DELIVERABLE REPORT D2.3

Ontology Initial Release

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:	EMBL-EBI
DATE:	15.1.2015
VERSION:	V 1.0

SEVENTH FRAMEWORK PROGRAMME

Call identifier	FP7-NMP-2013-SMALL-7
Document Type	Deliverable Report
WP/Task	WP2 – Task 2.4.
Document ID	eNanoMapper D2.3.
Status	Final

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	Authored by Janna Hastings, Egon Willighagen, Gareth Owen. Reviewed by Barry Hardy (DC).
Purpose of the Document	Report on the initial release of the ontology
Document History	1.First draft, 06/11/2014 2. Second draft, 05/12/2014 3. Third draft, 15/01/2015 4. Fourth draft, 26/01/2015 5. Completed version, 30/01/2015

TABLE OF CONTENTS

TABLE OF FIGURES

GLOSSARY

Abbreviation / acronym	Description
ENM	engineered nanomaterial
OWL	Web Ontology Language
BFO	Basic Formal Ontology
IAO	Information Artifact Ontology
NPO	NanoParticle Ontology
CHEMINF	Chemical Information Ontology
BAO	BioAssay Ontology
OBI	Ontology for Biomedical Investigations
PATO	Phenotype and Trait Ontology
UO	Unit Ontology
ENVO	Environment Ontology
ChEBI	Chemical Entities of Biological Interest
IRI	International Resource Identifier

1. EXECUTIVE SUMMARY

The eNanoMapper project aims to build an ontology and database to collate and describe data relevant for “safe by design” engineered nanomaterial development. Work Package 2 of this effort is developing and disseminating a comprehensive ontology for the nanosafety domain, encompassing nanomaterials and all information relating to their characterization, as well as relevant experimental paradigms, biological interactions and safety information. This deliverable report describes the initial release of the ontology, including the content and organization, and the technical and curatorial processes that have been followed to create it.

2. INTRODUCTION

Nanomaterials are materials in which the units have at least one dimension sized in the 1-100nm range. In addition to the wide diversity of natural nanomaterials available, advances in chemical synthesis techniques have led to an explosion in the number of engineered nanomaterials (ENMs) in recent years. Materials with structures in the nanoscale range often have unique optical, electronic, and mechanical properties, and as a result ENMs are being developed to meet specific application needs in diverse domains across the engineering and biomedical sciences (e.g. drug delivery). However, accompanying the proliferation of nanomaterials is a challenging race to understand and predict their possibly detrimental effects on human health and the environment.

The eNanoMapper project (www.enanomapper.net) is creating a pan-European computational infrastructure for toxicological data management for ENMs, based on semantic web standards and ontologies. eNanoMapper aims to develop a comprehensive ontology and annotated database for the nanosafety domain to address the challenge of supporting the unified annotation of nanomaterials and their relevant biological properties, experimental model systems (e.g. cell lines), conditions, protocols, and data about their environmental impact. Rather than starting afresh, the developing ontology is building on existing work, integrating existing ontologies in a flexible pipeline. The establishment of a universal standardisation schema and infrastructure for nanomaterials safety assessment is a key project goal, which will catalyse collaboration, integrated analysis, and discoveries from data organised within a knowledge-based framework. This framework will support the discovery of nanomaterial properties responsible for toxicity, and the identification of toxicity pathways and nano-bio interactions from linked datasets, ontologies, 'omics data and external data sources.

Ontologies are structured controlled vocabularies enhanced with explicit formal relationships between entities in support of advanced automated reasoning for inference and error detection. **Work Package 2** of the eNanoMapper project focuses on the development and dissemination of a comprehensive ontology for the nanosafety domain, encompassing nanomaterials and all information relating to their characterization, as well as relevant experimental paradigms, biological interactions and safety information.

