DELIVERABLE D5.3

User applications for importing NanoWiki data

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:	Maastricht University (UM)
DATE:	31.1.2015
VERSION:	V.1.0

Call identifier	FP7-NMP-2013-SMALL-7
Document Type	Deliverable Report
WP/Task	WP5 Community Outreach / Task 5.5 User application for importing NanoWiki data, Task 5.3 Integrated Testing
Document ID	D5.3
Status	Final Version

Partner Organisations	• Douglas Connect, GmbH (DC) • National Technical University of Athens (NTUA) • In Silico Toxicology (IST) • Ideaconsult (IDEA) • Karolinska Institutet (KI) • VTT Technical Research Centre of Finland (VTT) • European Bioinformatics Institute (EMBL-EBI) • Maastricht University (UM)
Authors	E.L. Willighagen Reviewed by Barry Hardy (DC)
Purpose of the Document	To report the outcomes of Task 5.3 on NanoWiki Data Application.
Document History	1. First draft: 2014-12-28 2. Second draft: 2015-01-25 3. Third draft: 2015-02-01 4. Fourth draft: 2015-02-08 5. Final version: 2015-02-21

TABLE OF CONTENTS

TABLE OF FIGURES

GLOSSARY

Abbreviation / acronym / namespaces	Description
RDF	Resource Description Framework
CCZero	Creative Commons Zero waiver
NCBI	National Center for Biotechnology Information
CML	Chemical Markup Language
SPARQL	SPARQL Protocol and RDF Query Language (recursive acronym)
SMW	Semantic MediaWiki
clo (namespace)	http://purl.obolibrary.org/obo/
bao (namespace)	http://www.bioassayontology.org/bao#
SMILES	Simplified molecular-input line-entry system
SDF	Structure Data File
XML	eXtensible Markup Language

1. SUMMARY

The eNanoMapper Community Outreach focuses on the implementation of a number of use cases defined by the community. One common use case is the import into the eNanoMapper database of data from another database. This use case therefore consists of exporting data from one database in some format, writing an importer that can handle that format and reading it into the knowledge base, with minimal or no information loss.

The goal of the work completed here was to expose a dataset originally initiated in the lab of the Karolinska Institutet by the current principal investigator of the Maastricht University team. This dataset, called NanoWiki, currently consists of an internal knowledgebase of physicochemical characterizations and biochemical assay results extracted from the literature for more than 300 nanomaterials, mostly metal oxide nanomaterials. The tasks entailed the work to expose this data via the first eNanoMapper data warehouse platform. The data from the knowledgebase was exported as Resource Description Framework (RDF) data and as such imported in the eNanoMapper database using a dedicated importer to map data to the data structure of the eNanoMapper database.

The results show that we successfully worked out this use case. It should be noted that the RDF export of NanoWiki was not using the eNanoMapper ontology, though during the task, the NanoWiki knowledgebase was updated to at least use ontologies in the process of being adopted by eNanoMapper (WP2). This is demonstrated with the structure for some studies showing the annotation of bioassays and cell lines particularly. The NanoWiki is now exposed via the http://data.enanomapper.net/ platform, serving the further purpose of demonstrating the functionality of the running development version of this platform. The data is available under a CCZero data license.

2. INTRODUCTION

2.1. DATA EXCHANGE AND INTEROPERABILITY

The use case of data exchange (partially or in full) is a very common one in the life sciences. In fact, many databases exist that integrate data from other databases, such as the NCBI Entrez databases for genes, and the PubChem database for chemical compounds. To make such integration and index possible, data exchange is required. This exchange means that the representation in one database needs to be converted into a representation in the other database. Of course, this is the central theme of the eNanoMapper project, but a common exchange format as interoperability layer is only yet in the making for nanosafety data. Moreover, sometimes it is just more efficient to make a dedicated convertor. This is particularly likely when the data exchange is only between two databases, and not between many.

However, for the eNanoMapper project it is of key importance to show that the developed knowledge base can, in fact, accept data from other databases, for whatever purpose we may want to do so. This can be indexing, as to provide a search machine for nanomaterials, or to verbatim make a particular dataset available to the public or a selected community.

2.2 EXPORT FORMATS

The formats in which data is exchanged between databases are abundantly available. They also vary in used technologies. For example, for chemical structures many chemical formats exist, though for nanomaterials the choices are much less than for small, organic structures. Examples of such formats include the MDL V2000 molfile and Structure Data File (SDF) formats (.mol and .sdf files) and the newer, more flexible and more exact Chemical Markup Language (CML). For relational databases an SQL dump is neither an uncommon nor an unreasonable export format. Furthermore, for unstructured data exchange, the Resource Description Framework (RDF) is becoming popular, which has the advantage that is nicely integrates ontology use.

Communities are attempting to standardize these formats. Sometimes this standardization is one that arises *de facto*, as is the case with the SDF format. This format is not Open (you cannot freely redistribute the specification), heavily supported, but also occasionally extended with custom extensions. A relevant standard exchange format for nanotoxicity data is ISA-TAB and ISA-TAB-Nano [1]. However, like RDF, this only provides a framework and not a specific format. Templates overcome this issue, but there is no community repository of such templates at this moment.

2.2.1 RDF/XML

Resource Description Framework (RDF) is, as the name suggests, a framework and not a format. Originally it was, however, when the RDF/XML serialization format was deeply integrated into the RDF specification. This has since been fixed, and several RDF serialization formats currently exist, of which RDF/XML is still one used. The XML in RDF/XML refers to eXtensible Markup Language a more semantic exchange format introduced in the last decade of the 20th century, following up on earlier markup languages like SGML. Other RDF serialization formats include NTriples, Turtle, and JSON-LD.

