Current Situation: available ontology and templates, interoperability

Egon Willighagen
Maastricht University, The Netherlands
@egonwillighagen
ORCID:0000-0001-7542-0286

Brussels, 25 January 2016, CC-BY

Interoperability

- Interfaces between things
  - Research (language, data, source code, …)
  - Databases (schema, API, …)
  - Tools (API, formats, …)

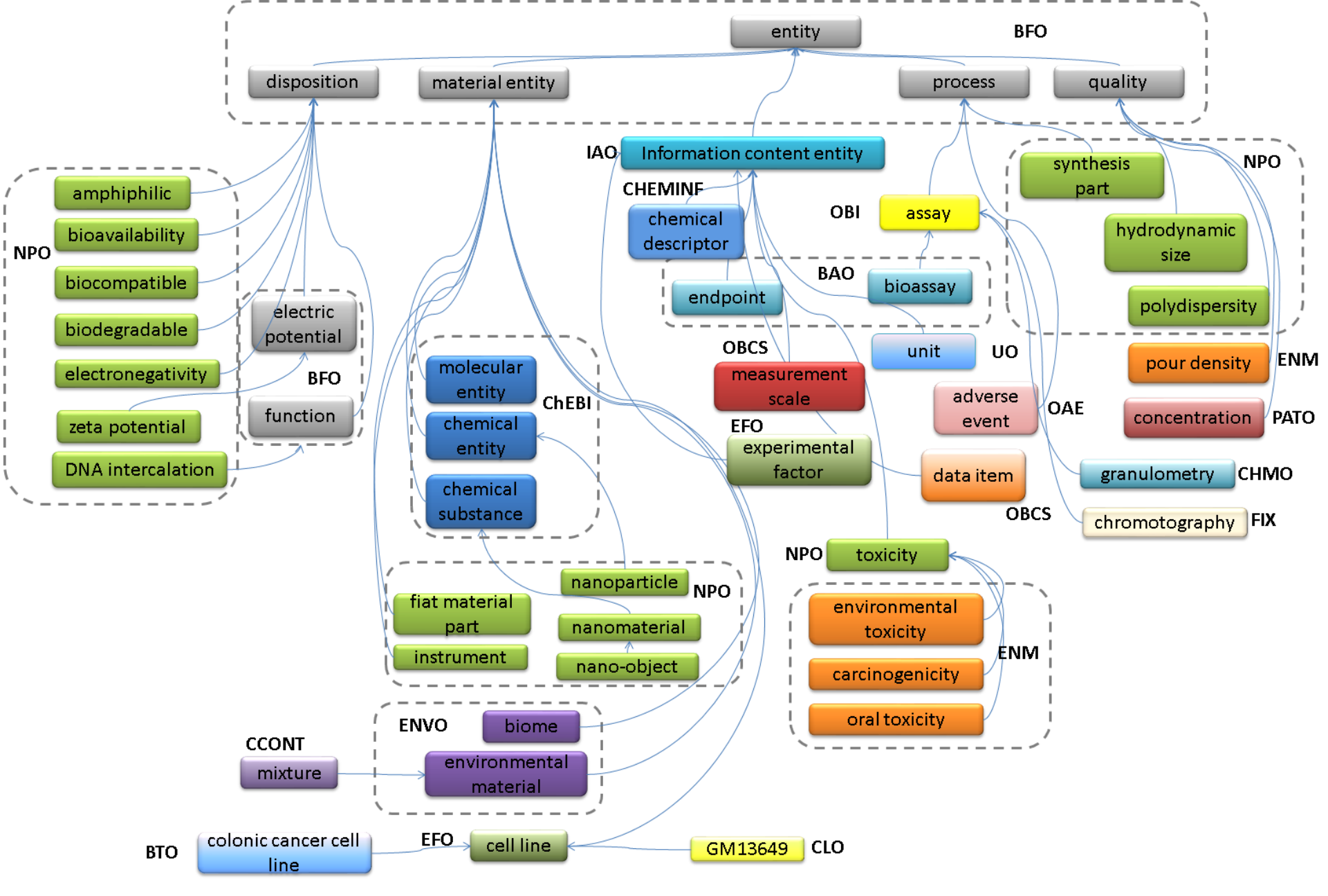
Image: C. Liu, Flickr, CC-BY-NC,
https://www.flickr.com/photos/lvsun/7942762686

