



**Proposal Task Force by: Thomas Exner, +49 171 3807485,
Thomas.exner@edelweissconnect.com**

Proposal: Ontology Inventory

<p>Problem definition</p>	<p>Problem description: Currently there are increasing needs of high quality data to feed into the evolving nanoinformatics workflows and for use in a regulatory and policy context. To achieve this, datasets need to be findable online and machine readable and able to be combined to larger datasets to be reused for the identification of hidden patterns and relationships as demonstrated, among others, by Labouta et al. (2019) during the meta-analysis of 93 peer-reviewed papers. Thus, datasets need to be annotated using established ontologies, along with communication between the various nanosafety databases/repositories. Currently, there is a number of parallel ongoing activities within the NSC projects for the development of data dictionaries (collections of terms), taxonomies and ontologies, which are focussed in a field-specific context. This potentially results in duplication of work and lack of communication between the various efforts, leading to lack of harmonisation and communication and to huge effort and time being wasted.</p> <p>Aims of activity:</p> <ul style="list-style-type: none"> • Generation of an inventory on the ongoing NSC projects and international activities to develop collections of terms, taxonomies and ontologies for dataset annotation and identification of limitations and additional needs. • Identification on the similarities and alignment potential to promote a harmonised response of the nanosafety community and promote data interoperability and reusability, as per the requirements of the Open data initiative of the European Commission.
<p>Stakeholder</p>	<ol style="list-style-type: none"> 1) Projects and individuals working on harmonization of data and the semantic annotation of datasets and are in need of extended dictionaries and ontologies to represent their results in a harmonized and interoperable way 2) Data curates, database managers and developers of databases and data management tools working on improved functionality for semantic annotation of datasets before or during upload, linking and enrichment of data across different data warehouses 3) Nanoinformatics experts in need of high quality datasets for the development and publication of better predictive tools and for the

	<p>implementation of the recommendations of the EU-US Roadmap Nanoinformatics 2030 (e.g. NanoSolveIT, NanoInformaTIX projects)</p>
<p>Approach</p>	<p>The open data initiative of the EU as well as the commitment of the nanosafety community to FAIR data sharing and the emergence of nanoinformatics is imposing strong and new requirements on the management of data. Only by allowing harmonized and interoperable searches, retrieval and enrichment across all nano-related databases, the produced results from the publicly funded projects can be exploited in a time-efficient way to guide risk assessment, registration and regulation of nanomaterials, the introduction of safe-by-design concepts in nanomaterial development and the development of high-quality and robust nanoinformatics tools for the implementation of the EU-US Roadmap Nanoinformatics 2030, driven forward, currently, by the NanoSolveIT and NanoInformaTIX projects.</p> <p>A technical requirement to allow the automatization of such search and retrieval processes is the semantic annotation of the data to be able to find the needed information in one database and link/enrich data across databases. Even if there are different tools and templates to support data annotation, the main burden of annotation is imposed onto the data providers and curators, which needs to be supported by training on the use of ontologies and relevant ontology browsers, look-up services and data annotations tools. Additionally, due to the diversity of data relevant for nanosafety research, existing ontologies and defined terminologies are lacking important terms for example in the area of physicochemical characterization and exposure and fate. Another issue, for existing terms, are field-specific definitions for commonly used terms, making them unusable and non-interoperable in a different context. To provide the support and extend the existing terminology, projects of the NSC have included ontology tasks or even complete ontology work packages in their descriptions of action. Additionally, WG F as well as the infrastructure project NanoCommons are working on the development of general concepts for ontology and semantic annotation.</p> <p>However, we still see a lack in the full, or at all, coordination of the activities. Different projects are creating new ontologies or extending existing ones independently and without informing or collaborating with each other. This leads to duplication of work and, in the worst case, the creation of competing and contradicting ontologies used in different data annotations efforts and data warehouses preventing the combination of the data without large human harmonization efforts. Similar needs for coordination of ontology development has been recognized in communities having longer relevant experience and has e.g. resulted in the creation of the Open Biological and Biomedical Ontology (OBO) Foundry. Therefore, we propose here a new task force with the goal to document the status quo of the ontology development activities in the NSC including expected outcomes (collection of terms, domain or application ontology, annotation tool or template). We will also identify needs for additional expertise and orthogonal</p>

	<p>developments and the alignment potential between the ongoing efforts. Additionally, the responsible persons and their background will be documented to be able to build a list of experts (extendable to neighbouring disciplines and international activities) and link people and projects with similar focus and questions. The task force is not meant to do ontology development itself but will only exist for a short time and formed by a small number of members to contact all projects and collect the information in the form of two questionnaires addressed at the project coordinator and ontology task leader, respectively. This should then be followed up by the creation of expert groups for term collection in specific fields of nanosafety research like exposure, hazard and physchem characterization, for grouping of these terms, build hierarchies and relationships to translate these defined terminologies into ontologies and for keeping these ontologies harmonized, interoperable and complementary. These groups could be either realized as additional task forces or become a wider task of WG F, in collaboration with the rest of the NSC WGs, with the long-term goal to establish a collective of ontology developers, domain experts and users that are committed to collaboration, adherence to shared principles and function as the coordination and standardization body, i.e. the Open (Nano)Material Ontology Foundry.</p> <p>We want to stress here that we see the need for creating this ontology activity inventory and some of the follow-up discussions outside of the normal WG F activities even if they hopefully will also stimulate the work of WG F and attract more participants to its meetings. The cross-WG nature of this effort, can also help re-invigorate activities in other NSC WGs as well. A freshly formed task force can have a critical and objective look at the existing approaches, new and alternative developments in other scientific areas and ontology expert groups and existing tools that are independent from specific requirements and database implementations, which are the main focus of WG F.</p>
<p>Deliverable</p>	<ul style="list-style-type: none"> • Report on the inventory of ontology activities, alignment potential and key experts involved
<p>Duration of the activity</p>	<p>6 months until end November 2020.</p>
<p>Dissemination</p>	<p>The report will be disseminated to the Steering Group and all projects of the NSC to inform the leaders and members of terminology/ontology tasks about similar activities and interlink them but also to provide contract details for requesting support to projects without specific tasks and expertise.</p> <p>To protect personal information, the list of experts will not be shared (only project details will be included in the inventory) but a system will be set up where these experts can contribute by responding to requests or answer questions (implemented as wiki, ticket system or similar) and can register to be informed about follow-up activities and calls for experts.</p> <p>Based on the results, we will look into the possibility of running a relevant webinar to present our findings. The nature of the webinar (NC internal, public) will be determined based on the finding and following consultation with the participating projects.</p>

Contribution by	Thomas Exner Iseult Lynch Dieter Maier Anastasios Papadiamantis Egon Willighagen
Conflict of Interest	None
Task force leader	Thomas Exner
Steering Group	
Steering Group Feedback	Approved
Coordination Team	
CT-Contact to bring it to the SG and to follow up with the proposer	E. Valsami-Jones