3. NANOWIKI EXPORT

3.1 HISTORY

The NanoWiki data started as a private knowledgebase for NanoQSAR research at the Karolinska Institutet in 2011. The focus of that NanoQSAR study was the toxicity of metal oxides, primarily nanomaterials and limited primarily to cell line experiments, but the knowledgebase also includes a few larger materials. The wiki had two purposes: first, to capture biological data suitable for NanoQSAR research and, second, to serve as an index for relevant literature on metal oxide nanomaterial safety, allowing finding relevant literature for a particular material. The choice of a wiki followed from its flexibility: it allows changing the database "schema" by just adding new or changing old wiki pages.

The wiki was set up as a light weight installation using MediaWiki and the Semantic MediaWiki (SMW) plugin to ensure that data could be semantically encoded, queried using SPARQL (using the RDFIO plugin), and exported. First, this installation was installed on top of the operating system directly (GNU/Linux) but later moved to a virtual machine (still using GNU/Linux).

Since then data has been collected for more than 300 materials, of which slightly over 200 are metal oxides, another 80 carbon nanotubes, and a few metal and alloy nanomaterials. As goes with experimental NanoQSAR literature, there are only few papers with studies on many nanomaterials. The NanoWiki, as a result, consists of more than 50 journal articles (and an occasional report) from more than 25 different scientific journals. However, not all studies from all papers are fully encoded with bioassay details. In fact, only 28 studies are defined, from even fewer articles. For these, only eleven are annotated as biological assays (assay type, assay method, and cell line). This annotation effort started as part of the task reported on here. However, when looking at the number of measurements recorded beyond those on biological end points and thus including physicochemical properties, the numbers are more favorable. Over 700 measurements are recorded.

Table 1 *Overview of the eight semantically annotated bioassays in NanoWiki.*

ID	Paper ID	Cell line	Assay Type	Assay Method
1	Cytotox2011Puzyn [2]	Escherichia coli cells	Cytotoxicity	
2	Field2011 A1 [3]	HaCaT	Cell Viability Assay	Cell number determination
3	Liu2011Cytotox [4]	BEAS-2B	Cell Viability Assay	PI uptake assay
4	Weissleder146 A1 [5]	PaCa2	Cellular uptake	
5	Berg2009 A1 [6]	AML12 cell	Cell Viability Assay	Cell number determination
6	Docter2014 A1 [7]	Caco-2 cell	Cell Viability Assay	MTT reduction assay
7	Docter2014 A2 [7]	Colon HT29	Cell Viability Assay	
8	Limbach2005 A1 [8]	MRC-9 cell	Cellular uptake	MTT reduction assay
9	Gerloff2009 A1 [9]	Caco-2 cell	Cell Membrane Integrity	LDH Release
10	Gerloff2009 A2 [9]	Caco-2 cell	Metabolic Activity Assay	Metabolic Activity
11	Gerloff2009 A3 [9]	Caco-2 cell	DNA Damage Assay	LDH Release

3.2 STRUCTURE

The database is based on a wiki and takes advantage of the "template" approach of MediaWiki installations. Combined with the SMW plugin, this provides an easy way to generate semantic web data. For the purpose of NanoQSAR various templates have been generated, which effectively define a sort of database schema. Templates were defined for key classes of this knowledge base: materials, coatings and coating parts, studies, measurements, methods, papers, journals, descriptors, end points, cell lines, and assays (method, type). Furthermore, there is a relation class, which is used to record claims in papers, such as that the aggregation correlates with the material's zeta potential.

The various concepts in NanoWiki are linked together via relations. If we focus on measurement, and therefore materials, assays, we get a structure as shown in **Figure 1**. Measurements are defined for materials against a particular endpoint. This structure allows for both biological assays as well as physicochemical property measurements. The assays are defined for a number of measurements and a number of materials. The latter information can be derived in two ways, as shown in the figure. The assay itself is further specified by its type, method, the cell line used, and the paper in which it was published. The citation information is not generalized, but limited to articles, as the focus of the NanoWiki was data in published journal articles.

