Investigation-Study-Assay) and ISA-TAB-Nano formats are cited in many projects, but not used due to several reasons, e.g. lack of user friendliness or the steep learning curve needed to adopt the format. Instead, variations of the ISA have appeared, such as the ISA-TAB-Logic templates, developed and released under open license by the NANoREG project. The ISA-TAB-Logic templates are Excel files, with a number of fields describing the material sample, physchem characterisation and bioassays. Various flavours of Excel templates are used or considered for adoption by several NSC projects. We have cleaned, annotated and created configuration files for the publicly released templates, enabling direct import into the database through a web browser or programmatically. Ontology annotation is ongoing, with close collaboration with eNanoMapper WP2. We developed tools to automatically generate ISA-JSONv1 (the new specification of ISA) from existing nanomaterial safety data. We also developed *Material Schema extension* for the new ISA-JSON specification, which is the counterpart of the “material file” in the ISA-Tab-Nano specification. To allow for easy adoption, we have created a documentation site, describing the developments above, accessible via the eNanoMapper tutorials repository at Github <https://github.com/enanomapper/tutorials/tree/master/DataTemplates> and via menu “Data templates” of the search application at <https://search.data.enanomapper.net>

Being able to read the Excel spreadsheets and write the internal data model into ISA-JSON files accomplishes the goal of automatically generating the ISA files, and enables exporting query results from the database in any desired format. The templates provided are coupled with configuration files, enabling import into the database through a simple web form or programmatically via Application Programming Interface. The use of the tools and templates described in this deliverable is also foreseen within the NSC projects enabling harmonized collection of data, which will eventually be compatible across projects.

2. INTRODUCTION

ISA (Investigation-Study-Assay)¹[1] 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. ISA-Tab is the legacy format, relying on tab delimited files. During the last year a new serialisation was designed and implemented by the same team at University of Oxford, namely ISA-JSON, relying on JSON files (JavaScript Object Notation). The ISA-JSON release (October 2016) includes a JSON schema and an ecosystem of tools used for creating, validating and visualizing documents. 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) and an attempt to extend the ISA-Tab format by introducing a separate file for describing the (nano)material components. The ISA-Tab-Nano is documented in a publication [2] and in the US Nano WG wiki², which include example spreadsheets, but no tools to parse the files and to enforce the specification. For this reason, the practical use of ISA-Tab-Nano is not straightforward, as demonstrated by the efforts of the FP7 NanoPuzzles project [3] and the introduction of "ISA-Tab-logic" templates by the FP7 NANoREG project.

With the previous experience of formatting experimental data as ISA-Tab files manually being very time consuming [4], we have been exploring various solutions to improve the user friendliness. Taking into account the observation that the majority of EU NanoSafety Cluster projects prefer to prepare their experimental data using custom spreadsheet templates, we have previously reported the pragmatic approach of the eNanoMapper WP3 team, implementing a configurable Excel parser, able to parse diverse spreadsheets into the same internal data model and export the data from this data model into different formats [5]. This allows us to provide format converters (Figure 1).

¹ <http://isa-tools.org/>

² <https://wiki.nci.nih.gov/display/icr/isa-tab-nano>

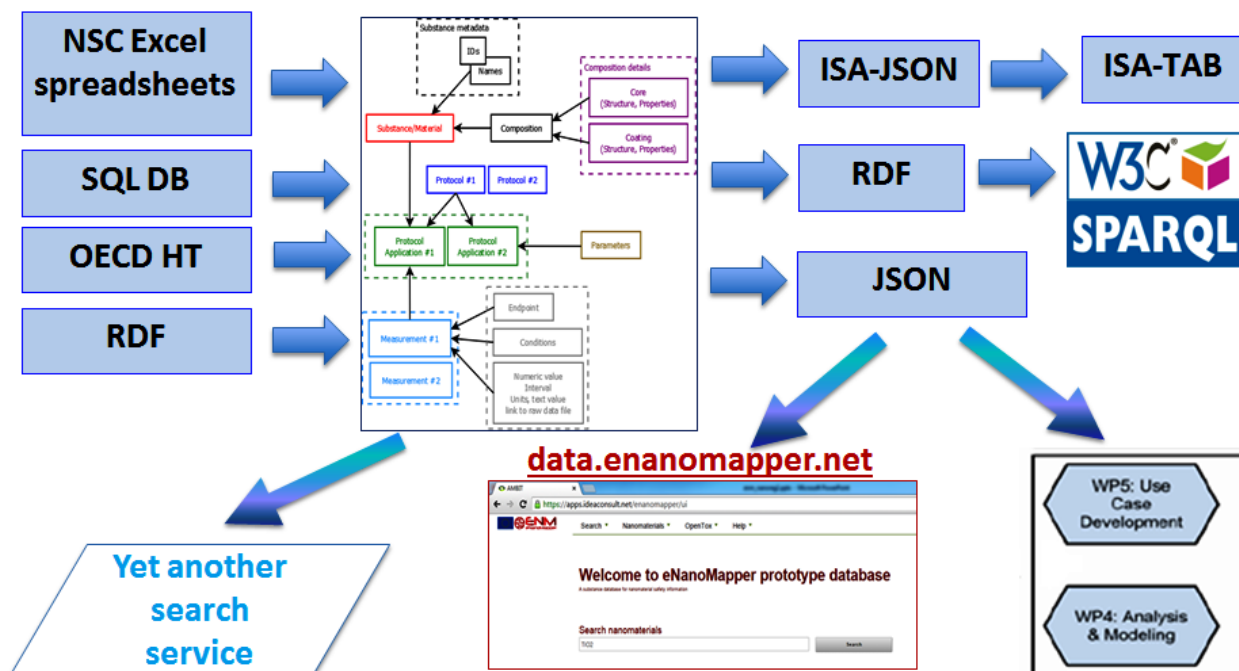


Figure 1. The common eNanMapper data model enables conversions between different formats.

The new ISA-JSON specification is designed around the concept of “core” ISA schema and “extensions”. It is expected that the University of Oxford team develops and maintains the core, while different communities develop the extensions. The support for nanomaterials fits very well in this scheme as an extension, and here we report the development of a (nano)material extension for ISA-JSONv1. We have implemented tools for converting the internal model of the eNanMapper database into ISA-JSONv1 and provide tools for generating ISA-JSON files. If needed, the ISA-JSON files can be translated into legacy ISA-TAB via the tools provided by the ISA team.

Finally, we report a large effort of cleaning, configuring and providing Excel spreadsheet templates for data entry, based on the publicly released NANOREG templates. These templates are now coupled with JSON configuration files, enabling direct import into the database and are ready to use by the data providers. The templates and related documentation is available at <http://ambit.sf.net/enanmapper/templates>. The review of the content of the templates (whether the fields included are all necessary and sufficient to describe the experiment) is not a subject of this deliverable, but was initiated by several efforts and meetings on NanoSafety Cluster level as well as a dedicated task in the NanoSafety Cluster Working group 4 (Databases) and discussions on the same topics in various projects, considering the use of the templates.

3. INVESTIGATION-STUDY-ASSAY

ISA (Investigation-Study-Assay) 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 extension³, 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).

The ISA group started migration in 2015 to a new JSON format, instead of tab delimited files as more flexible and efficient way to store the experimental meta-data. Currently, there are ISA v1 and ISA v2 JSON versions. The ISA v2 is based solely on JSON, where the backward compatibility with the ISA v1 and ISA-TAB format is preserved via ISA-tools project utilities.