eNanoMapper

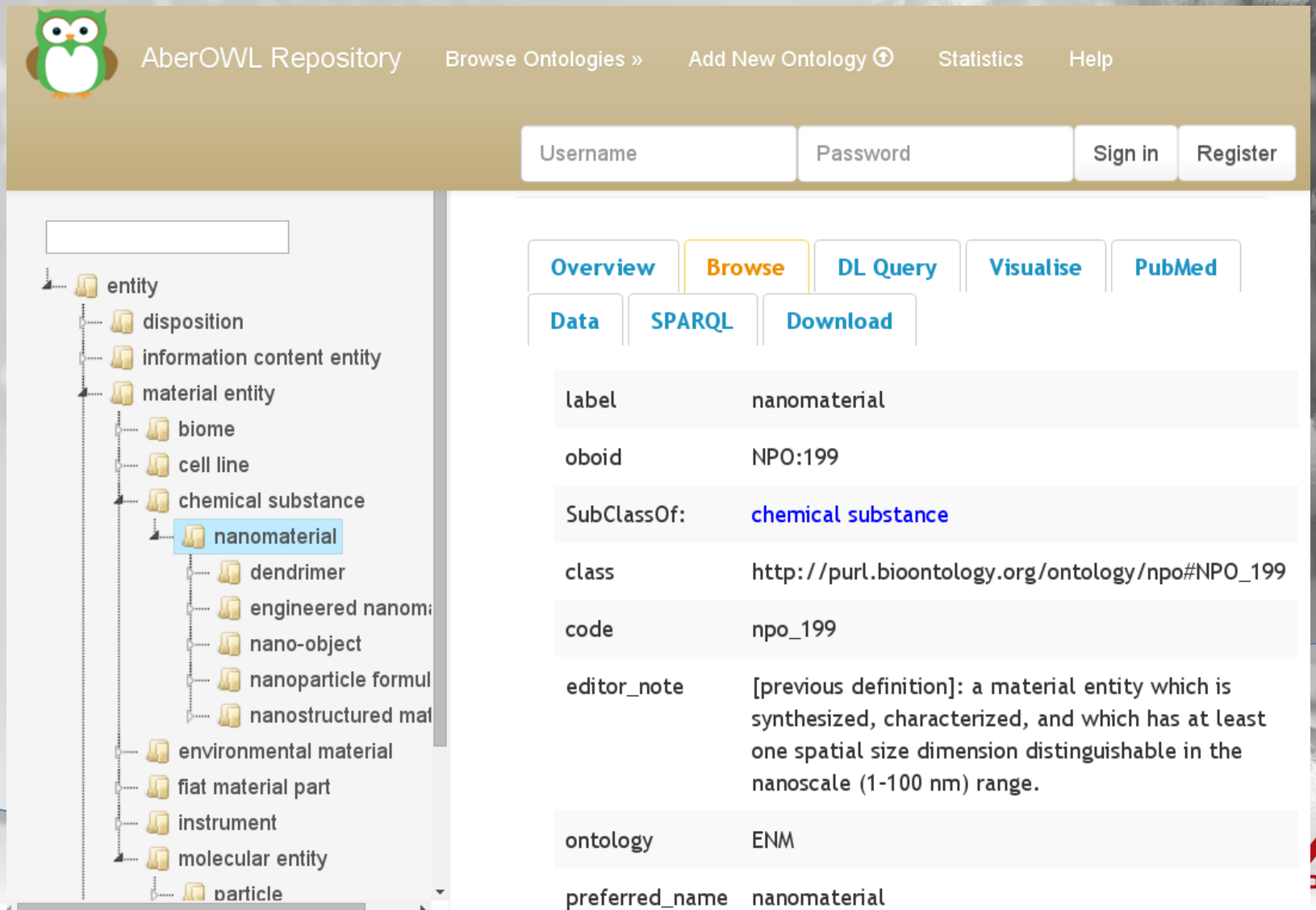
The Ontology

- Interoperability by using the same language
- Allows linking of:
  - the same things
  - similar things (by taking advantage of is-a and part-of relations)
- Bridges between data and publications