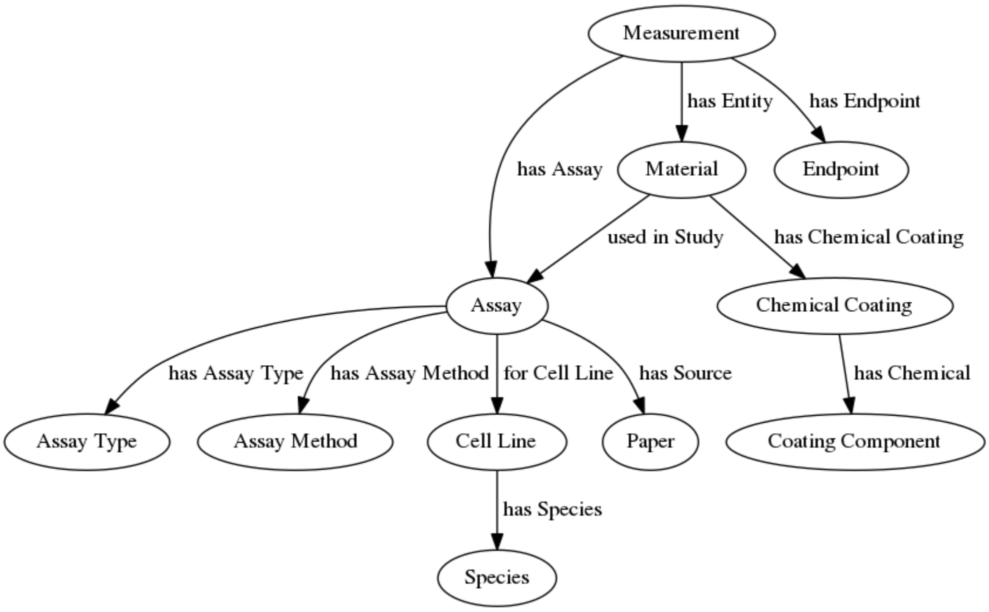

***Figure 1** Outline of the structure of the data in the NanoWiki.*

3.3 RDF AND ONTOLOGIES

The Semantic MediaWiki provides a standardized route for exporting the data, as RDF/XML, but does require the information to be encoded. The MediaWiki template approach is used by the NanoWiki and using this approach most information from the templates are stored as semantic information. The exact details are not part of this work, and further details will be omitted (available with the author, if interested). The outcome, however, is used and depicted in **Error! Reference source not found.**, where it shows facts from the *Liu2011Cytotox* page. Besides the ability to download facts from individual wiki pages, SMW also provides a bulk download which was used in this task.

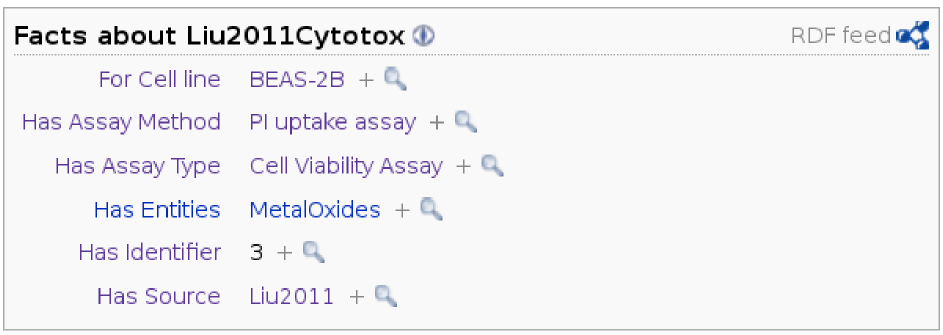

***Figure 2** Visualization of Semantic MediaWiki of "facts" semantically stored in the wiki pages, ready to be exported as RDF.*

The SMW platform has a generic approach to exporting RDF. This includes a very general URI scheme for wiki pages. Because both concepts and properties are both represented by wiki pages, the approach is almost the same for both. A downside for this is that these generic URIs do not look nice. However, there is a workaround by equating wiki pages to or subclassing them to ontology terms. It must be noted that this ontology work is ongoing and was not part of the NanoWiki prior to the start of eNanoMapper.

To exemplify this annotation work, we here list some annotation. For the cell lines the NanoWiki defines equivalence ("Original URI") and subclassing ("Subclass Of") referring to the Cell Line ontology (namespace :*clo* [10]):

Cell line	Has Species	Original URI	Subclass Of
Prostate DU 145	Human	clo:CLO_0002840	
PaCa2	Human	clo:CLO_0007726	
MRC-9 cell	Human	clo:CLO_0007870	
Leukemia RBMI-8226	Human	clo:CLO_0008873	
HaCaT	Human		clo:CL_0000000
Escherichia coli cells	Escherichia coli		clo:CL_0000000
Colon HT29	Human	clo:CLO_0004283	
Breast MCF7	Human	clo:CLO_0007606	
BEAS-2B	Human	clo:CLO_0001925	
AML12 cell	Mus musculus		clo:CLO_0001716

Similarly, for assay types the BioAssay ontology is used (namespace: *bao* [11]):

Assay Type	Original URI	Subclass Of
Toxicity Assay	bao:BAO_0002189	
Physicochemical Characterization		
Oxidative Stress Assay	bao:BAO_0002168	
Genotoxicity Assay	bao:BAO_0002167	
Cytotoxicity Assay	bao:BAO_0002993	
Cell Viability Assay	bao:BAO_0003009	
Cell Growth Assay	bao:BAO_0002100	
DNA Damage Assay		bao:BAO_0002167
Metabolic Activity Assay		bao:BAO_0003009
Cell Membrane Integrity Assay		bao:BAO_0003009

And for assay methods too:

Assay Method	Original URI	Subclass Of
PI uptake assay		bao:BAO_0000167
MTT reduction assay	bao:BAO_0002457	
Tetrazolium Salt Cleavage		bao:BAO_0000166
Cell number determination	bao:BAO_0000572	
Lactate Dehydrogenase Activity		bao:BAO_0000167
Fpg-modified Comet Assay		bao:BAO_0000127

4. ENANOMAPPER IMPORT

Using the RDF exported from the NanoWiki as explained in section 3, the data can be imported into the eNanoMapper infrastructure. This needs a module that understands the NanoWiki RDF and translates this into concepts of eNanoMapper. This is done by the importer, which is outlined here.