The ISA team develops and maintains documentation⁴ and tools⁵ supporting the new ISA ISA-JSON formats. 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. The JSON schemas describe various components of the 3 layers of ISA platform (i.e. investigation, study, assay, sample, source, ontology annotation, etc.). The schemas are available at GitHub:

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

The ISA-JSON specification was officially released at the end of October 2016 <http://isa-tools.org/2016/10/release-of-the-isa-specs/> [6]

³ <https://wiki.nci.nih.gov/display/icr/isa-tab-nano>

⁴ <https://isatools.readthedocs.io/en/latest/>

⁵ <https://github.com/ISA-tools/isa-api>

3.1 ISA-JSON NANO EXTENSION (NANO MATERIAL JSON SCHEMA)

We have developed a Substance/Material JSON schema which is a nanomaterial extension of ISA-JSON v1, the counterpart of the ISA-Tab-Nano format. The schema is available at the *enanomapper/isa-api* fork at GitHub in folder “.../isa_model_version_1_0_schemas/material”, a new subfolder of the folder containing the core ISA schema (Figure 2):

https://github.com/enanomapper/isa-api/tree/master/isatools/schemas/isa_model_version_1_0_schemas/material

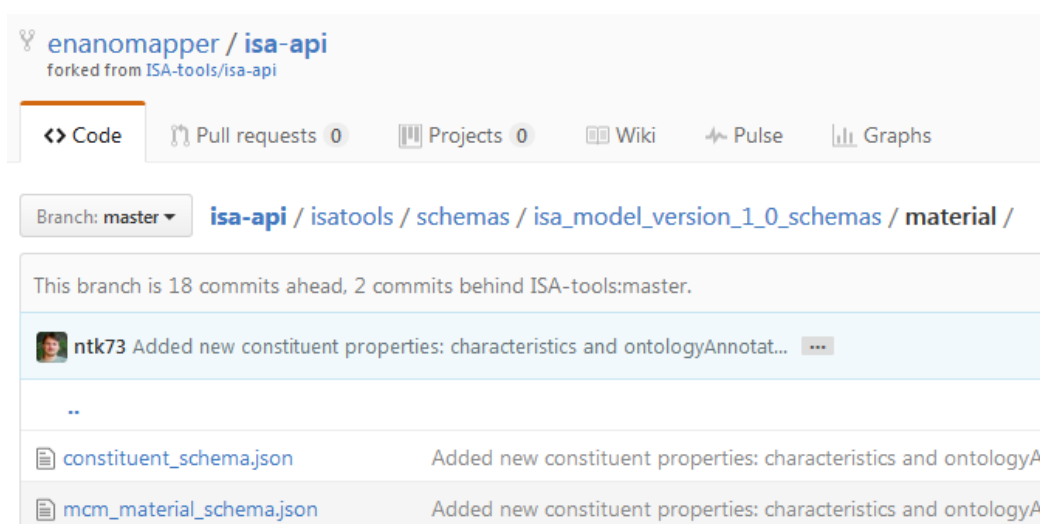


Figure 2. eNanoMapper isa-api fork at GitHub repository. The new JSON-schema for nanomaterial description.

The main JSON schema ([mcm_material_schema.json](#)) defines a material (substance) with several basic properties: ``id``, ``lotIdentifier``, ``name`` etc (Figure 3). The material composition is an array of constituents.

```
$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}
  ► constituents {2}
  ► description {1}
  ► synthesis {1}
  ► designRationale {1}
  ► intendedApplication {2}
  ► characteristics {2}
  ► mcmType {2}
  ► chemicalName {2}
  ► dataFiles {2}
```

Figure 3. The JSON schema for Material/Substance consists of the basic nominal nanomaterial characteristic and array of material components

A separate helper JSON schema ([constituent_schema.json](#)) is implemented for definition of all components of the nanomaterial (Figure 4). The composition of a nanomaterial may contain one or several components. Each component has a role (core, coating, etc.) and linkages to other constituents. The linkage describes the relation between two components. For example, two components may be covalently bonded, one being embedded or encapsulated within another constituent etc.

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

Figure 4. JSON schema for material component. The component is represented by basic description and an array of linkages to other components

3.2 ISA-JSON EXPORT

The export to ISA-JSONv1 is enabled for each data collection (<https://data.enanomapper.net/bundle>) in the eNanoMapper database, indicated by the ISA icon (Figure 5).

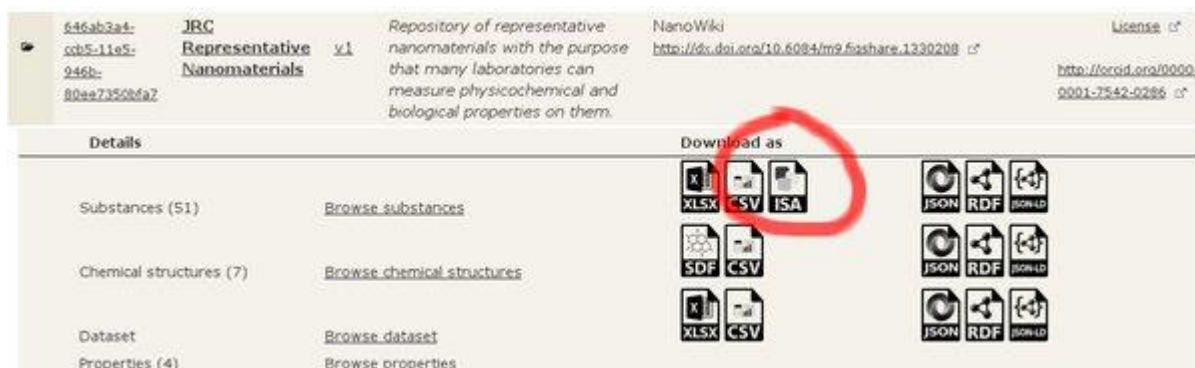


Figure 5. eNanoMapper web page with download links in different formats

Converting files into ISA-JSON format is also possible without import into the database. More information on how to convert input file in supported formats into ISA-JSON or RDF is available at <http://ambit.sourceforge.net/enanomapper/templates/convertor.html>

ISA-JSON export software tools are available at:

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

3.2.1 ISA-JSON IMPLEMENTATION

The initial support for exporting data into ISA-JSON format was previously reported in D3.2. In the reporting period we continued the improvement of the ISA-JSON export library and developed new features. The basic workflow for the ISA-JSON export is shown in Figure 6.

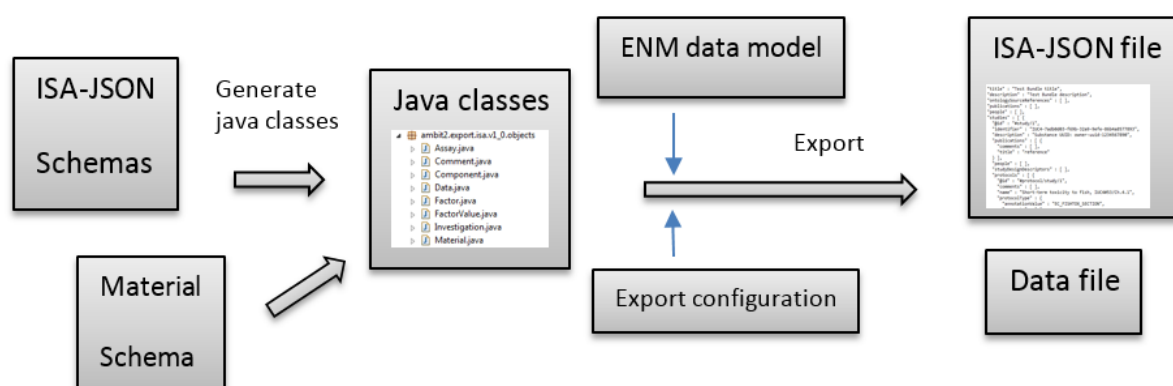


Figure 6. ISA export workflow: (i) ISA-JSON schemas and Material schema are converted to Java classes, (ii) eNanoMapper data records are mapped into ISA model; (iii) data is serialized as a ISA-JSON file and a data file.