eNM
eNanoMapper


entity
BFO
disposition
material entity
process
quality
IAO
Information content entity
CHEMINF
chemical descriptor
OBI
assay
BAO
bioassay
endpoint
NPO
synthesis part
hydrodynamic size
polydispersity
amphiphilic
bioavailability
biocompatible
biodegradable
electronegativity
zeta potential
DNA intercalation
electric potential
BFO
function
molecular entity
chemical entity
chemical substance
ChEBI
OBCS
measurement scale
unit
UO
pour density
ENM
adverse event
OAE
concentration
PATO
EFO
experimental factor
data item
granulometry
CHMO
OBCS
chromotography
FIX
NPO
toxicity
nanoparticle
NPO
fiat material part
nanomaterial
instrument
nano-object
environmental toxicity
carcinogenicity
oral toxicity
ENM
ENVO
biome
CCONT
mixture
environmental material
BTO
colonic cancer cell line
EFO
cell line
GM13649
CLO

BioPortal and AberOWL



Difference between v1 and v2 of eNM ontology

v1 eNM			v2 eNM			Difference v1 and v2 eNM	
Ontology	# classes		Ontology	# classes		Ontology	# classes added
BAO	329		BAO	351		BAO	22
BFO	6		BFO	6		BFO	0
CHEBI	26		CHEBI	26		CHEBI	0
CHEMINF	376		CHEMINF	376		CHEMINF	0
ENM	42		ENM	56		ENM	14
ENVO	438		ENVO	438		ENVO	0
IAO	1		IAO	20		IAO	19
NPO	420		NPO	426		NPO	6
OAE	1921		OAE	1921		OAE	0
OBI	632		OBI	742		OBI	110
owl	21		owl	5		owl	-16
PATO	33		PATO	37		PATO	4
UO	310		UO	310		UO	0
total	4555		BTO	34		BTO	34
			CCONT	12		CCONT	12
			CHMO	71		CHMO	71
			CLO	428		CLO	428
			EFO	1322		EFO	1322
			FIX	17		FIX	17
			OBCS	92		OBCS	92
			total	6690		total added	2135