This report describes the initial release of the eNanoMapper ontology. The ontology has been created and made available in a public version control repository (GitHub) as an OWL language file. Parts of the content are re-used from existing ontologies, for which we have implemented a fully automated scripting process in order to reduce manual overhead while staying up to date with the evolution of source ontologies. Other parts of the ontology have been manually curated by the eNanoMapper team. In some cases, we have contributed content back to the source ontologies we re-use (thus helping to ensure longer-term sustainability of the full suite of interoperating ontologies). Each of these aspects is described in this report, starting with an overview of the content and organization of the resulting ontology, through the development and maintenance processes, and finally the availability and use of the ontology.

3. ONTOLOGY CONTENT

For the M12 initial release of the ontology, it contains 4,454 classes, the bulk of which have been imported from externally available existing ontologies following the custom 'slimmer' import process described in Section 4 below.

These classes are organized beneath a light-weight organizing upper level which is a subset of the Basic Formal Ontology (BFO, Grenon & Smith 2004). The root of this ontology is the most generic class `entity', which denotes any type of thing in reality, which we also adopt as our most upper term. The next level of the classification then includes the classes 'disposition', 'material entity', 'process' and 'quality'. The additional upper level class 'information content entity' is taken from the Information Artifact Ontology (IAO). This slim-line upper level is illustrated in Figure 1.

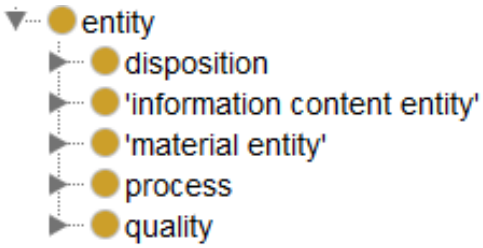

Figure 1 The light-weight upper level for the eNanoMapper ontology

Beneath each of these organizing classes, relevant content from different ontologies has been assembled. An overview of the different sections of the content and the ontologies that they have been imported from is presented in Figure 2.