The ISA data model is defined by JSON schemas which describe various components of the 3 layer of the ISA platform (i.e. investigation, study, assay, as well as sample, source, ontology annotation, etc.).

On the basis of ISA-JSON schemas we generate Java classes, implementing the ISA data model (see Figure 6). The Java objects are filled with the data either from the eNanoMapper database or from input files. These objects are serialized as JSON files (the actual ISA-JSON format files). During the first implementation, reported in D3.2, we used the jsonschema2pojo software tool (<http://www.jsonschema2pojo.org/>) to generate java classes from ISA-JSON schemas. With the constant updates and increasing complexity of the schemas the jsonschema2pojo tool turned out to be inefficient for the latest schema versions. In response, we developed our own tool for java class generation⁶. Additionally, the utilities for storage of the actual measurements data were improved (ExternalDataFileManager, the extra layer to the ISA platform).

⁶ <https://svn.code.sf.net/p/ambit/code/trunk/ambit2-all/ambit2-apps/ambit2-export/src/test/java/ambit2/export/isa/codeutils/>

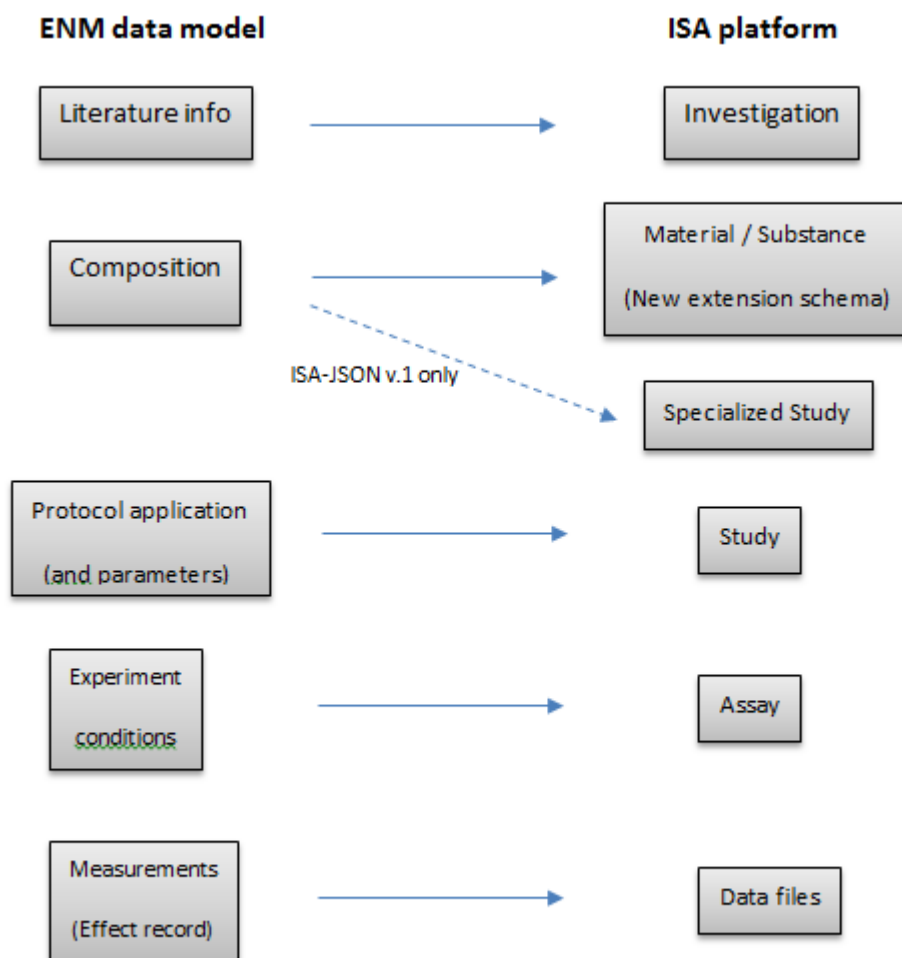


Figure 7 Mapping between eNanoMapper and ISA data models

Figure 7 summarizes the basic mapping between eNanoMapper data model and ISA platform. The correspondence between the three major layers of ISA platform to the eNanoMapper data model was established. Recently the ISA team announced the new isa-api validation and conversion tools which are currently being used for testing the generated by eNanoMapper ISA-JSON files. The isa-api tools allow converting ISA-JSON files into legacy ISA-TAB files and vice versa. Being able to write the internal data model into ISA-JSON files accomplishes the goal of automatically generating the ISA files, and enables exporting query results from the database in a standard and validated ISA format. The ISA-JSON v1 files can be converted to the legacy ISA-TAB files via the Python tools available at <http://isa-tools.org/>.

4. SPREADSHEET DATA TEMPLATES

With previous experience of formatting experimental data as ISA-TAB files manually being very time consuming, we have been exploring various solutions to improve the user friendliness of data preparation. Taking into account the observation that the majority of EU NanoSafety Cluster projects prefer to prepare their experimental data using custom spreadsheet templates, the eNanoMapper WP3 developed tools to directly read the spreadsheets and import into the database and/or convert into ISA-JSON or semantic formats.

The development of the configurable Excel parser enabling import of NanoSafety Cluster Excel templates was reported previously. The parser is embedded in the web application, allowing to upload NSC spreadsheets through a web form. The tool was improved and enhanced with new features. A command line application enabling file conversion without importing into the database was also developed. The configuration metadata for the parser is defined in a separate file, mapping the custom spreadsheet structure into the internal eNanoMapper storage components. Details on the configuration syntax and a user guide with examples is available at <http://ambit.sourceforge.net/enanomapper/templates/convertor.html>.

4.1 “ISA-TAB LOGIC” TEMPLATES

The NANoREG project released a large set of Excel templates under open license, available at <http://www.nanoreg.eu/media-and-downloads/templates>. As of October 2016, there are templates for 31 phys-chem, 27 in-vitro and 35 in-vivo assays. While not strictly following the ISA-TAB and ISA-TAB-Nano formats, the NANoREG templates have been designed around *ISA-Tab-logic*, i.e. structuring the data in investigation-study-assay related groups. The Excel format was selected to ensure user friendly means of data logging for the NANoREG project partners. The NANoREG templates are organized as one spreadsheet per assay, multiple NM can be entered as rows. One Excel file may contain more than one assay, measuring the same endpoint. The metadata is organised in several groups:

Sample information: Contains information about the NM (including names, ID, supplier, vial number and replicate number, as well as dispersant). The reporting organisation, operator and date of the experiment are also in this section.

Module (phys-chem, in-vitro or in-vivo), the endpoint (e.g. cell viability), and the assay name (e.g. “Alamar blue”).

Method and instrument information: Any parameter describing the experiment, including cell lines, instrument, controls, time points, concentrations. These differ widely across different experiments. This section includes a subgroup “size distribution”, providing placeholders for size distribution measured for the sample (including details on the dispersion protocol and dispersion medium). These fields are (almost) constant across all templates.

Results: Several columns to specify measurement outcomes, along with uncertainty.

SOP: (reference to the protocol)

Statistics of the number of fields, terms and assays is presented in Table 1.