Internal and External Validation

Internal

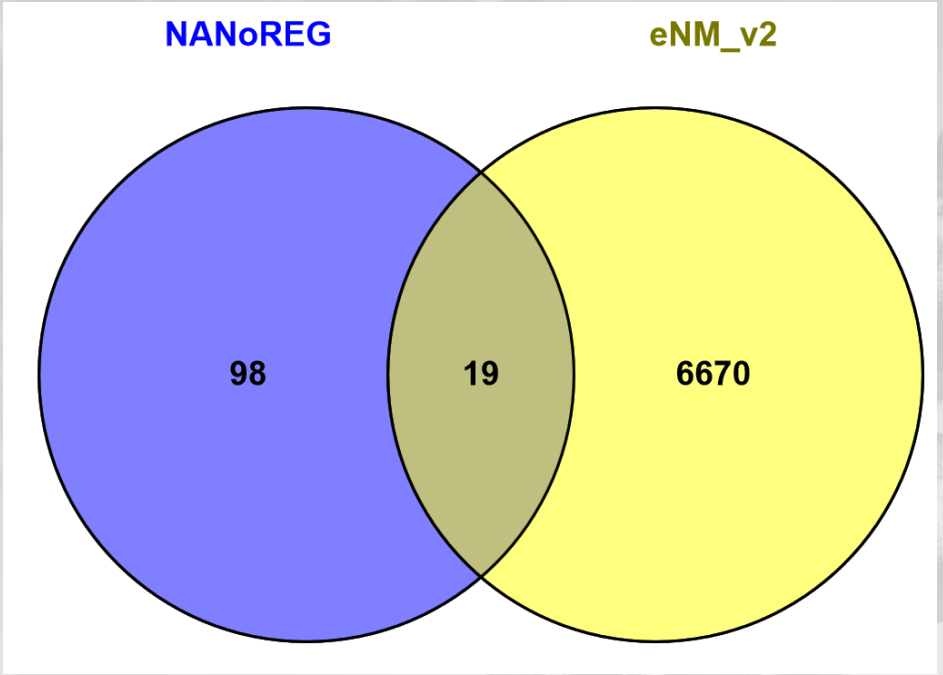
- Use Cases
- RDF serialization

External

- NANoREG WP6
- UDS
- NECID
- Templates

eNanoMapper

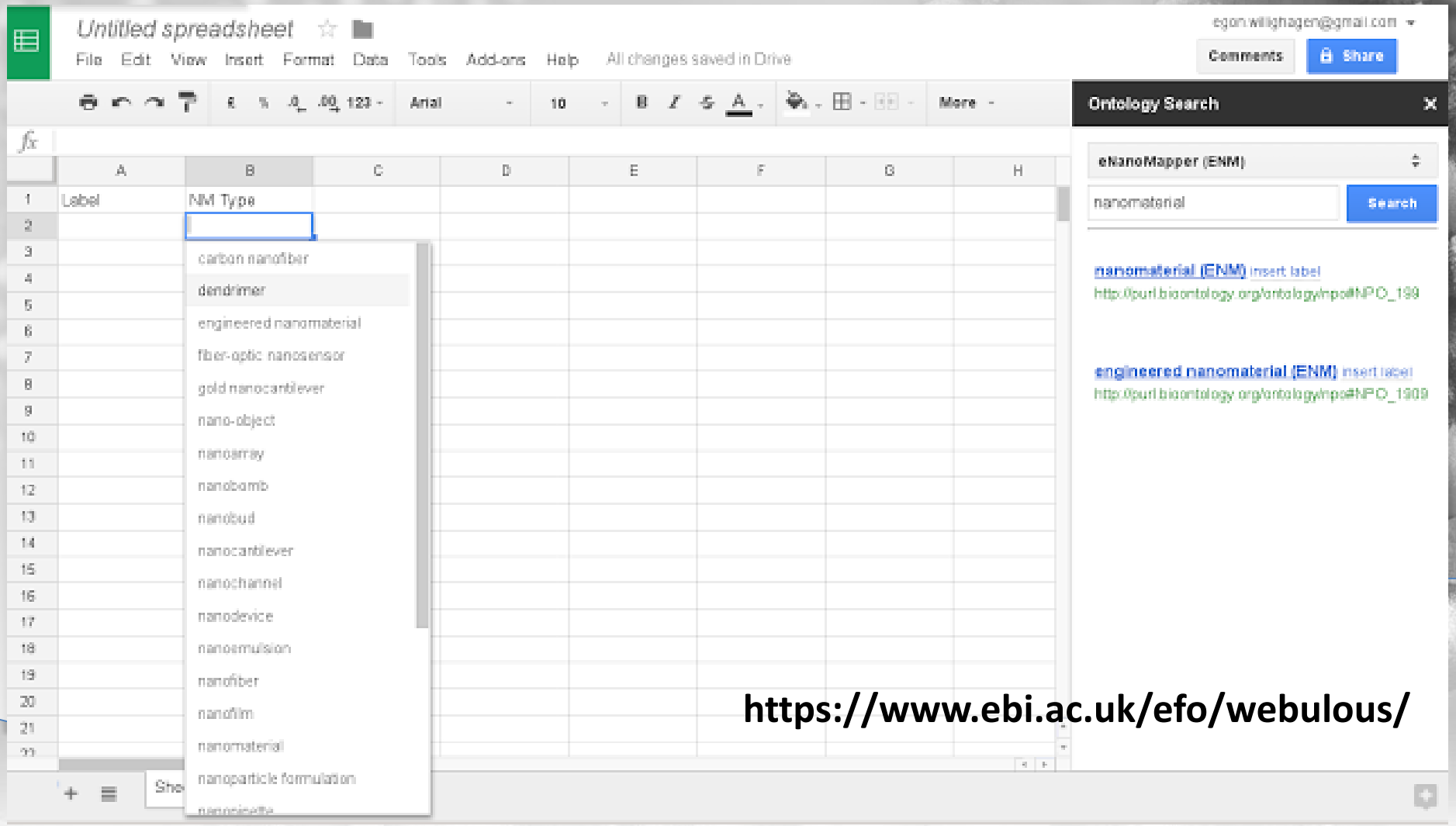
Overlap URIs NANoREG and eNM v2



- Most of the other 98 items are mapped to URIs (but not in v2)
- Ongoing work

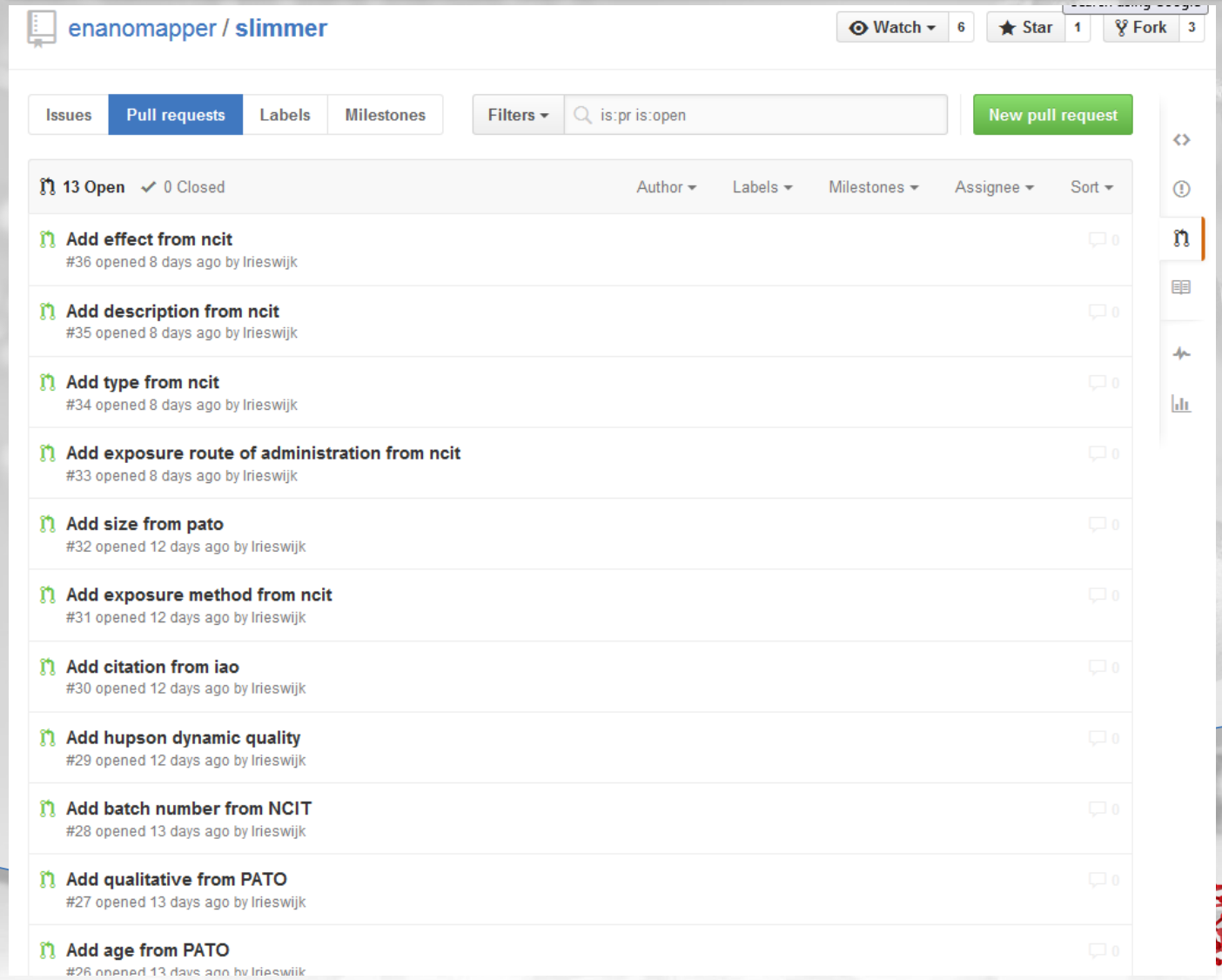
eNM
eNanoMapper

OWL/RDF → Interoperability