4.1 STRUCTURE OF THE IMPORTER

The importer is written in Java as a file reader extending both the iterating reader API of the Chemistry Development Kit (CDK; [12]) by subclassing the "*DefaultIteratingChemObjectReader*" and AMBIT's raw reader, "*IRawReader*" [13]. These APIs allow the eNanoMapper platform, which is implemented by the IDEA partner in eNanoMapper as an extension of the AMBIT platform, to read these files into the system, and reduces this problem otherwise to mapping the NanoWiki RDF content to internal concepts, which is outlined in the next section.
The source code of this reader is found on GitHub at https://github.com/vedina/loom/ with the name *NanoWikiRDFReader* in the (Java) package *net.idea.loom.nm.nanowiki* in the *loom-nm* module. This project is not specific for importing NanoWiki data and includes a lot of other code. The project has a Maven build file that details how the software is compiled. These Maven build instructions also define the *loom-nm* in which this reader resides.
Minimal testing is also provided by this loom project, in the class *NanoWikiRDFTest*. This JUnit-based test suite tests that it extracts the basic structure and that measurements and references are read into the eNanoMapper system.

4.2 MAPPING OF THE INPUT DATA TO THE ENANOMAPPER DATA STRUCTURE

Mapping of concepts was done manually during the development of the importer, involving both the IDEA and the UM teams. Because the model evolved during the task, so did the set of mappings. The importer outlined in Section 4.1 makes use of SPARQL queries to extract information from the data exported from the NanoWiki. These queries aggregated the data, but does not do the mapping yet: it allows for iterating over relevant concepts in the NanoWiki RDF, after which the importer uses the AMBIT Java classes to translate this to database calls. The following table shows how various concepts are mapped using this approach:

NanoWiki concept	AMBIT concept	REST API
Material	Substance	/substance/{id} [1]
Coating Component	Substance Composition/Component, Chemical structure	/substance/{id}/composition [2]
Assay	Protocol, ProtocolApplication	/substance/{id}/study [3]
Endpoint	EffectRecord.endpoint	effects.endpoint in /substance/{id}/study
Paper	LiteratureEntry	Citation in /substance/{id}/study

[1] http://enanomapper.github.io/API/#!/substance_1/getSubstanceByUUID
[2] http://enanomapper.github.io/API/#!/substance_1/getSubstanceComposition
[3] http://enanomapper.github.io/API/#!/substance_1/getSubstanceStudy

Where needed, the NanoWiki RDF is complemented with additional information. Prominently, the core components of materials in the NanoWiki do not have a SMILES description of the chemical composition, unlike coating materials that do. Therefore, the importer adds this SMILES information.

4.3 RESULTING DATA SET IN ENANOMAPPER

The resulting data set loaded into eNanoMapper is described in this section. We here show a few screenshots of aspects of the NanoWiki data as exposed by the AMBIT platform.

4.3.1 THE DATA SET

The data set page in AMBIT is depicted in **Figure 3**. This front page shows the name of the data and further provenance information.

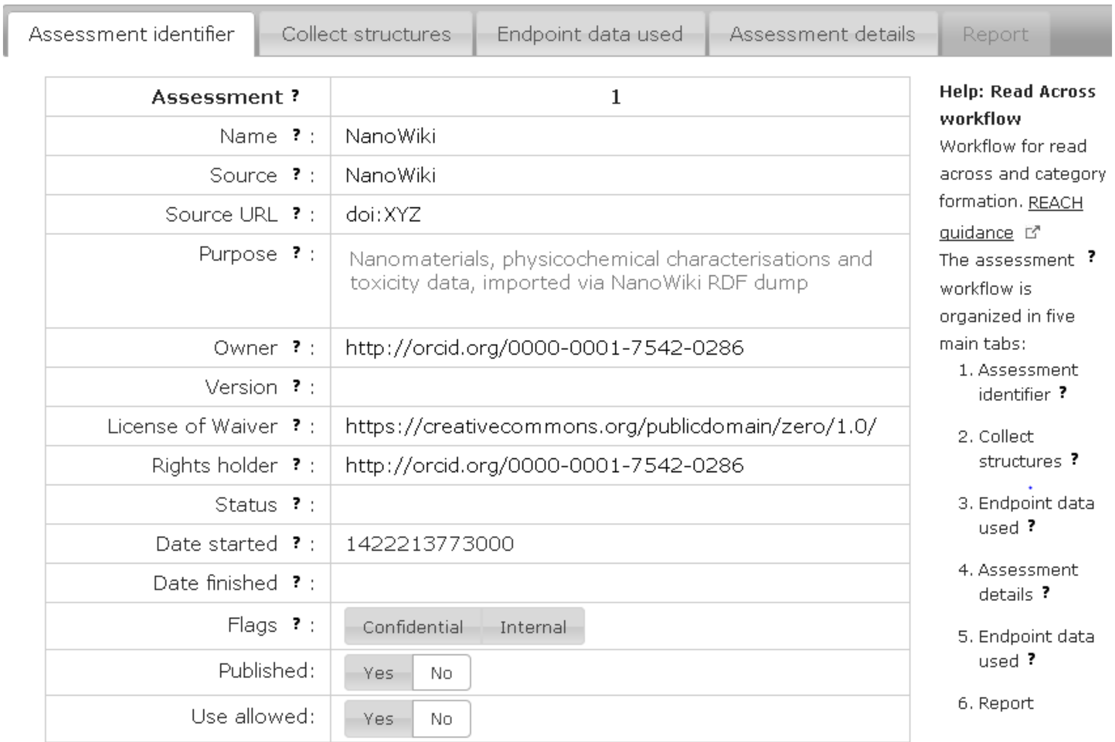

Figure 3 Screenshot of the NanoWiki data set in AMBIT.

4.3.2 A NANOMATERIAL

A material is stored in the database as a substance, and as substances are in AMBIT, it has general information, a composition, and properties. These bits of information are split up over three tabs, shown in **Figure 4**, **Figure 5**, and **Figure 6**.