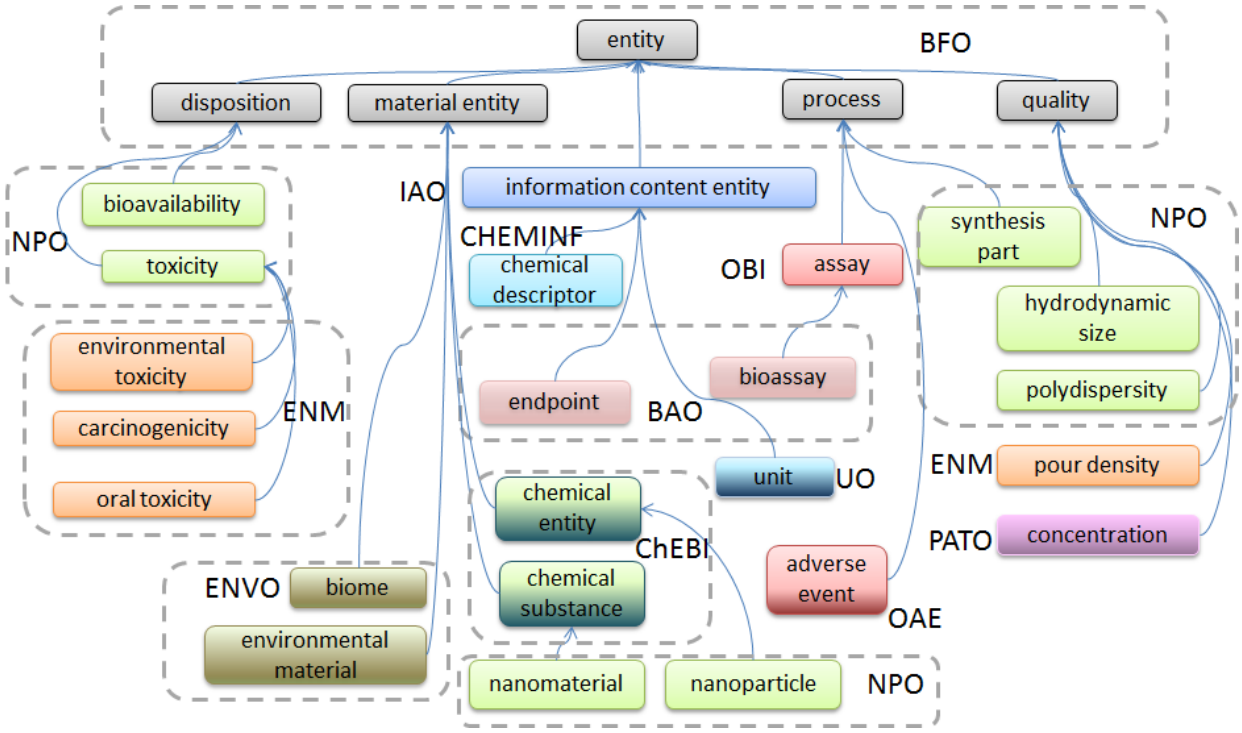

Figure 2 An overview of content, with examples, in eNanoMapper, showing the external ontologies from which classes were imported. Manually annotated content is tagged as 'ENM'.

3.1. DISPOSITIONS

Dispositions include classes such as 'amphiphilic', 'biodegradable' and 'zeta potential' (imported from the NanoParticle Ontology, NPO, Thomas et al., 2011). The class 'toxicity' from the NPO is also included. The NPO, however, contained only a small subset of toxicity classes. For this reason, the toxicity classes have been manually extended with content annotated in the eNanoMapper namespace (http://purl.enanomapper.org/onto/ENM_00000xx). The classes which are imported from NPO include 'immunotoxicity' and 'cytotoxicity'. Those which are manually annotated include 'toxicity to microorganisms', 'genetic toxicity' and more. A detailed subset of the toxicity classes included in the resulting composite ontology is illustrated in Figure 3. Notice that the integration is seamless between the imported content and the manually annotated content.

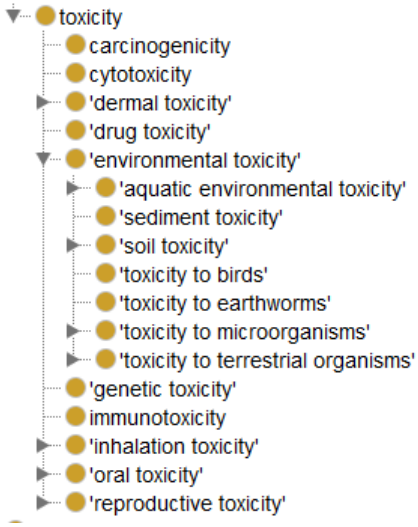

Figure 3 A view over a subset of the toxicity classes included in the eNanoMapper ontology

4.2. INFORMATION CONTENT ENTITIES

Information content entities are units of information of given types which are typically about something. Note that even though we have included it as an upper level grouping class, information content entity is not itself included directly in BFO but rather in the dedicated ontology IAO. There are many different types of information content entity of relevance in nanomaterial safety research. Various different types of chemical descriptors are typically used in the characterization of nanomaterials. These are imported to eNanoMapper from the CHEMINF ontology (Hastings et al., 2011). They include 'atom count', 'gravitational index', 'polar surface area', 'logP', 'vapour pressure', 'melting point' and others. Descriptors are typically the result of measurements or calculations. Some additional descriptors needed for nanomaterial safety research have been contributed back to CHEMINF. Furthermore, 'endpoint' is included with content primarily imported from the BioAssay Ontology (BAO, Vempati et al., 2012). Types of endpoints include concentration response endpoints such as IC80 and LC50, time endpoints such as Tmax, percentage endpoints such as 50 percent activation, and physical endpoints such as pKa. Some of the endpoints are also included in CHEMINF as descriptors; in due course these will be disambiguated with overlaps resolved by means of explicitly removing one class in the slimming process while annotating one or more additional synonyms to the remaining class to ensure the same level of knowledge representation. In addition to the imported endpoints, a small number of toxicology-relevant endpoints have been manually annotated. These include 'skin sensitization'. Another type of information content entity included in the eNanoMapper ontology is 'parameter', which is imported from the NPO. The NPO classes included in this section partly overlap with other classes included from other sources. In due course, these are to be merged with the descriptor and endpoint classes. Finally, a wide and comprehensive range of units are imported from the Unit Ontology (UO, Gkoutos et al., 2012) classified beneath 'information content entity'. These include nanometer, kilobyte, picovolt and so on. Some of the imported types of unit are illustrated in Figure 4.