https://www.ebi.ac.uk/efo/webulous/

Pull requests within GitHub



eNanoMapper

Templates

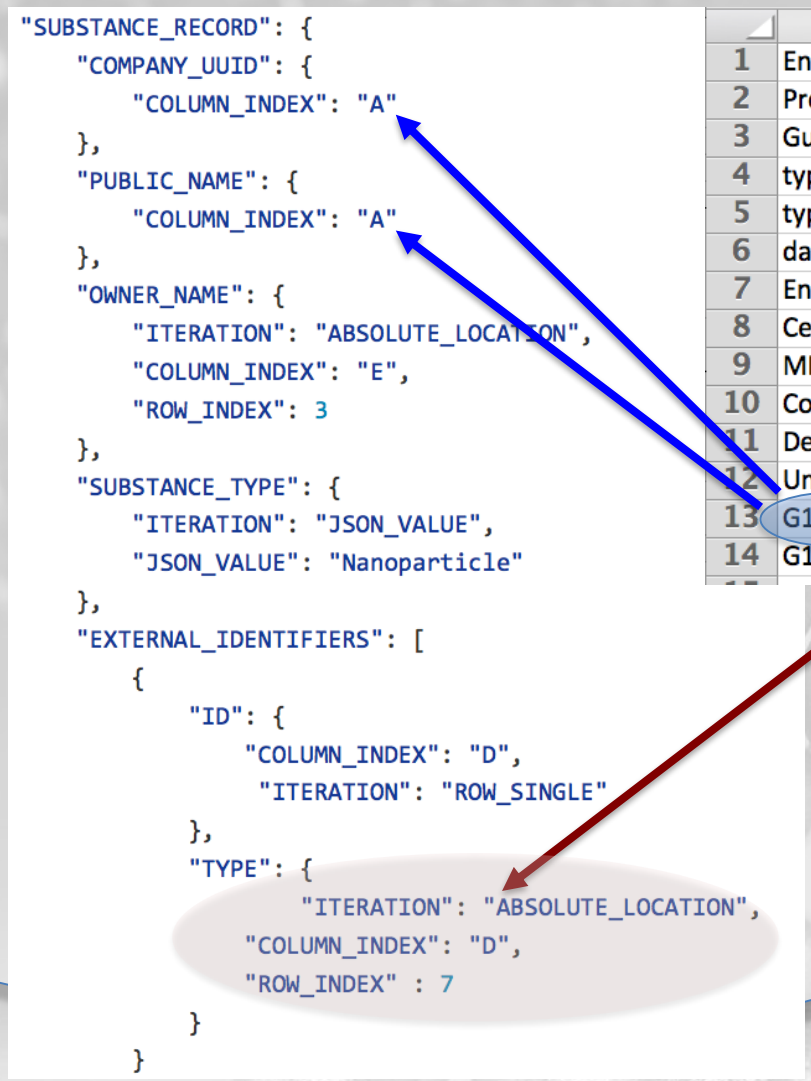
Spreadsheet serializations

- Advantages:
  - tools to edit are abundant
  - clear visual representation of the data
- Disadvantages (that can be overcome):
  - very free format
  - need standardization

Supported Templates

Row Labels	Column Labels: Count of Template(s) Confidential	Count of Template(s) Public	Count of Data file(s) Confidential	Count of Data file(s) Public
Under development				
NanoPUZZLES		24		24
Richard Marchese Robinson		24		24
Supported				
MARINA		8		
Peter Ritchie		8		
MODENA		1		1
Andrea Haase		1		
Thomas Puzyn				1
ModNanoTox				1
IST				1
NanoREG		23		1
Bengt Fadeel				1
Sara Totaro		23		
Included in eNanoMapper database				
MARINA		12		12
Karolinska Institutet		12		12
Grand Total	32	36	3	37

