eNanoMapper Updates and its Collaborations with the Community

Authors: Egon Willighagen, Friederike Ehrhart, Linda Rieswijk, Nina Jeliazkova, Chris Evelo, Lucian Farcal, Barry Hardy, Haralambos Sarimveis

February this year marked the completion of the first draft versions of three key components of the eNanoMapper platform: the database, ontology and modelling software. This update highlights these three contributions to the community and some of their applications to solve community challenges. Furthermore, we briefly discuss our new Associate Partner program and other outreach and dissemination activities.

Database - software and demonstration server: http://data.enanomapper.net/
The database software was recently described in detail in a publication in the Beilstein Journal of Nanotechnology (*Jeliazkova et al., 2015*), extending on the conference paper for the nanoinformatics symposium in Belfast in 2014 (*Jeliazkova et al., 2014*). These papers describe the database functionality, demonstrate the default graphical user interfaces (see Figure 1), focusing on the data formats supported to import nanosafety data, and the Application Programming Interface (API). The import and upload functionality currently supports a variety of NanoSafety cluster Excel templates, OECD HT (IUCLID .i5z files) and a custom RDF format for importing NanoWiki. ISA-Tab-Nano import and export is under active development. Furthermore, the importer can easily be extended to support more spreadsheet formats and templates, using the configurable parser with specific JSON configuration developed in the past six months (see https://github.com/enanomapper/nmdataparser/).

The API (whose interactive documentation can be found at http://enanomapper.github.io/API/) supports the automated listing of nanomaterials and their physicochemical and biological properties, searching, and automated upload of data. The API is based on REST and JSON and can be used by any programming language. Various client libraries are currently released in development versions for JavaScript (https://github.com/enanomapper/ambit.js) and the R statistical environment (being internally tested). The software can be downloaded from http://ambit.sourceforge.net/enanomapper.html and installation instructions can be found at the same location.

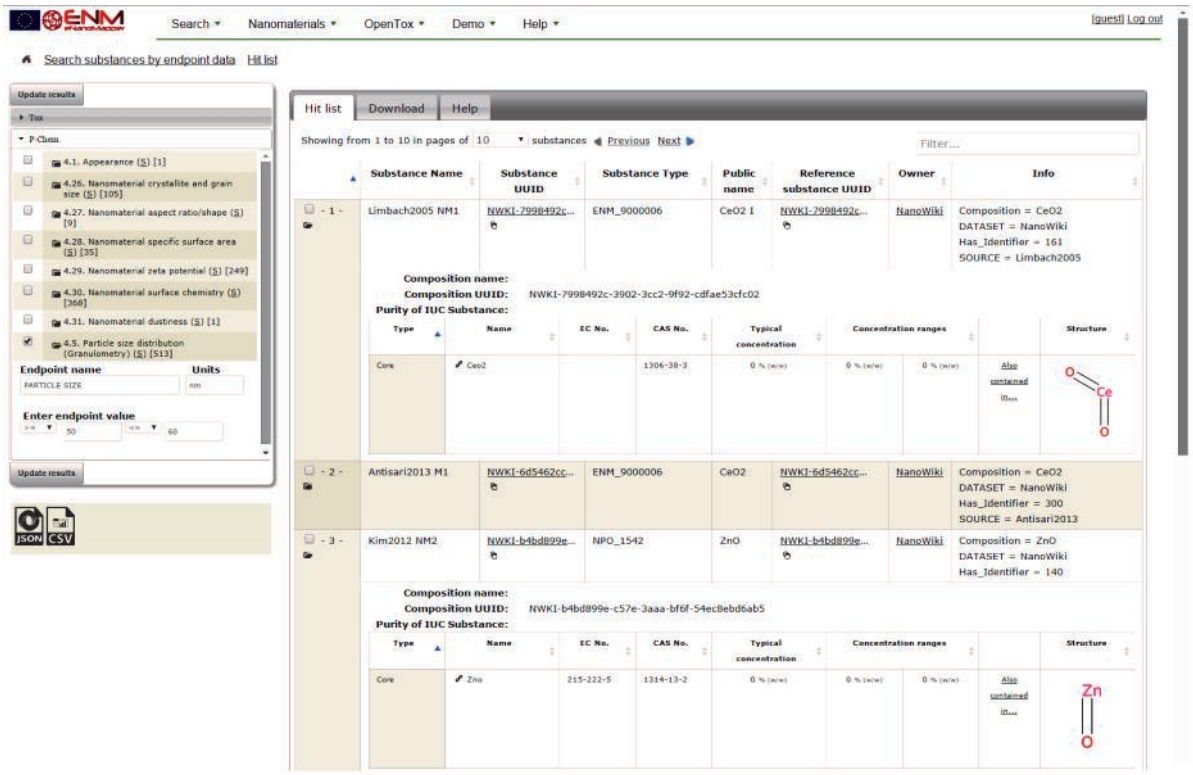

Figure 1. Search functionality in the eNanoMapper database software, showing the results for a search on nanomaterials with a size between 50 and 60 nm

Supported Community Data Sets: the eNanoMapper database server can be set up by NSC projects to host data from their project (please contact us if you would like to try this option). Currently, the platform has been tested for reading templates and data from (partners of) the FP7-ModNanoTox, FP7-MARINA (see the next section), COST Action MODENA and FP7-NANoREG projects.