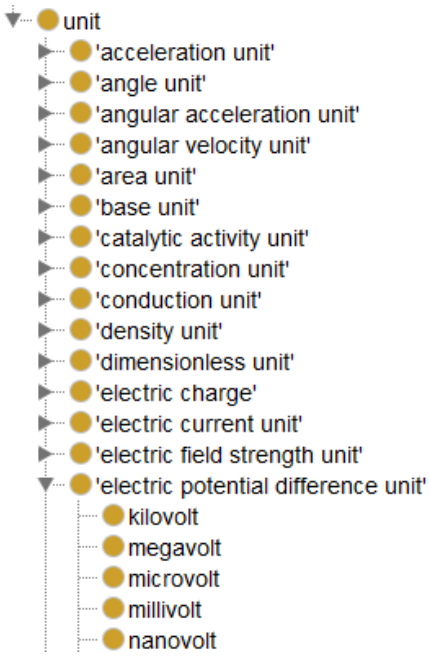

Figure 4 Some of the imported types of unit in the eNanoMapper ontology

4.3. MATERIAL ENTITIES

Beneath 'material entity' we find various different types of classes. We include, of course, classes for types of nanomaterial – e.g. 'engineered nanomaterial', 'nanotube', 'metal oxide nanoparticle', biodegradable nanoparticle' and 'surface-functionalized nanoparticle'. The bulk of the classes of nanomaterial and nanoparticle type are imported from the NPO, but some are also imported from ChEBI (Hastings et al., 2013). The components of nanomaterials are also included, containing classes for nanoparticle parts such as shell and core.

In order to adequately annotate experiments describing the environmental impact of nanomaterials, we have included a subset of the Environment Ontology (ENVO, Buttigieg et al., 2013) beneath material entity. These include, for example, environmental materials such as soil, rock and sand, and whole biomes such as the terrestrial and the aquatic biome.

The instruments which are typically used in nanomaterial characterization or safety experiments are also included beneath material entity. These classes can be used in annotation of experimental protocols where the instrumentation is an important component. They are imported from the BAO and from the NPO. A subset of the instrument classes is illustrated in Figure 5.

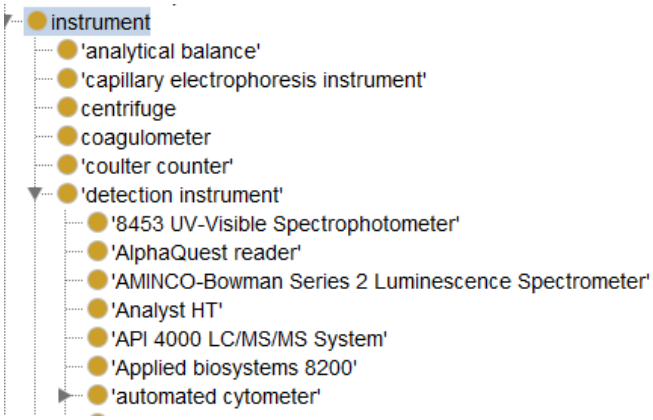

Figure 5 A subset of instrument classes included in eNanoMapper

4.4. PROCESSES

Beneath 'process' the classes typically describe experimental methods that are used in either the synthesis of or testing the properties or interactions of nanomaterials. Here we find the 'assay' class that synthesizes the assay descriptions from the OBI (Brinkman et al., 2010) and the BAO ontologies. We also find the process of functionalization of a nanoparticle from the NPO, and synthesis methodology also from the NPO. Types of assays are quite extensive, with the different types of assay covering the range of biomedical experimentation. Toxicology assays are not very highly represented, although there are some; in the future we may manually annotate these types of experiments as needed for our data annotation activities. Some of the assay types included in the ontology are illustrated in Figure 6.