Excel Files (NanoSafety Cluster)



```
"SUBSTANCE_RECORD": {
    "COMPANY_UUID": {
        "COLUMN_INDEX": "A"
    },
    "PUBLIC_NAME": {
        "COLUMN_INDEX": "A"
    },
    "OWNER_NAME": {
        "ITERATION": "ABSOLUTE_LOCATION",
        "COLUMN_INDEX": "E",
        "ROW_INDEX": 3
    },
    "SUBSTANCE_TYPE": {
        "ITERATION": "JSON_VALUE",
        "JSON_VALUE": "Nanoparticle"
    },
    "EXTERNAL_IDENTIFIERS": [
        {
            "ID": {
                "COLUMN_INDEX": "D",
                 "ITERATION": "ROW_SINGLE"
            },
            "TYPE": {
                    "ITERATION": "ABSOLUTE_LOCATION",
                "COLUMN_INDEX": "D",
                "ROW_INDEX" : 7
            }
        }
```

	A	B	C	D	E	F	G
1	EndpointCa	Core compo	Surface mo	External Ide	Cell.association		
2	Protocol				ICP-AES		
3	Guideline				doi: 10.1021/nn4		
4	type_of_study						
5	type_of_method				ICP-AES		
6	data_gathe	Description			Perkin-Elmer		
7	Endpoint	Element	Abbreviate	Classificatic	Net cell associatic	Net cell ass	Net cell ass
8	Cell				A549	A549	A549
9	MEDIUM						
10	Condition						
11	Designation				Mean	SD	N
12	Units				mL/ug(Mg)	1/ug(Mg)	
13	G15.AC	[Au]	AC	Anionic	0.02751	0.01654	3
14	G15.AHT	[Au]	AHT	Cationic	0.49705	0.08013	3

eNanoMapper

Need: Linking Data

Linking the same nanomaterial between database

- Same material
  - same label and from the same paper
  - same representational material
- Similar material / material classes
  - chemical similarity
  - biological identity similarity

eNM eNanoMapper

Clear Reuse Statement

- The first step in interoperability is availability

Library

Hastings *et al. Journal of Biomedical Semantics* (2015) 6:10
DOI 10.1186/s13326-015-0005-5

JOURNAL OF BIOMEDICAL SEMANTICS

RESEARCH Open Access

eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment

Janna Hastings[1*], Nina Jeliazkova[2], Gareth Owen[1], Georgia Tsiliki[3], Cristian R Munteanu[4,5], Christoph Steinbeck[1] and Egon Willighagen[5]

Abstract

Engineered nanomaterials (ENMs) are being developed to meet specific application needs in diverse domai the engineering and biomedical sciences (e.g. drug delivery). However, accompanying the exciting prolifera novel nanomaterials is a challenging race to understand and predict their possibly detrimental effects on hu health and the environment. The eNanoMapper project (www.enanomapper.net) is creating a pan-Europea computational infrastructure for toxicological data management for ENMs, based on semantic web standar ontologies. Here, we describe the development of the eNanoMapper ontology based on adopting and exte existing ontologies of relevance for the nanosafety domain. The resulting eNanoMapper ontology is available purl.enanomapper.net/onto/enanomapper.owl. We aim to make the re-use of external ontology content se and thus we have developed a library to automate the extraction of subsets of ontology content and the as the subsets into an integrated whole. The library is available (open source) at http://github.com/enanomap slimmer/. Finally, we give a comprehensive survey of the domain content and identify gap areas. ENM safet boundary between engineering and the life sciences, and at the boundary between molecular granularity a granularity. This creates challenges for the definition of key entities in the domain, which we also discuss.

Keywords: Nanomaterial, Safety, Ontology

Background

Nanomaterials are materials in which the individual components are sized roughly in the 1-100 nanometer range in at least one dimension, although an exact definition is still being debated [1,2]. Particles in this size range display special properties having to do with their very large ratio of surface area to volume [3]. Natural nanomaterials include viral capsids and spider silk. Recent years have seen an explosion in the development of engineered nanomaterials (ENMs) aiming to exploit the special properties of these materials in various domains including biomedicine (e.g. as vehicles for drug delivery), optics and electronics [3].