Table 1 NANoREG template statistics

Excel file	Max number of fields per template	Number of terms	Number of assay templates
INVITRO_BARRIERCROSSING.xlsx	66	535	9
INVITRO_CTA.xlsx	45	53	1
INVITRO_CYTOTOXICITY.xlsx	45	197	4
INVITRO_GENOTOXICITY.xlsx	51	175	4
INVITRO_IMMUNOTOX.xlsx	50	151	3
INVITRO_NMDEPOSITION.xlsx	36	128	3
INVITRO_OCCULARIRRITATION.xlsx	41	44	1
INVITRO_VIABILITY.xlsx	73	492	9
INVIVO_ALI.xlsx	32	71	2
INVIVO_BAL.xlsx	57	153	3
INVIVO_BIODISTRIBUTION.xlsx	39	44	1
INVIVO_ECOTOX.xlsx	56	223	4
INVIVO_GENOTOX.xlsx	51	175	4
INVIVO_IMMUNOTOX.xlsx	59	384	7
INVIVO_INHALATION.xlsx	34	153	4

INVIVO_ORGANBURDEN.xlsx	50	58	1
INVIVO_PATTERNDISTRIBUTION.xlsx	55	168	3
INVIVO_PBPBPK.xlsx	48	58	1
INVIVO_PULMONARY.xlsx	46	203	5
INVIVO_TOXLONGTERM.xlsx	37	164	4
PCHEM_BATCHDISPERSION.xlsx	34	38	1
PCHEM_COMPOSITIONCHEMICAL.xlsx	55	187	4
PCHEM_CRYSTALLINEPHASE.xlsx	28	66	2
PCHEM_DENSITY.xlsx	28	64	2
PCHEM_DUSTINESS.xlsx	44	217	4
PCHEM_IEP.xlsx	42	46	1
PCHEM_POTENTIOMETRY.xlsx	36	43	1
PCHEM_SIZE.xlsx	82	352	7
PCHEM_SOLUBILITY.xlsx	49	140	3
PCHEM_SURFACECHEMISTRY.xlsx	31	137	4
PCHEM_VSSA.xlsx	34	40	1
PCHEM_ZETAPOTENTIAL.xlsx	40	45	1
Grand Total	82	5004	5004

The fields in each template are grouped in several sections, described above (e.g. sample information, size distribution). The list of all sections is presented in Table 2. The section “sample information” is expanded and lists individual fields.

Table 2 Groups of fields in NANoREG templates

Group of fields	Number of fields
analytical parameters - final exposure media	8
analytical parameters - initial exposure media	9

endpoint_assay	311
header0	421
header1	114
INVALID	2
method and instrument information	1328
results	577
sample information	1454
size distribution	671
size distribution - quality parameters	19
sop	61
(blank)	25
Total	5000

4.1.1 CONFIGURATION FILES FOR THE NANOREG TEMPLATES

Within the eNanoMapper project (WP3 Database) the templates are analysed, cleaned and regenerated (Figure 8). The original templates are created manually and over a long time (the duration of the NANOREG project) and contain minor errors like typos, slightly different naming for the same type of field, as well as units merged with the field name. With automatic extraction of the fields from all templates it became possible to fix these minor errors (semi-automatically), write units into a separate row and synchronise the naming across the templates. Additionally, for some fields expecting predefined answers (e.g. yes/no or phys-chem/in-vitro/in-vivo for the “module” field) the regenerated templates provide pickup lists. The pickup lists could be extended to other fields, as long as we progress with ontology annotation and identification of the fields which can benefit from a list of predefined ontology terms.

Last but not least, the JSON configurations can now be generated together with the templates, based on list of extracted and annotated fields. The JSON configurations are essential in order to enable import into the database.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	SAMPLE INFORMATION												
2	Replicate number	NM id code	NM chemistry (core)	Cas number	Vial number	Supplier	Material state	Use of dispersant	Dispersant reference	Sample name	Reporting organisation	Operator	Date of preparation
3													
4													
5													
6													
	N	O	P	Q	R	S	T	U	V	W	X		
1	METHOD AND INSTRUMENT INFORMATION												
2	Date of analysis	Module	Endpoint	Assay name	Dispersion protocol	Size distribution	Size distribution analysis technique	Dispersion medium	Concentration	Incubation time	Mean hydrodynamic diameter	Associated uncertainty (sd)	
3									mg/ml	h	nm		
4													
5													
6													
	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH			
1													
2	Associated uncertainty (sd)				Cell type	Pass number	Culture medium composition	Plate	Volume of incubation medium/well	Seeding	Number of technical replicates per conditions		
3													
4													
5													
6													
	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	
1	RESULTS												SOP
2	Number of technical replicates	Calculated exposure	Exposure time	Positive	Positive control test	IC50 (positive control)	Nps tested concentration	Corresponding dose/surface	Percentage of viability	nps tSd	IC50	References to SOPs	
3													
4													
5		µg/cm ²	h		µm		µg/ml	µg/cm ²			µg/ml		
6													

Figure 8. The NANoREG template for endpoint cell viability , MTS assay

The clean templates and JSON configuration files are available at GitHub repository:

<https://github.com/enanomapper/nmdataparser/tree/master/enmconvertor/src/site/resources/templates>

The fields are being matched with ontologies and annotated with the help of WP2 partners (several issues at <https://github.com/enanomapper/ontologies/issues>). Once the annotation is completed, we will re-generate the templates to include ontology terms.

A user friendly documentation is developed and is accessible via the eNanoMapper tutorials repository at Github:

<https://github.com/enanomapper/tutorials/tree/master/DataTemplates>

and as a menu “Data templates” of the search application at <https://search.data.enanomapper.net/> (Figure 9).

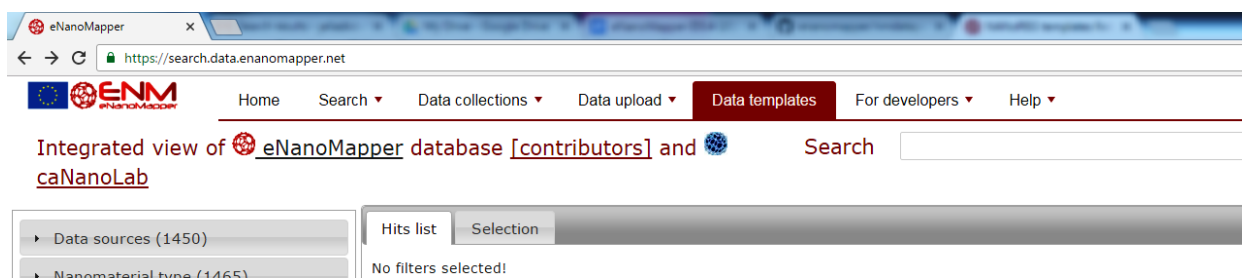


Figure 9 The data templates documentation menu at <https://search.data.enanmapper.net>

The documentation web site (Figure 10) consists of sections for downloading the templates and configuration files for Physicochemical characterisation, In-vitro assays and in-vivo assays, description and download links for the ISA-JSON material schema and converters as well as [interactive display](#) of all fields, cleaned values and annotations.