IUC Substance | Composition | P-Chem (6)

IUC Substance name:	Shaw51 34
IUC Substance UUID:	NWKI-47d254de-8242-3374-9ed3-085227d473cc
IUC Public name:	PNP26
Legal entity:	NanoWiki
Legal entity UUID:	NWKI-9f4e86d0-c85d-3e83-8249-a856659087da
Type substance composition:	MetalOxide
IUC Substance Reference Identifier	
CAS:	?
EC:	?
Chemical name:	?
IUPAC name:	?
UUID:	NWKI-47d254de-8242-3374-9ed3-085227d473cc
IUC Flags:	Alternative Identifier = NP34 Coating = PVA-2DPEG-2DED-2DAminoSPARK680 Composition = Fe2O3 DATASET = NanoWiki Has_Identifier = 64

Figure 4 *Screenshot of the general information about one of the nanomaterials from Shaw et al.* [14].

IUC Substance | Composition | P-Chem (6)

Composition name: Composition Name 47D254DE824233749ED3085227D473CC
Composition UUID: NWKI-47d254de-8242-3374-9ed3-085227d473cc
Purity of IUC Substance:

Type	Name	EC No.	CAS No.	Typical concentration	Concentration ranges		
Coating	Amino_spark680			0 % (w/w)	0 % (w/w)	0 % (w/w)	Also contained in...
Coating	Ed			0 % (w/w)	0 % (w/w)	0 % (w/w)	Also contained in...
Coating	Peg			0 % (w/w)	0 % (w/w)	0 % (w/w)	Also contained in...
Coating	Pva			0 % (w/w)	0 % (w/w)	0 % (w/w)	Also contained in...
Core	Fe2o3			0 % (w/w)	0 % (w/w)	0 % (w/w)	Also contained in...

Showing 1 to 5 of 5 entries

Figure 5 *Screenshot of the chemical composition of one of the nanomaterials from* Shaw et al. [14]. *Each component has associated chemical information that is not shown here but is linked to. Each component is associated with a SMILES string allow chemical substructure searching on the materials.*

IUC Substance | Composition | P-Chem (6)

Filter… Expand all | Collapse all

PNP26

4.5 Particle size distribution (Granulometry) (1)

Test Material Form	Distributi type	Endpoint	Value	Reference	Guideline	Owner	UUID
-	-	PARTICLE SIZE	= 20	DOI		-	NWKI-3dd176d2-8…

Showing 1 study(s) (1 to 1) Previous Next

4.29 Nanomaterial zeta potential (1)

Type of method	Test Material Form	Endpoint	Result	Reference	Guideline	Owner	UUID
-	-	ZETA POTENTIAL	= -4.3	DOI		-	NWKI-3cbe101f-b…

Figure 6 *Screenshot of the physicochemical properties of the nanomaterials from Shaw* et al. [14].

4.3.3 TOXICITY INFORMATION

The toxicity data from NanoWiki is provided in a fourth tab, if present. NanoWiki does not have cytotoxicity data for only a small selection of papers. Even for the eight annotated assays, not all have toxicity described, for various reasons. Being out of scope of the NanoQSAR research is one of them. Serving as an index to search for materials based on material composition and type of experiment for many studies was often the main goal. However, when toxicity information was recorded, because the data set was large enough for NanoQSAR studies, it is presented in the *Tox* tab, as depicted in **Figure 7**.

IUC Substance | Composition | P-Chem (4) | Tox (5)

Filter… Expand all | Collapse all

Micron

8.100 Cell Viability Assay (5)

Reference	Cell line	Doses/concentrations	Endpoint	Result	Result (text)	Guideline	Owner	UUID
2011	HaCaT	= 10 mg/L	Percentage Viable Cells	= 101	-		Chemosphere	NWKI-1bcd287a-291e-4173-…
2011	HaCaT	= 500 mg/L	Percentage Viable Cells	= 98	-		Chemosphere	NWKI-5c9e9e91-0c88-4faa-…
2011	HaCaT	= 7000 mg/L	Percentage Viable Cells	= 74.3	-		Chemosphere	NWKI-72a4393b-5dd7-4bcf-…
2011	HaCaT	= 100 mg/L	Percentage Viable Cells	= 95	-		Chemosphere	NWKI-ae63ad42-ee3c-450a-…
2011	HaCaT	= 1000 mg/L	Percentage Viable Cells	= 92	-		Chemosphere	NWKI-fb9a42a3-ce97-45fc-…

http://dx.doi.org/10.1016/j.chemosphere.2011.04.067

Showing 5 study(s) (1 to 5) Previous Next

Figure 7 *Screenshot of the dose-response toxicological data of hafnium oxides on the HaCaT cell line, as measured by the percentage viable cells. The data was taken from the paper from Field* et al. *in Chemosphere* [15].

5. INTEGRATED TESTING

For the integrated testing, we outlined a few use cases that were anticipated for this data set and that match the functionality of the current database platform. First, we wanted to manually validate that information was successfully copied from the NanoWiki into the AMBIT platform. Second, we want to summarize the NanoWiki data using d3.js and the eNanoMapper API. Third, we wanted to test if we could recover nanomaterials by searching on their chemical composition.

5.1 MANUAL VALIDATION

Manual validation was done, and some of the results are shown in Section 4. We compared counts of materials and papers, which provided a basic indication that all content from the wiki was correctly exported as RDF and imported into the system. More detailed examination was done by comparing the number of measurements for a selection of materials. An example is **Figure 7** where the toxicity for hafnium oxides is shown. This validation was repeated a number of times, in an iterative process as it resulted in updates of the NanoWiki database structure itself, finalized in the structure reported in Section 3.2. Things that changed during this iterative validation process are listed in the below table.