Counterbalancing the many possible benef oped nanotechnology, nanoparticles also p risks to human and environmental health [4]. these dangers, regulatory bodies are calling atic and thorough toxicological and safety in into ENMs with the objective of feeding kno predictive tools which are able to assist re designing safe nanomaterials. Evaluating an the possible dangers of different nanomateri assembling a wealth of information on those the composition, shape and properties of th nanoparticles, their interactions with biolog across different tissues and species, and th behaviour into the natural environment. Th arising from different disciplines with highl neous requirements, methods, labelling an practices. Regulatory descriptions of ENMs those needed for nanoQSAR analyses. Saf ments may also vary under different condition

*Correspondence: hastings@ebi.ac.uk
[1] European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom
Full list of author information is available at the end of the article

BEILSTEIN JOURNAL OF NANOTECHNOLOGY

The eNanoMapper database for nanomaterial safety information

Nina Jeliazkova[*1], Charalampos Chomenidis[2], Philip Doganis[2], Bengt Fadeel[3], Roland Grafström[3], Barry Hardy[4], Janna Hastings[5], Markus Hegi[4], Vedrin Jeliazkov[1] Nikolay Kochev[1,6], Pekka Kohonen[3], Cristian R. Munteanu[7,8], Haralambos Sarimveis Bart Smeets[7], Pantelis Sopasakis[2,9], Georgia Tsiliki[2], David Vorgrimmler[10] and Egon Willighagen[7]

Full Research Paper

Address:
[1]Ideaconsult Ltd., Sofia, Bulgaria, [2]National Technical University of Athens, School of Chemical Engineering, Athens, Greece, [3]Karolinska Institutet, Stockholm, Sweden, [4]Douglas Connect GmbH, Zeiningen, Switzerland, [5]European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom, [6]Department of Analytical Chemistry and Computer Chemistry, University of Plovdiv, Plovdiv, Bulgaria, [7]Department of Bioinformatics, NUTRIM, Maastricht University, Maastricht, The Netherlands, [8]Computer Science Faculty, University of A Coruna, A Coruña, Spain, [9]IMT Institute for Advanced Studies Lucca, Lucca, Italy and [10]In silico toxicology GmbH (IST), Basel, Switzerland

Email:
Nina Jeliazkova* - jeliazkova.nina@gmail.com

* Corresponding author

Beilstein J. Nanotechnol. **2015,** *6,* 1609–1634.
doi:10.3762/bjnano.6.165

Received: 31 March 2015
Accepted: 03 July 2015
Published: 27 July 2015

This article is part of the Thematic Series "Nanoinformatics for environmental health and biomedicine".

Guest Editor: R. Liu

Keywords:
database; EU NanoSafety Cluster; nanoinformatics; nanomaterials; nanomaterials ontology; NanoQSAR; safety testing

Abstract

Background: The NanoSafety Cluster, a cluster of projects funded by the European Commision, identified the need for a c tional infrastructure for toxicological data management of engineered nanomaterials (ENMs). Ontologies, open standards, ar operable designs were envisioned to empower a harmonized approach to European research in nanotechnology. This provides a number of opportunities and challenges in the representation of nanomaterials data and the integration of ENM i tion originating from diverse systems. Within this cluster, eNanoMapper works towards supporting the collaborative safety ment for ENMs by creating a modular and extensible infrastructure for data sharing, data analysis, and building computatio cology models for ENMs.

Results: The eNanoMapper database solution builds on the previous experience of the consortium partners in supporting data through flexible data storage, open source components and web services. We have recently described the desig

News from the NSC Projects

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.

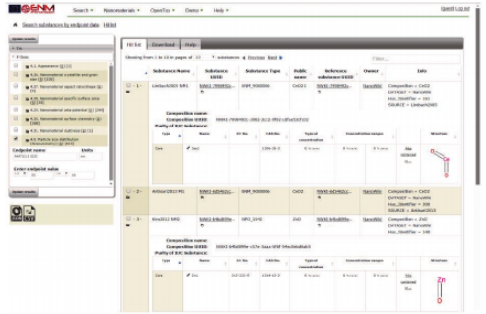


Figure 1. Search functionality in the eNanoMapper database software, showing the results for a search on nanomaterials with a size between 50 and 60 nm