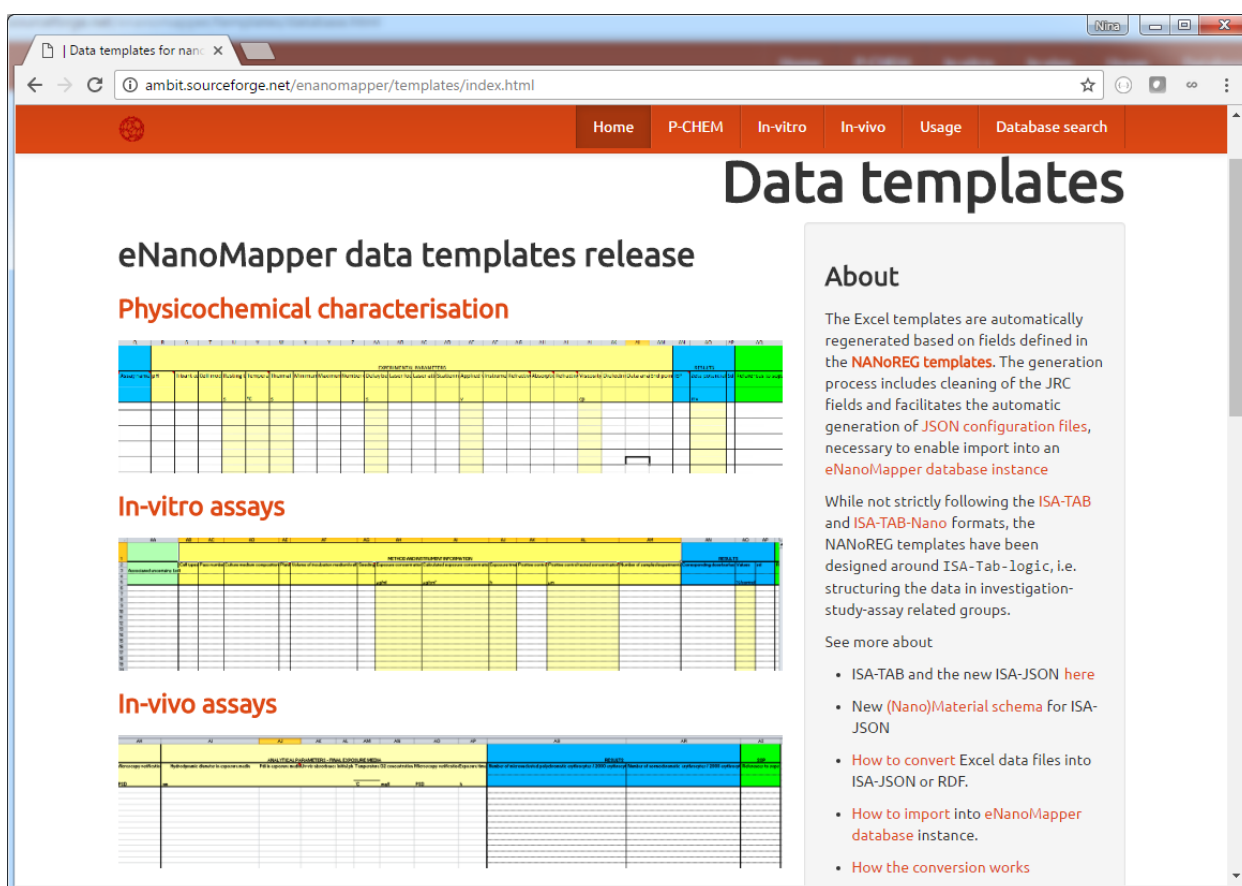


Figure 10 Documentation site providing templates and configuration files download

4.2. IOM TEMPLATES

A different type of template, developed by the Institute of Occupational Medicine (IOM) (<http://www.iom-world.org/>) has been used by a number of NSC projects (NANOMMUNE, NANOTEST, ENPRA, MARINA, NANOSOLUTIONS). We reported previously (D3.2) an implementation of NMDataParser features for support of Excel sheets with data blocks (i.e. new JSON section EFFECT_BLOCKS). The import of IOM templates into eNM database is based on EFFECT_BLOCKS syntax. IOM templates contain four major Excel sheets. The first sheet includes information about the sample and the metadata of the experiment (see Figure 11). Most of the data from the first sheet is imported using the standard JSON configurations which define a mapping between spreadsheet fields and eNM database.

38	Dispersion agent:	water with 0,05% BSA							
39									
40	Aids used to disperse - Y / N:	Sonication:	Y	Vortexing:	Y	Stirring:	N		
41									
42	Treatment concentration series (C) (µg/ml):	C1	C2	C3	C4	C5	C6	C7	C8
43	Alter or add as necessary	0	1	5	10	25	50	75	100
44									
45	CELL LINE/TYPE								
46									
47	Short-Name:	HMDM							
48	Full specific name (note any line variants or related IDs):	Primary human monocyte-derived macrophages (HMDM)							
49	Supplier:	Karolinska University Hospital, Stockholm, Sweden - buffy coats from healthy adult blood donors							
50									
51	CELL CULTURE CONDITIONS								
52									
53	Medium (Supplier/Lot No.):	RPMI (Sigma, R0833)							
54	Serum (Supplier/Lot No.):	FBS (Gibco, 10500)							
55	Serum concentration in growth medium:	10%							
56	Serum concentration in treatment medium:	0,05% Bovine Serum Albumin (Sigma, A2058)							
57									
58	TIMELINE								
59									
60	Time points (hours - or specify any other units):	T1	T2						
61	Alter or add as necessary	6	24						
62									

Figure 11 IOM template: Test conditions sheet

Some data fields from the IOM template are defined as dynamic arrays with varying length e.g. concentration series and time points as shown in Figure 11 (rows 43 and 61 respectively). We have implemented additional JSON sections VARIABLES and VARIABLE_MAPPINGS to support dynamic length data structures. Section VARIABLES defines single value variables and variables of type array both used in EFFECT_BLOCKS sections.

```

"VARIABLES": {
  "C": {
    "IS_ARRAY": true,
    "TRIM_ARRAY": true,
    "ITERATION": "ABSOLUTE_LOCATION",
    "SHEET_INDEX": 1,
    "COLUMN_INDEX": "B",
    "COLUMN_INDICES": [
      2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
    ]
  }
}

```

```

    ],
    "ROW_INDEX": 43
  },
  "TimeDesignations": {
    "IS_ARRAY": true,
    "TRIM_ARRAY": true,
    "ITERATION": "ABSOLUTE_LOCATION",
    "SHEET_INDEX": 1,
    "COLUMN_INDEX": "B",
    "COLUMN_INDICES": [
      2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
    ],
    "ROW_INDEX": 60
  },
  "TimePoints": {
    "IS_ARRAY": true,
    "TRIM_ARRAY": true,
    "ITERATION": "ABSOLUTE_LOCATION",
    "SHEET_INDEX": 1,
    "COLUMN_INDEX": "B",
    "COLUMN_INDICES": [
      2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
    ],
    "ROW_INDEX": 61
  }
},
"VARIABLE_MAPPINGS": [
  {
    "NAME": "Time",
    "KEYS_VARIABLE": "TimeDesignations",
    "VALUES_VARIABLE": "TimePoints"
  }
]

```

The time points (T) and concentrations (C) are defined in the first sheet and only referred to in the data sheets. This is supported by the VARIABLE_MAPPINGS section in the configuration, e.g. TimeDesignations are mapped onto TimePoints thus in the database values 6 and 24 are stored instead of “T1” and “T2”.

The experimental measurements are described in three different levels of detail accordingly in excel sheets “Raw Data”, “Test Results” and “Test Summary”. The configuration file specifies which (or all) of these sheets may be imported into the database.

The import configuration principle is the same for each one of these sheets. The EFFECT_BLOCKS section is used to configure simultaneous reading of many effects grouped in sub-blocks of measurement values in accordance to the variations of the experimental factors where for each sub-block, a group of values and associated parameters are configured.