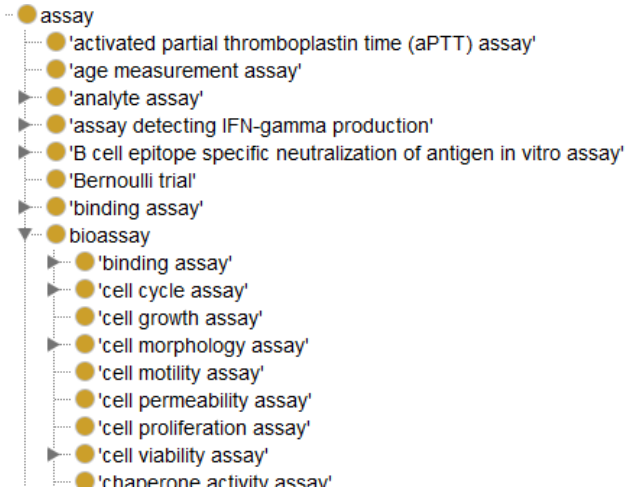

Figure 6 A small subset of the types of assay included in eNanoMapper from OBI and BAO

To enable safety-relevant annotation we have furthermore included a subset of the ontology for adverse events (OAE, He et al., 2014) which captures medically relevant adverse events that may occur on exposure to a given substance.

4.5. QUALITIES

The final upper-level term is 'quality', and beneath this class are various intrinsic properties of chemical entities such as nanoparticles, including mass, porosity, shape and size. Many of these qualities are

imported from the PATO ontology, others are imported from the NPO. We have manually added content to eNanoMapper to represent additional properties of nanoparticles not yet covered by the NPO classes, such as 'dustiness'. We anticipate these classes may in due course be assimilated into the NPO.

In this initial release, we have focused only on including classes needed for annotation of data. We have imported the relationships used in our source ontologies, but not yet attempted to merge and restructure the use of relationships to arrive at a coherent and unambiguous set of relationships. These improvements are due to follow in a subsequent release. Similarly, we have imported the annotations on the classes as they were present in the original ontologies, and not yet attempted to harmonize between, e.g. use of rdfs:comment for definitions and use of iao:definition for definitions.

During our upcoming reviews and revisions we will focus on enhancing the coverage and the quality of the ontology offering as well as putting the ontology to use for (e.g.) annotation of real data. All feedback to the ontology will be collected via the GitHub issue tracker as described in D2.1.

4. DEVELOPMENT AND MAINTENANCE

4.1. ONTOLOGY SLIMMER FOR EXTERNAL CONTENT

External content is included from pre-existing ontologies including the NPO, UO and others as surveyed in our earlier report D2.2. In order to minimize the cost of re-use of content but allow flexibility in how that content is arranged in the composite ontology, as well as be able to seamlessly stay up-to-date with source ontology releases, we developed an ontology slimming library. The library extracts subsets of existing ontologies according to a defined set of instructions, which basically consist of IRIs to include (with or without descendants), IRIs to exclude (by default the remainder of the ontology is excluded other than what is explicitly included), and finally the ability to specify superclass IRIs for subset contents which may be in other ontologies. It is the latter facility that allows us to use the slimmer library to compose the external content like jigsaw puzzle pieces into an overarching whole.

For example, the instruction IRI file for the NanoParticle Ontology import is reproduced in Figure 7.