Updates on the data.enanomapper.net instance: in the past six months a new 'bundle' has been added, containing experimental data from *in vitro* experiments performed at Karolinska Institutet within FP7-MARINA (see Bundle 4; http://data.enanomapper.net/ui/assessment?bundle_uri=https://apps.ideaconsult.net/enanomapper/bundle/4).

Ontology: http://purl.enanomapper.net/onto/enanomapper.owl

The first eNanoMapper ontology was released initially in March 2015, consisting of more than 4500 terms. It was downloaded 500 times from BioPortal alone (http://bioportal.bioontology.org/ontologies/ENM). In September 2015 we released the second version of the ontology, adding almost 2000 classes from new ontologies (e.g., EFO, CCONT and OAE) to the eNanoMapper ontology. Regular updates are planned at six month intervals (*Hastings et al., 2015*).

Community uptake of the ontology: the usability of the ontology has been tested against FP7-NANoREG data. Specialized terms used in NANoReg schema were mapped to existing URIs from the eNanoMapper ontology. Currently, a limited number of the terms can be mapped one-to-one to URIs present in the eNanoMapper ontology. One reason for this limited overlap is the lack of exact matching terms found in the ontology sources re-used within the eNanoMapper ontology. Therefore, additional ontology sources are being explored to be added to the next version of the ontology. Moreover, in some cases, specialized terms originating from the database schema need to be mapped to multiple URIs. Different approaches are currently being explored to solve this problem.

Other community interactions, where the ontology is practically being tested, include collaborations with the UDS and the ISA-Tab communities. eNanoMapper co-organized (*Rumble et al., 2015*) and participated in workshops with these communities in the past months.

Modelling

The first draft version of the eNanoMapper modelling web application has been completed in accordance with the modelling APIs which are documented with swagger at http://enanomapper.ntua.gr:8080/jaqpot/swagger/. The current version of the web application is fully integrated with the eNanoMapper database, the R and Python programming languages and the Weka library of data mining algorithms. It consists of tools for calculating nano-specific descriptors (image, quantum mechanical, gene ontology and Chemistry Development Kit descriptors for nanomaterials), software packages and web services for importing, merging and creating datasets and for building nanoQSAR models (implementations of statistical and machine learning algorithms such as PLS, LASSO, Decision trees, Linear regression, SVM, Clustering etc.) and methods for analyzing biological data. The eNanoMapper computational infrastructure also includes an independent web application for calculating image descriptors (http://enanomapper.ntua.gr:8880/imageAnalysis) and an R package that automates the generation of optimal and fully validated QSAR models (https://github.com/enanomapper/RRegrs/tree/master/RRegrs) (*Tsiliki et al., 2015*). The involvement of users outside eNanoMapper is supported, and they can enrich the toolset available by integrating their own algorithms and nanoQSAR models. eNanoMapper is actively participating in the MODENA COST Action and is engaged with the NSC Harmonisation Initiative project group activities on computational modelling and on modelling requirements and standards.

Associate Partner Program

The eNanoMapper consortium has started its Associate Partner Program (http://www.enanomapper.net/associate-partner-program) to extend its routes to engage with third parties (companies, projects and individuals), an approach picked up from other projects where such programs helped spread solutions. Our current associate partners are represented on the eNanoMapper website http://www.enanomapper.net/tags/associate-partners and include a small and large industrial organization and a academic partner. Please, do not hesitate to contact any of our partners if you are interested in our work and this program.

Application Overview

We have recently started to summarise specific solutions, such as outlined in the above section on our website. This section of our website is a useful place to continue to explore the solutions we offer to the community. You can view this overview at http://enanomapper.net/applications and you will find three categories, matching the ontology, database and modelling work done in our project.

Upcoming Webinars

29 October 2015 (4:00 pm, CET) "*Enriching protein corona fingerprints using gene ontology information - an integration technique*"
See http://www.enanomapper.net/library/webinar-announcement
Recordings of previous eNanoMapper webinars: http://www.enanomapper.net/library/webinars

Recent component releases

Hastings et al. eNanoMapper ontology v2.0,
http://purl.enanomapper.net/onto/enanomapper.owl
Jeliazkova et al. ambit2 2.7.4, **2015**. http://ambit.sourceforge.net/downloads.html
Willighagen et al. NanoWiki v1, **2015**, doi:10.6084/m9.figshare.1330208
Tsiliki et al. RRegrs 0.0.4. **2015** doi:10.5281/zenodo.21946
Willighagen et al. ambit.js 0.0.2. **2015** doi:10.5281/zenodo.16517

Abbreviations

- **API** Application Programming Interface

A way computer programs talk to one another. Can be understood in terms of how a programmer sends instructions between programs. The API specifies how software components should interact. A good API makes it easier to develop a program by providing all the building blocks. A programmer then puts the blocks together.