1	TEST RESULTS								
2									
3									
4		Replicate 1					Replicate 2		
5		T1	Titanium Dioxide	Average raw data	100-(Exp/Max*100)	Cell viability (%)	T1	Titanium	Average raw data
6			0	0.336	85.429	100.000		0	0.289
7			1	0.449	80.554	94.294		1	0.323
8			5	0.427	81.495	95.395		5	0.328
9			10	0.468	79.716	93.312		10	0.317
10			25	0.443	80.788	94.568		25	0.347
11			50	0.292	87.357	102.257		50	0.288
12			75	0.412	82.147	96.158		75	0.325
13			100	0.291	87.394	102.301		100	0.426
14			Intracellular LDH Control	1.970					2.365
15									
16		Replicate 1					Replicate 2		
17		T2	Titanium Dioxide	Average raw data	100-(Exp/Max*100)	Cell viability (%)	T2	Titanium	Average raw data
18			0	0.349	84.653	100.000		0	0.510
19			1	0.348	84.678	100.029		1	0.688
20			5	0.393	82.718	97.715		5	0.771
21			10	0.461	79.702	94.152		10	0.849
22			25	0.458	79.854	94.331		25	0.808
23			50	0.458	79.821	94.293		50	0.711
24			75	0.690	69.605	82.224		75	0.683
25			100	0.671	70.458	83.231		100	0.657
26			Intracellular LDH Control	1.923					1.943
27									

Figure 13 IOM template: Test results sheet

In the case of the sheet “Test results” (see Figure 13), the effects block contains 4 sub-blocks defined by varying the replicate and time points. For each sub-block, a group of values is defined by varying the concentrations. In the database each measurement value is stored together with 3 parameters: time point, replicate and concentration.

1	TEST SUMMARY				
2	Titanium Dioxide				
3	NM-103 (TiO2)				
4		T1	Average	St Dev	
5			NM-103 (TiO2)		
6			0	100.00	0.00
7			1	94.80	3.53
8			5	96.18	1.89
9			10	95.27	3.07
10			25	95.90	1.51
11			50	100.11	2.12
12			75	96.68	1.63
13			100	98.99	4.25
14					
15		T2	Average	St Dev	
16			NM-103 (TiO2)		
17			0	100.00	0.00
18			1	97.85	6.21
19			5	94.27	6.68
20			10	90.83	7.19
21			25	91.32	5.76
22			50	92.85	2.74
23			75	85.48	4.88
24			100	86.74	5.01
25					

Figure 14 IOM template: Test summary sheet

In the case of the sheet “Test summary” (see Figure 14), the effects block contains 2 sub-blocks defined by varying only the time points. For each sub-block, a group of values is defined by varying the concentrations. In the database each measurement value is stored together with 3 parameters: time point, replicate and concentration.

The configuration files for IOM templates are developed on case by case basis, as the templates are confidential.

5. NANOREG DATA TRANSFER

Most of the data generated by the NANoREG project is entered via a web entry tool, into a MySQL database, both developed by the Dutch Organization for Applied Scientific Research (TNO). The TNO database design is based on templates developed by the European Joint Research Center (JRC) for assays performed in NANoREG. The rest of the data is entered into Excel files, following the NANoREG templates, described in the previous sections. The Excel files are available at the NANoREG installation of the CIRCABC system. Upon an agreement between NANoREG and eNanoMapper, the NANoREG data is transferred to the eNanoMapper database. WP3 received several versions of the TNO experimental database as MySQL dump (July, September and October 2016).

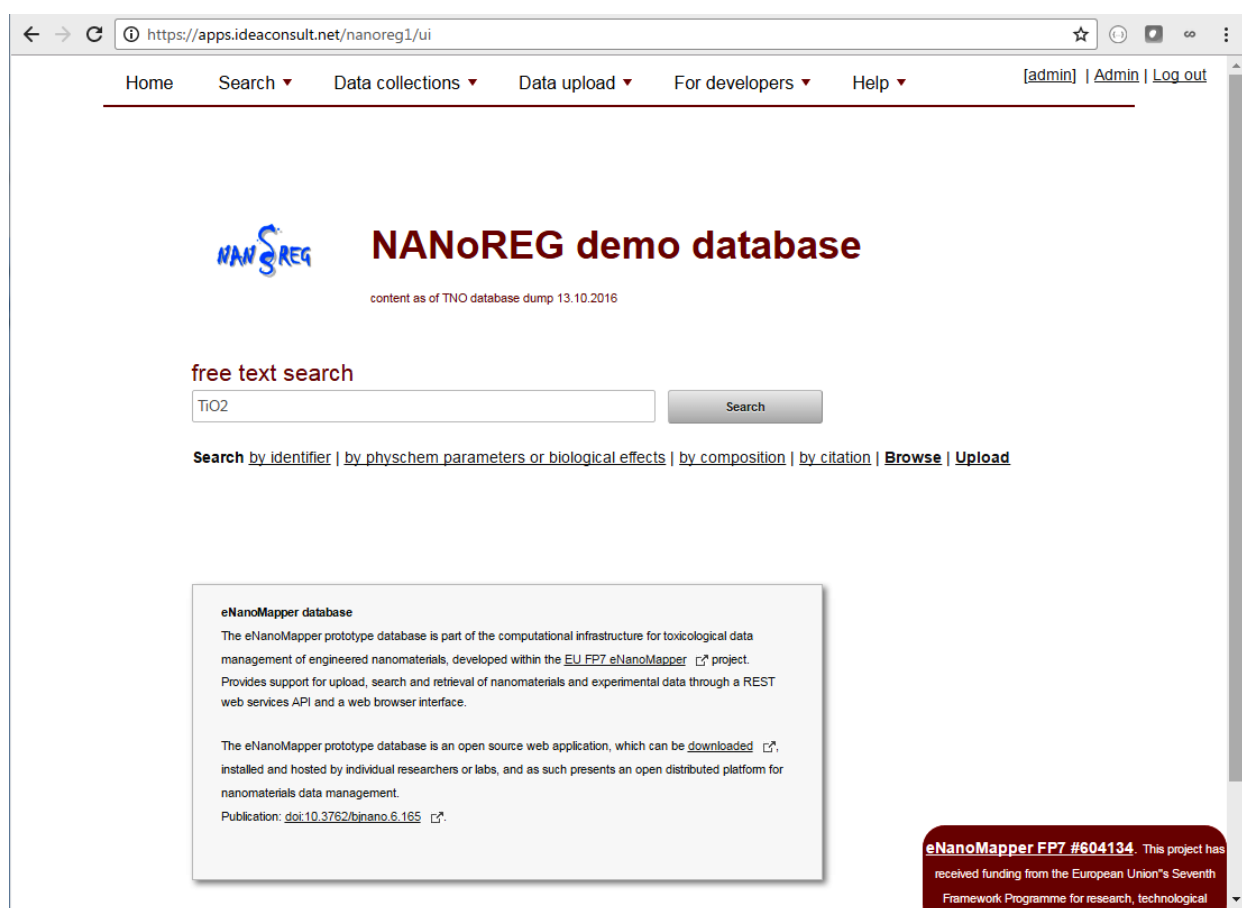


Figure 15 NANoREG database application front page

The conversion from the TNO SQL dump and import into the eNanoMapper database is performed via custom SQL scripts. Technical matters have been discussed between Ideaconult & TNO over email and teleconferences.

An instance of the eNanoMapper database and the search application were installed at <https://apps.ideaconsult.net/nanoreg1> and <https://sandbox.ideaconsult.net/search/nanoreg1> (Figure 15 and Figure 16 respectively). The credentials are distributed for testing purposes to partners from NANoREG, eNanoMapper and CEINT, Duke University, US.