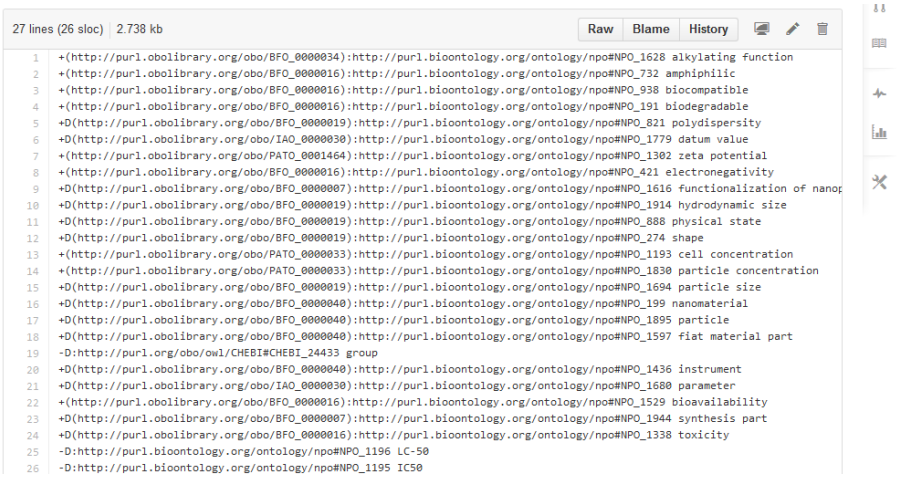

Figure 7 The IRIS file giving the instructions for the import of the NanoParticle Ontology

This instruction specifies that classes such as 'biocompatible' and 'particle' should be included with their descendants and mapped to parent classes within the light-weight upper level (itself a subset of BFO). Furthermore, some classes are explicitly excluded, when they appear in branches that are otherwise imported, such as 'IC50' which is removed here in an effort to prevent the overlap with BAO that would otherwise occur.

As a general rule, we have attempted to only include external content that is "primary" in the given ontology. If they in turn import secondary content, then we exclude that from our import, rather importing it as a module from the source ontology itself. For example, we exclude BFO, UO, ChEBI and PATO content that is included in NPO; instead we import modules from BFO, UO, ChEBI and PATO directly. We believe that this approach is much easier to maintain as we do not duplicate and perpetrate "stale" content. Of course, the automation of the slimming is key to our solution to sustainability – if all this picking and choosing branches and leaves from different ontologies had to be done manually it would have been prohibitively time consuming.

To achieve a unified whole ontology, each imported branch is assigned a classification parent which is chosen as the "best fit" for the content of the branch, harmonizing across different ontologies where applicable. The included content needs to be classified beneath at least one of the BFO upper level terms (disposition, material entity etc.) and better yet if it can be included in an even more specific location beneath a class from another imported ontology.

A primary objective of this exercise is to reduce and remove duplication across the different ontologies to achieve a unified whole. If the various ontologies – NPO, BAO, ChEBI, OBI, etc. – are imported into a composite whole without this type of targeted scripting approach, the resulting duplication is immense, breaking tools that rely on unique label-to-id mappings and significantly reducing the usability of the result for curators. We use our duplication checking script (described in D2.1) to check for duplicates across the imported set of ontologies. If we find duplication we specifically exclude one of the duplicated IRIs from the import of that ontology's module. We need to make a choice which to keep and which to exclude. For example, as described above, we exclude 'IC50' from the NPO import because it is present in BAO. It should be noted that this decision is often not easy, as different ontologies have different advantages and disadvantages. One advantage of NPO is that it is well annotated with metadata, including multiple synonyms and a lengthy, helpful definition for almost all classes. BAO, by contrast, is metadata-light with almost no synonyms and many missing definitions. In this case, therefore, by favouring the BAO class for IC50 we lose metadata. However, this is compensated by the fact that BAO contains many – most – of the relevant endpoint classes for biological experiments, while NPO has relatively few. Part of the effort involved in developing a reuse-oriented ontology such as eNanoMapper is in submitting content and requests to contributing source ontologies to improve their offerings. To this end we hope to work with the BAO developers to improve their representation of metadata.

4.2. CURATED AND CONTRIBUTED CONTENT

The automated content extracted from the slimmer is linked to from the file enanomapper-auto.owl. This imports all the content that is included in subset (slim) files in the external/ folder. On the other hand, manual content is curated into files in the internal/ folder and is assembled in the top-level enanomapper.owl file. The separation is important because the –auto file and relevant subsets may be completely overwritten at all times so keeping any manual content in them would be problematic.