Observation	**Change in NanoWiki during T5.5**
Assay information is not semantically annotated. It is only given as plain text in wiki pages.	Assay information can now be defined using templates, including assay type, assay method, and cell line information. This is done via new wiki templates and eight assays have been updated to use these new templates.
Measured properties are inconsistently represented. Sizes are reported in the Material template, while biological properties use a Measurement template.	All measured properties are now using the Measurement template, including experimental conditions under which the measurement was performed. The latter means that the NanoWiki can now report, for example, the medium used for particle size determination, and report the pH at which a zeta potential was measured.
Mappings are ad hoc and hardcoded in the source code of the importer.	This was addressed by an increasing number of mappings of concepts in the NanoWiki with community ontologies. Obviously, this is work in progress as this overlaps with work in WP2 about ontologies. Further updates are to be expected in this area.
Missing units for experimental data.	Some changes have been made manually. Additionally, we plan to update the importer to report problems like this, as a kind of computer-assisted curation.

5.2 VISUALLY SUMMARIZING NANOWIKI DATA

Because the current validation is manual and would not cover all data without a larger effort, we also wanted to summarize content of the data set visually in a set of summarizing diagrams. The diagrams are made with the d3.js library (http://d3js.org) and a custom JavaScript client library for the AMBIT server, called *ambit.js*. We here summarize the particle size in Figure 8 and zeta potential distributions in Figure 9. Of course, the chemical/toxicological interpretation is limited, because of the diversity of nanomaterials in the database.

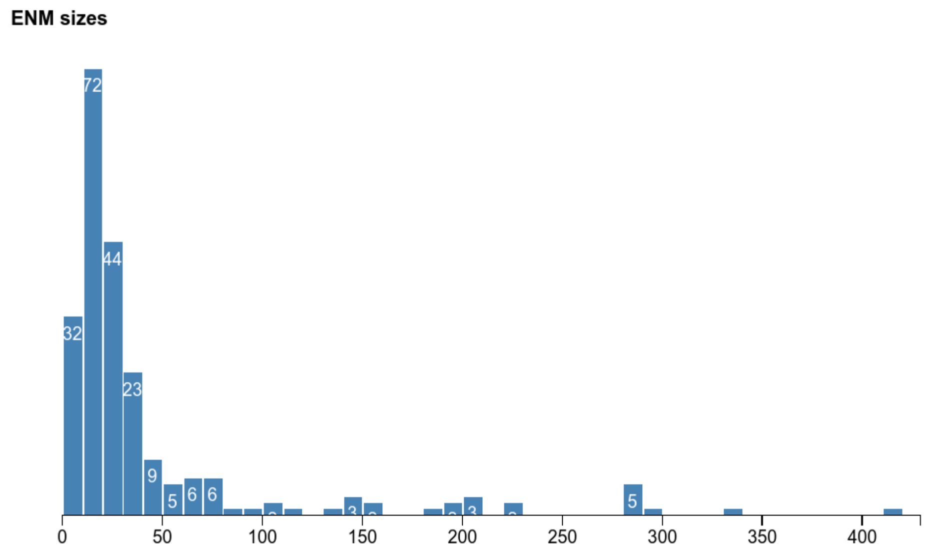

Figure 8 *Distribution of particle sizes (DLS and TEM) of nanomaterials from the NanoWiki data set, extracted from the eNanoMapper API and visualized in a HTML page with the JavaScript-based d3.js library.*

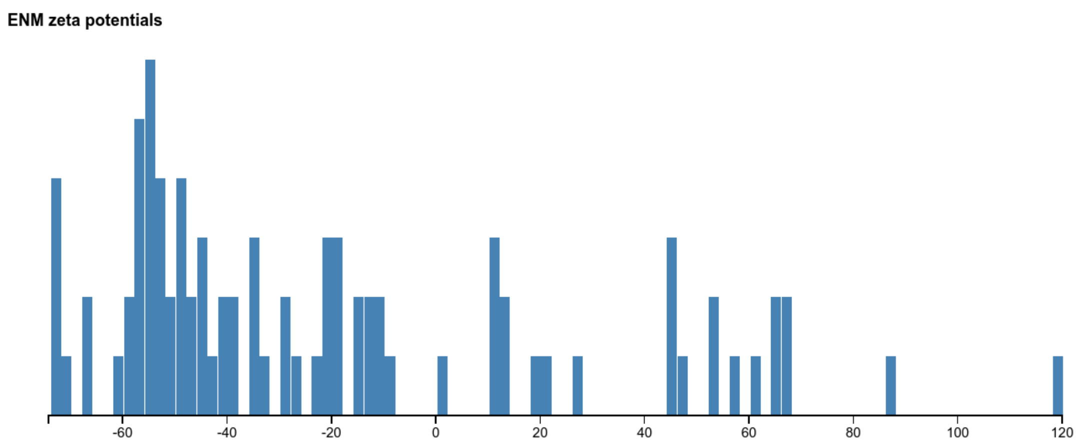

Figure 9 *Distribution of zeta potentials of nanomaterials from the NanoWiki data set, extracted from the eNanoMapper API and visualized in a HTML page with the JavaScript-based d3.js library.*

5.3 SEARCHING NANOMATERIALS BASED ON THEIR CHEMICAL COMPOSITION

A further test was performed by doing substructural searches on the database, which tests if the structural information is properly imported. As an example, Figure 10 shows a substructure search on 2-mercaptoethanesulfonate (MES) which is a component of a chemical coating of one of the nanomaterials in the NanoWiki. The results show an additional nanomaterial from another data set, highlighting a transcription error in the NanoWiki made during the extraction of data from the article by Harper *et al.* [16].