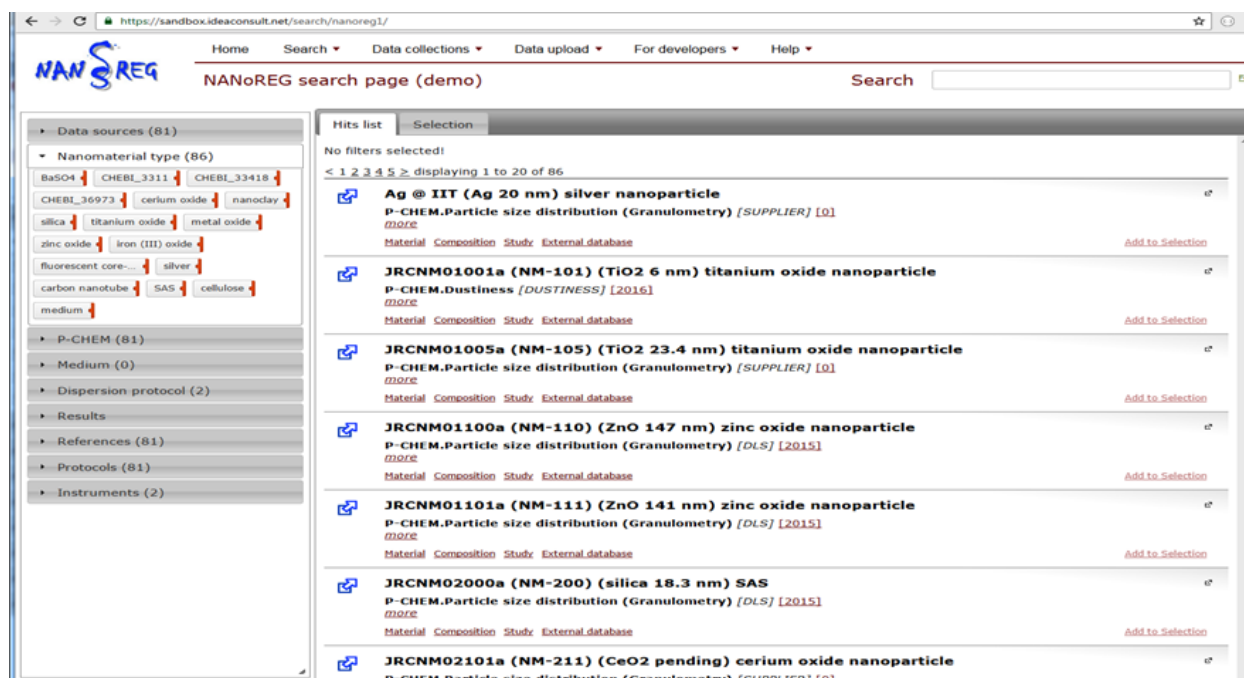


Figure 16 Search application with NANoREG data

The search application relies on search services and user interface previously reported in D3.2 and D5.5. During the reporting period the search application was considerably updated according to user feedback and included in the list of applications reported in D5.7. A user guide is available at eNanoMapper tutorial repository⁷. The NANoREG specific search application is set up for testing purposes during the data transfer, to allow the fine tuning of the user interface. Apart from the specific search interface, the NANoREG data will be integrated in the main search application at <http://search.data.enanomapper.net>.

⁷ https://github.com/enanomapper/tutorials/tree/master/Entering_and_analysing_nano_safety_data

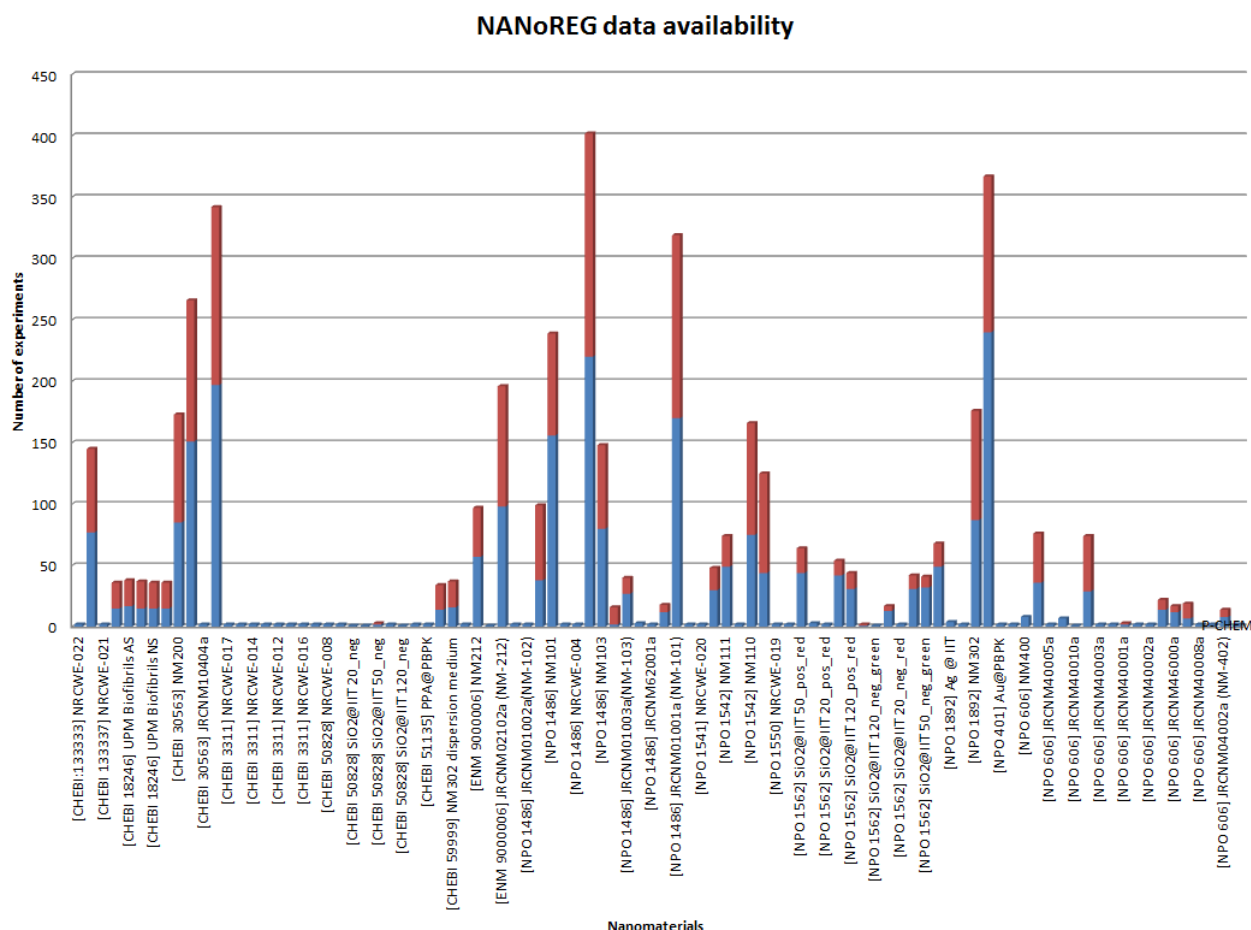


Figure 17 NANoREG data availability

Besides the SQL dump, a large amount of NANoREG data is provided as Excel files, provisionally following the NANoREG templates. Statistics of the files content is provided in ANNEX 1. NANoREG EXCEL FILES OVERVIEW. Cleaning, configuring and importing into the database are ongoing.