- **JSON** JavaScript Object Notation

JSON is a lightweight data-interchange format (http://json.org/). It is easy for humans to read and write. It is easy for machines to parse and generate. It is based on a subset of the JavaScript Programming Language, Standard ECMA-262 3rd Edition, December 1999. JSON is a text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and many others.

- **OECD HTs** The OECD Harmonized Templates

The OECD HTs are structured (XML) data formats for reporting safety-related studies on chemical substances. The OECT HTs http://www.oecd.org/ehs/templates/) and the supporting IT tool (IUCLID5, http://iuclid.eu) are used in a regulatory context, for preparation of substance dossiers for REACH and for other regulatory frameworks operating in Europe; as well as by the JRC NanoHub database

- **RDF** The Resource Description Framework

RDF is one of the core common standards and data exchange formats of the Semantic Web. RDF it is based upon the idea of making statements about resources (in particular web resources) in the form of subject–predicate–object expressions. These expressions are known as triples in RDF terminology.

- **REST** Representational state transfer

REST is an abstraction of the architecture of the World Wide Web; more precisely, REST is an architectural style consisting of a coordinated set of architectural constraints applied to components, connectors, and data elements, within a distributed hypermedia system. REST ignores the details of component implementation and protocol syntax in order to focus on the roles of components, the constraints upon their interaction with other components, and their interpretation of significant data elements. OpenTox web services are based on the REST architecture.

- **URI** Uniform Resource Identifier

The URI is an identifier for a term in an ontology using the Web Ontology Language (and any other RDF-based system). It typically looks like a URL, such as used for a common website address.

References

Tsiliki G, Munteanu CR, Seoane JA, Fernandez-Lozano C, Sarimveis H, Willighagen E. RRegrs: an R package for computer-aided model selection with multiple regression models, Journal of Cheminformatics, **2015**, 7:46 doi:10.1186/s13321-015-0094-2

Hastings J, Jeliazkova N, Owen O, Tsiliki G, Munteanu CR, Steinbeck C, Willighagen E. *eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment*, Journal of Biomedical Semantics, **2015**, 6:10 doi:10.1186/s13326-015-0005-5

Jeliazkova N, Chomenidis C, Doganis P, Fadeel B, Grafström R, Hardy B, Hastings J, Hegi M, Jeliazkov V, Kochev N, Kohonen P, Munteanu CR, Sarimveis H, Smeets B, Sopasakis P, Tsiliki G, Vorgrimmler D, Willighagen E. The eNanoMapper database for nanomaterial safety information. Beilstein J. Nanotechnol. **2015**, 6, 1609-1634 doi:10.3762/bjnano.6.165

Jeliazkova, N., Doganis, P., Fadeel, B., Grafstrom, R., Hastings, J., Jeliazkov, V., Kohonen, P., Munteanu, C. R., Sarimveis, H., Smeets, B., Tsiliki, G., Vorgrimmler, D., Willighagen, E., The first eNanoMapper prototype: A substance database to support safe-by-design. In: Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on. IEEE, pp. 1-9. doi:10.1109/bibm.2014.6999367

Rumble J, Willighagen E. Comparing Approaches to the Description of Nanomaterials. CODATA Blog, Ed. S. Hodson, **2015**, http://codata.org/blog/2015/08/07/comparing-approaches-to-the-description-of-nanomaterials/

Fadeel B, Fornara A, Toprak MS, Bhattacharya K. *Keeping it real: the importance of material characterization in nanotoxicology*. Biochemical & Biophysical Research Communications. **2015**, doi:10.1016/j.bbrc.2015.06.178

eNanoMapper Deliverable Reports http://www.enanomapper.net/library/reports

eNanoMapper Meetings

eNanoMapper held two virtual meetings on 16 and Tues 20 October, to discuss developments on eNanoMapper deliverables, and also discuss progress and needs on other projects requiring input from eNanoMapper e.g., on data, templates, ontology, etc.

As a result, the consortium were able to develop concrete common view on harmonisation and interoperability supporting a knowledge infrastructure for the cluster and specific actions between projects: supporting linked data exchange between databases and providing data with a common ontology to applications. The outcomes of the meeting will contribute towards a concrete statement and commitment plan for the eNanoMapper meeting in November in Brussels.

eNanoMapper Meeting in Brussels

The eNanoMapper project is convening a meeting in Brussels on 23rd November DG RTD, Covent Garden, Place Rogier, Brussels. This is a key meeting, at which the consortium plan to present and discuss the progresses made by the project selected under the call "*NMP.2013.1.3-2 Nanomaterials safety assessment: Ontology, database(s) for modelling and risk assessment*." eNanoMapper has reached mid-term and now needs feedback from the Community.

It is expected that a maximum of 2 representatives per project (ideally people dealing with data production or data analysis and dealing with data storage) will attend.