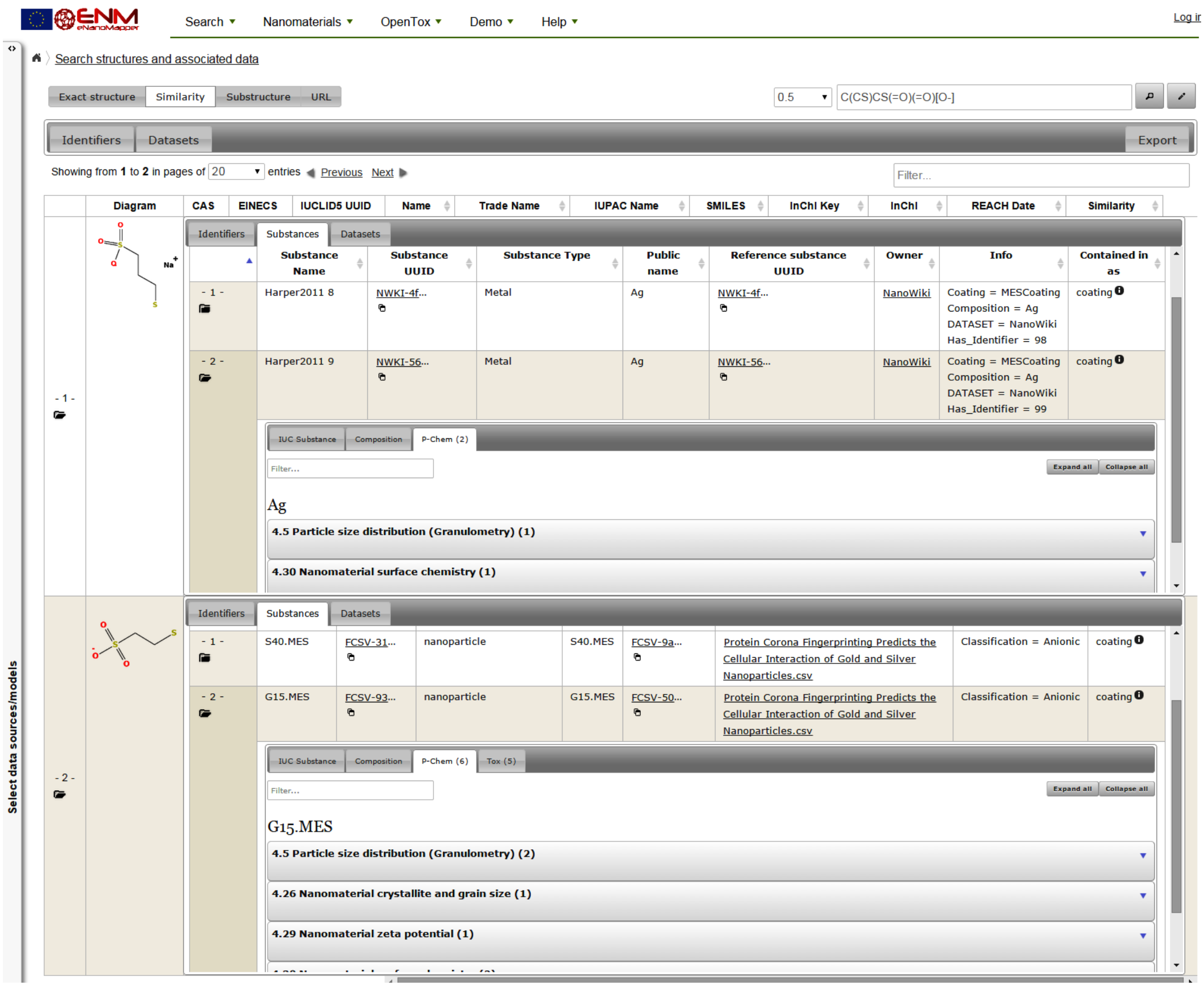

__Figure 10__ Results of a chemical structure similar search for 2-mercaptoethanesulfonate (MES) showing two nanomaterials with such a coating, including one from the NanoWiki database. The search results also show an inconsistency in the structural representation in the NanoWiki.

6. CONCLUSIONS

We report here on the successful export of the data from the NanoWiki knowledgebase in a format that can be read into the eNanoMapper knowledgebase. The importer was written for this RDF/XML format, and mappings have been developed to map RDF predicates to the appropriate database “fields”. The result was tested, with mixed results, using various manual testing approaches. In doing so, we extended the NanoWiki with further semantic annotation to match the direction the eNanoMapper ontology, particularly the annotation of nanomaterials as chemical substances, and annotation of experimental details around the bioassays with appropriate ontology terms from the BioAssay (BAO) Ontology (e.g. for end points and method details) and the Cell Line Ontology (CLO).