6. CONCLUSION

The major developments reported in this deliverable are as follows:

- A *Material Schema extension* for the new ISA-JSON specification, which is the counterpart of the “material file” in the ISA-Tab-Nano specification;
- Cleaning and creation of configurations of the publicly available NANOREG templates, enabling direct import into the database;
- Development and open access to tools for exporting the internal data structures of the eNanoMapper database into ISA-JSON files;
- Development and open access to tools for converting between multiple input formats (including Excel templates) into ISA-JSON and semantic formats;
- Ontology annotation is ongoing, with close collaboration with eNanoMapper WP2;
- Creation of a documentation site, describing the developments above;
- The documentation site may serve as an entry point for templates download and related information. It is accessible via the eNanoMapper [tutorials repository](#) at Github and via menu “Data templates” of the search application at <https://search.data.enanomapper.net>

Being able to read the Excel spreadsheets and write the internal data model into ISA-JSON files accomplishes the goal of automatically generating the ISA files, and enables exporting query results from the database in any desired format. The sets of Excel templates are based on existing NanoSafety Cluster templates that are used or considered for adoption by several NSC projects, such as caLIBRAte and NanoReg2, which are currently implementing the eNanoMapper database data model for collection of data. The templates are coupled with JSON configuration files, enabling import into the database through a simple web form or programmatically via Application Programming Interface. The use of the tools and templates described in this deliverable is also foreseen within the projects enabling harmonized collection of data, which will eventually be compatible across projects.

7. BIBLIOGRAPHY

- [1] 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.
- [2] 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.
- [3] 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.
- [4] P. Kohonen, E. Benfenati, D. Bower, R. Ceder, M. Crump, K. Cross, R. C. Grafström, L. Healy, C. Helma, N. Jeliaskova, V. Jeliaskov, S. Maggioni, S. Miller, G. Myatt, M. Rautenberg, G. Stacey, E. Willighagen, J. Wiseman, and B. Hardy, "The ToxBank Data Warehouse: Supporting the Replacement of In Vivo Repeated Dose Systemic Toxicity Testing," *Mol. Inform.*, vol. 32, no. 1, pp. 47–63, Jan. 2013.
- [5] 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.
- [6] S.-A. Sansone, P. Rocca-Serra, A. Gonzalez-Beltran, and D. Johnson, "ISA Model and Serialization Specifications 1.0." Zenodo, 2016.

ANNEX 1. NANoREG EXCEL FILES OVERVIEW

The NANoREG templates for phys-chem characterisation consist of 10 files, describing 20 assays.

WP2 - phys chem	Number of fields
crystalline phase_NR_20160828.xlsx	149
WAXD	30
XRD	119
crystallite size_NR_20160828.xlsx	349
SAXS	201
WAXD	34
XRD	114
NRCWE XRF_data_2016-08-31.xlsx	513
WDXRF	513
NRCWE-biomolecule interaction_2016-08-31.xlsx	1888
BSA	115
IL-6	591
IL-8	591
LDH	591

Results_TEM_PTA_D2.10 Protocol(s) for size-distribution analysis of primary MNM	657
object.xlsx	
Coda Cerva_PTA	169
Coda Cerva_TEM	488
SEM_NMs size.xlsx	394
assay_SEM	394
WP2_31_BiomoleculeInteraction_20160831.xlsx	154
BSA	109
IL-6	15
IL-8	15
LDH	15
WP2_31_Coating_TGA_20160801.xlsx	190
plain template	190
WP2_31_ElementalAnalysis_CHN_150716.xlsx	260
elemental composition_CHN	260
WP2_31_FunctionalGroupsCoating_ATR-FTIR_20160718.xlsx	158
coating_ATR-FTIR	158
Total	4712

The NANoREG templates for in vitro experiments consist of 7 files, describing 11 assays.

WP5 - in vitro	Number of fields
NPL_WP5_46_Solubility_Colorimetry_v9.xlsx	536

assay_colorimetry	215
batch dispersion	321
WP5_Chromosomal damage in vitro_49FIOH.xlsx	384
Chromosomal damage in vitro	384
WP5_Cytotoxicity in vitro_49FIOH.xlsx	696
Cytotoxicity in vitro_BEAS 2B	318
Cytotoxicity in vitro_THP-1	378
WP5_DNA damage in vitro.xlsx	498
DNA damage in vitro	498
WP5_DNA damage in vitro_49FIOH.xlsx	498
DNA damage in vitro	498
WP5_mRNAexpression in vitro_49FIOH.xlsx	756
mRNA expression in vitro_IL-1B	378
mRNA expression in vitro_TNF-a	378
WP5_protein_secretion in vitro_49FIOH.xlsx	872
Protein secretion in vitro_IL1b	436
Protein secretion in vitro_TNFa	436
Total	4240

The NANoREG templates for in vitro experiments are listed in the following table.

WP4 - in vivo	Number of fields
Lung_Distribution_-_CRM_based_on_T1.5_template_4.4.xlsx	74

Tabelle1	74
Lung_Distribution_-_IBM_based_on_T1.5_template_4.4.xlsx	55
Tabelle1	55
WP4_BAL cell counts_49FIOH.xlsx	642
BAL cell counts	642
WP4_DNA damage in vivo_49FIOH.xlsx	646
DNA damage in vivo - BAL cells	323
DNA damage in vivo - lung cells	323
WP4_ecotox. D.magna_TEK_JRCNM02000a.xlsx	86
MN assay in vivo - D.magna	86
WP4_ecotox. D.magna_TEK_JRCNM02102a_1.xlsx	86
MN assay in vivo - D.magna	86
WP4_ecotox. D.magna_TEK_JRCNM02102a_2.xlsx	80
MN assay in vivo - D.magna	80
WP4_ecotox. D.magna_TEK_JRCNM03000a_1.xlsx	79
MN assay in vivo - D.magna	79
WP4_ecotox. D.magna_TEK_JRCNM03000a_2.xlsx	87
MN assay in vivo - D.magna	87
WP4_ecotox. D.magna_TEK_JRCNM04001a.xlsx	86
MN assay in vivo - D.magna	86
WP4_ecotox. P.subcap_TEK_JRCNM01003a_1.xlsx	86
MN assay in vivo - microalgae	86

WP4_ecotox. P.subcap_TEK_JRCNM01003a_2.xlsx	87
MN assay in vivo - microalgae	87
WP4_ecotox. P.subcap_TEK_JRCNM01003a_3.xlsx	87
MN assay in vivo - microalgae	87
WP4_ecotox. P.subcap_TEK_JRCNM02102a_1.xlsx	86
MN assay in vivo - microalgae	86
WP4_ecotox. P.subcap_TEK_JRCNM02102a_2.xlsx	87
MN assay in vivo - microalgae	87
WP4_ecotox. P.subcap_TEK_JRCNM02102a_3.xlsx	86
MN assay in vivo - microalgae	86
WP4_ecotox. P.subcap_TEK_JRCNM03000a_1.xlsx	86
MN assay in vivo - microalgae	86
WP4_ecotox. P.subcap_TEK_JRCNM03000a_2.xlsx	87
MN assay in vivo - microalgae	87
WP4_long term effect SYSTEMIC GENOtoxicity_ENEA.xlsx	433
COMET - sistemic genotox	129
lung genotoxicity	40
Micronuclei - sistemic genotox	132
PIG-A sistemic genotox	132
WP4_mRNAexpression in vivo_49FIOH.xlsx	1020
mRNA expression in vivo_IFN-g	120
mRNA expression in vivo_IL-10	120

mRNA expression in vivo_IL-13	220
mRNA expression in vivo_IL-1b	220
mRNA expression in vivo_IL-6	120
mRNA expression in vivo_TNF-a	220
WP4_systemic effects bone marrow_49FIOH.xlsx	322
MN assay in vivo - bone marrow	322
WP47_lung_histology_genotoxicity.xlsx	341
lung genotoxicity	69
organ histology	272
Total	4729