So far, we have manually annotated toxicological endpoints and toxicity classes in the file internal/endpoints.owl. Each class has at minimum an eNanoMapper ID (in the ENM_00000xx namespace), a unique label, and a textual definition. Some classes have additionally been allocated multiple synonyms.

We anticipate a substantial portion of the content we annotate will not be direct in the eNanoMapper ontology but rather contributed back to the source ontologies. For CHEMINF, for which eNanoMapper project partners are also contributors, we have already started to directly annotate nanomaterial relevant descriptors. We aim to regularly prepare contributions back to NPO as well.

5. AVAILABILITY AND USE

The ontology is located at http://purl.enanomapper.org/onto/enanomapper.owl and may be downloaded in full from the development area at https://github.com/enanomapper/ontologies. All licensing information for the source ontologies and the eNanoMapper ontology is reproduced in that development area – we have only built on open source components and our ontology is also open source (CC-BY 3.0). The testing reports for the ontology building workspace are available in our Jenkins installation at http://jenm.bigcat.maastrichtuniversity.nl/.

The ontology file may be downloaded and opened using the Protégé tool as described in the eNanoMapper report "Framework and Infrastructure for Ontology development, versioning and dissemination" (D2.1). In addition, for ease of reference for the general audience we will submit the ontology to BioPortal which offers online browsing and searching. We also plan to publicize the released ontology via the dissemination channels in use for the project, e.g. the project website, Twitter account etc.

We are now entering the first review phase (internal review) and soon will be soliciting formal feedback from the community (external review).

6. CONCLUSION

The initial release of the eNanoMapper ontology is an important milestone in the overall project, since it allows the community early insight into our efforts, methodology and results, such that we can reach the widest possible audience of users and make sure that we address their concerns based on their direct experiences with using the ontology. As noted in this report, the initial ontology release represents a starting point rather than being itself a complete solution to the needs of the domain, and the next months and years will accordingly see solid, user and annotation-driven growth and improvement in the ontology. We have scheduled an internal review process (with accompanying ontology modifications) followed by an external review process, but notwithstanding the schedule we will already gladly accept feedback from any interested member of the community.

7. BIBLIOGRAPHY

Brinkman, R.R., Courtot, M., Derom, D., Fostel, J.M., et al. (2010) Modeling biomedical experimental processes with OBI. J Biomed. Semantics 22;1 Suppl. 1:S7.

Buttigieg, P.L., Morrison, N., Smith, B., Mungall, C.J., Lewis, S.E. and the ENVO Consortium (2013), The environment ontology: contextualising biological and biomedical entities. Journal of Biomedical Semantics 4:43.

Grenon P. and Smith B. (2004) SNAP and SPAN: Towards Dynamic Spatial Ontology. Spatial Cognition and Computation, 4:1, 69-103.

Gkoutos G.V., Schofield P.N. and Hoehndorf R. (2012) The Units Ontology: a tool for integrating units of measurement in science. Database 2012 : bas033.

Hastings, J., Chepelev, L., Willighagen, E., Adams, N., Steinbeck, C. and Dumontier, M. (2011) The Chemical Information Ontology: provenance and disambiguation for chemical data on the biological semantic web. PLoS ONE, 6(10): e25513.

Hastings J., de Matos P., Dekker A., Ennis M., Harsha B., Kale N., Muthukrishnan V., Owen G., Turner S., Williams M. and Steinbeck C. (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res. 41:D456-63.

He Y., Sarntivijai S., Lin Y., Xiang Z., Guo A., Zhang S., Jagannathan D., Toldo L., Tao C. and Smith B. (2014) OAE: The Ontology of Adverse Events. Journal of Biomedical Semantics 2014, 5:29.

Thomas, D.G., Pappu, R.V., and Baker, N.A. (2011) NanoParticle Ontology for Cancer Nanotechnology Research. J Biomed In-form. 44(1): 59–74.

Vempati, U.D., Przydzial, M.J., Chung, C., Abeyruwan, S., et al. (2012) Formalization, annotation and analysis of diverse drug and probe screening assay datasets using the BioAssay Ontology (BAO) PLoS One 7(11):e49198.