6.1 OUTLOOK

This deliverable reports on a small task (**T5.5** was one man month worth of work) and on the initial integrated testing (**T5.3**). The latter, however, is not for this use case alone, and is a task in progress. As the database (**WP3**) and ontology (**WP2**) evolve, we will take advantage of further integrated testing. For this use case, particularly, things that can later be done are comparing read data with data from other databases, reporting on the use of minimal reporting standards, etc. Already initiated is the annotation of content with the ontology; this process will be a continuous one, as the input data may frequently change, requiring new ontology terms. Here an integrated test could be to validate if the input data uses ontology terms not currently part of the eNanoMapper ontology. We further anticipate the NanoWiki data to be further “tested” by taking advantage of Linus’s Law: "given enough eyeballs, all bugs are shallow" [17].

BIBLIOGRAPHY

[1] 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.

[2] T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T. P. Dasari, A. Michalkova, H.-M. Hwang, A. Toropov, D. Leszczynska, and J. Leszczynski, "Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles.," *Nat. Nanotechnol.*, vol. 6, no. 3, pp. 175–8, Mar. 2011.

[3] J. A. Field, A. Luna-Velasco, S. A. Boitano, F. Shadman, B. D. Ratner, C. Barnes, and R. Sierra-Alvarez, "Cytotoxicity and physicochemical properties of hafnium oxide nanoparticles.," *Chemosphere*, vol. 84, no. 10, pp. 1401–7, Sep. 2011.

[4] R. Liu, R. Rallo, S. George, Z. Ji, S. Nair, A. E. Nel, and Y. Cohen, "Classification NanoSAR development for cytotoxicity of metal oxide nanoparticles.," *Small*, vol. 7, no. 8, pp. 1118–26, Apr. 2011.

[5] R. Weissleder, K. Kelly, E. Y. Sun, T. Shtatland, and L. Josephson, "Cell-specific targeting of nanoparticles by multivalent attachment of small molecules.," *Nat. Biotechnol.*, vol. 23, no. 11, pp. 1418–23, Nov. 2005.

[6] J. M. Berg, A. Romoser, N. Banerjee, R. Zebda, and C. M. Sayes, "The relationship between pH and zeta potential of ∼30 nm metal oxide nanoparticle suspensions relevant to in vitro toxicological evaluations," *Nanotoxicology*, vol. 3, no. 4, pp. 276–283, Dec. 2009.

[7] D. Docter, C. Bantz, D. Westmeier, H. J. Galla, Q. Wang, J. C. Kirkpatrick, P. Nielsen, M. Maskos, and R. H. Stauber, "The protein corona protects against size- and dose-dependent toxicity of amorphous silica nanoparticles," *Beilstein J. Nanotechnol.*, vol. 5, pp. 1380–1392, Aug. 2014.

[8] L. K. Limbach, Y. Li, R. N. Grass, T. J. Brunner, M. A. Hintermann, M. Muller, D. Gunther, and W. J. Stark, "Oxide Nanoparticle Uptake in Human Lung Fibroblasts: Effects of Particle Size, Agglomeration, and Diffusion at Low Concentrations," *Environ. Sci. Technol.*, vol. 39, no. 23, pp. 9370–9376, Dec. 2005.

[9] K. Gerloff, C. Albrecht, A. W. Boots, I. Förster, and R. P. F. Schins, "Cytotoxicity and oxidative DNA damage by nanoparticles in human intestinal Caco-2 cells," *Nanotoxicology*, vol. 3. pp. 355–364, 2009.

[10] S. Sarntivijai, Y. Lin, Z. Xiang, T. F. Meehan, A. D. Diehl, U. D. Vempati, S. C. Schürer, C. Pang, J. Malone, H. Parkinson, Y. Liu, T. Takatsuki, K. Saijo, H. Masuya, Y. Nakamura, M. H. Brush, M. A. Haendel, J. Zheng, C. J. Stoeckert, B. Peters, C. J. Mungall, T. E. Carey, D. J. States, B. D. Athey, and Y. He, "CLO: The cell line ontology," *J. Biomed. Semantics*, vol. 5, p. 37, 2014.

[11] U. Visser, S. Abeyruwan, U. Vempati, R. P. Smith, V. Lemmon, and S. C. Schürer, "BioAssay Ontology (BAO): a semantic description of bioassays and high-throughput screening results.," *BMC Bioinformatics*, vol. 12, p. 257, Jan. 2011.

[12] C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, and E. Willighagen, "The Chemistry Development Kit (CDK): an open-source Java library for Chemo- and Bioinformatics.," *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 2, pp. 493–500, Mar. 2003.

[13] N. Jeliazkova and V. Jeliazkov, "AMBIT RESTful web services: an implementation of the OpenTox application programming interface.," *J. Cheminform.*, vol. 3, no. 1, p. 18, Jan. 2011.

[14] S. Y. Shaw, E. C. Westly, M. J. Pittet, A. Subramanian, S. L. Schreiber, and R. Weissleder, "Perturbational profiling of nanomaterial biologic activity," *PNAS*, vol. 105, no. 21, pp. 7387–7392, 2008.

[15] J. A. Field, A. Luna-Velasco, S. A. Boitano, F. Shadman, B. D. Ratner, C. Barnes, and R. Sierra-Alvarez, "Cytotoxicity and physicochemical properties of hafnium oxide nanoparticles," *Chemosphere*, vol. 84, pp. 1401–1407, 2011.

[16] S. L. Harper, J. L. Carriere, J. M. Miller, J. E. Hutchison, B. L. S. Maddux, and R. L. Tanguay, "Systematic evaluation of nanomaterial toxicity: utility of standardized materials and rapid assays.," *ACS Nano*, vol. 5, no. 6, pp. 4688–97, Jun. 2011.

[17] "Linus's Law," *Wikipedia*, 2015. https://en.wikipedia.org/wiki/Linus%27s_Law.