



The European Nanotechnology Community Informatics Platform: Bridging data and disciplinary gaps for industry and regulators

Grant Agreement No 731032

Deliverable Report 2.6

Deliverable	D2.6 - Action Plan for Year 3
Work Package	WP2: Community Building
Delivery date	M25 - 31 January 2020
Lead Beneficiary	UK Centre for Ecology & Hydrology (UKCEH)
Nature of Deliverable	Report
Dissemination Level	Confidential (CO)

Submitted by Lee Walker (UKCEH), Beatriz Alfaro Serrano (BNN), Thomas Exner (EwC), Martin Himly (PLUS), Harry Sariveis (NTUA), Laurent Winckers and Egon Willighagen (UM) and Iseult Lynch (UoB)

Revised by Thomas Exner (7P9) and Georgia Melagraki (NovaM)

Approved by Iseult Lynch (UoB)



Table of contents

Summary	3
List of Abbreviations	3
Introduction	4
Action Plan for Year 3	6
1. Ontologies	6
2. Development of a structural identifier for nanomaterials - Nano-InChI	8
3. and 4. Electronic notebooks and Protocols - templates and workflows	9
4. Data warehouse	9
5. Models and modelling tools	10
6. Training	11
7. Interactions with the EOSC and the ESFRI roadmap	13
8. Long-term sustainability of NanoCommons outputs	14
Conclusions	16
References	17
Annex 1 – Topics for interaction with identified EU/national projects	18
Annex 2 - List of EU-funded projects interacting with NanoCommons	19

Summary

This report presents the action plan for year 3 of the NanoCommons project as a continuation of actions planned for years 1 and 2 as part of the description of action in the Grant Agreement. It is guided by the activities successfully started and a SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis performed on the results of the first two years. This analysis showed that the services established as part of the NanoCommons infrastructure address important needs of the community especially with respect to metadata standards, semantic annotation of datasets for harmonisation of terminology prior to upload of datasets into NanoCommons KnowledgeBase, electronic lab notebooks (ELNs), data storage solutions and FAIR (Findable, Accessible, Interoperable and Re-usable) data sharing as well as nanoinformatics and risk assessment workflows.

User acceptance and uptake was demonstrated with multiple Transnational Access (TA) projects (12 applications received, 11 accepted, 1 transferred to WP8: NA2 - Training aligned to TA / JRA) based on the solutions provided by NanoCommons in these areas. However, gaps and areas needing strengthening were also identified and were added as foci for year 3 in the action plan. This includes continuous and increased support of ontology development and semantic annotation including new tools, development of a line notation as structural representation of nanomaterials and separate different nanoforms, increase training activities especially for young researchers and outreach to more projects and initiatives in the nanosafety area but also to other infrastructures organised and integrated into the European Open Science Cloud (EOSC).

List of Abbreviations

EC - European Commission

ELNs - electronic lab notebooks

EOSC - European Open Science Cloud

ESFRI - European Strategy Forum on Research Infrastructures

EUON - EU Observatory for Nanomaterials

FAIR - Findable, Accessible, Interoperable and Re-usable

H2020 - Horizon 2020

InChI - (IUPAC) International Chemical Identifier

JRA - Joint Research Activities

NA - Networking Activities

NInChI - Nanomaterials extension to InChI (to be developed by NanoCommons)

NMs - nanomaterials

NPO - NanoParticle Ontology

OECD - Organisation for Economic Cooperation and Development

PBPK - physiological based pharmacokinetic (modeling)

RDF - Resource Description Framework

SMILES - Simplified Molecular-Input Line-Entry System

SWOT - Strengths, Weaknesses, Opportunities and Threats

TA - Transnational Access

YNS - Young NanoScientists group of the European NanoSafety Cluster

Introduction

To be considered a successful research infrastructure, NanoCommons needs to demonstrate its value to the community, and its ability to address end-user and stakeholder needs in a cost-effective and streamlined manner. Based on our initial survey of the market needs during the proposal development, period one activity (2018-2019, years 1 and 2) has focused on establishing the core infrastructure and communication with the wider stakeholder community, and the roll-out of the 1st call for Transnational Access (TA) in the first part of 2019. To be able to flexibly react on the community uptake and the stakeholder feedback on the implemented infrastructure layout and the TA offerings, the activities for the second part of the project were only conceptually described in the project's description of action and two deliverables were integrated, which should report on the detailed plans for the following year 3 and 4, respectively. The report presented here is the first of these reports and is outlining the plan of action for year 3, that supplements what was already planned as part of the core developmental work with additions and adaptations to include the learnings from the first two years and address and focus on the most pressing needs of the community at the time of writing.

For the purposes of NanoCommons, as a research infrastructure, *Users with a capital U* are defined as external individuals who apply for access to a specific service or tool via the 6-monthly calls for TA. The community more broadly consists of developers of tools and services that could be integrated into NanoCommons to enhance the service provision and provide end-users of the services and tools with a better experience, and stakeholders for the nanosafety research e-infrastructure, which span the nanosafety research community, nano-enabled industry, regulators and the wider ecosystem including inter-governmental organisations such as the Organisation for Economic Cooperation and Development (OECD). Figure 1 shows an overview of these groupings and their interactions with and relationship to NanoCommons.

NanoCommons has developed a number of routes via which stakeholders and the wider community can interact with the NanoCommons team. The Helpdesk for TA Users was described in Deliverable Report D7.1, while the GitHub Issue tracker, targeting software and modelling tool developers, and the Community suggestion box, targeting experimental users and wider stakeholder groups, are described and presented in Deliverable Report D2.3.

Based on the ongoing [survey of community needs](#)¹, which can be accessed via a link on the NanoCommons project website, a good overview of additional developmental needs is emerging. A summary of the workshops and events that the NanoCommons consortium has organised and/or participated in can be found in Deliverable Reports D2.1 (1st Annual conference and nano-exploitation day, stakeholder workshop and User call) and D2.5 (Annual conference & nano-exploitation day, stakeholder workshop with TA calls and demonstrations). All usage information and feedback collected via these routes was integrated into a SWOT analysis and resulted in the action plan presented here to react to, and develop solutions to address the identified gaps.

¹ <https://www.surveymonkey.co.uk/r/PK2KXWW>

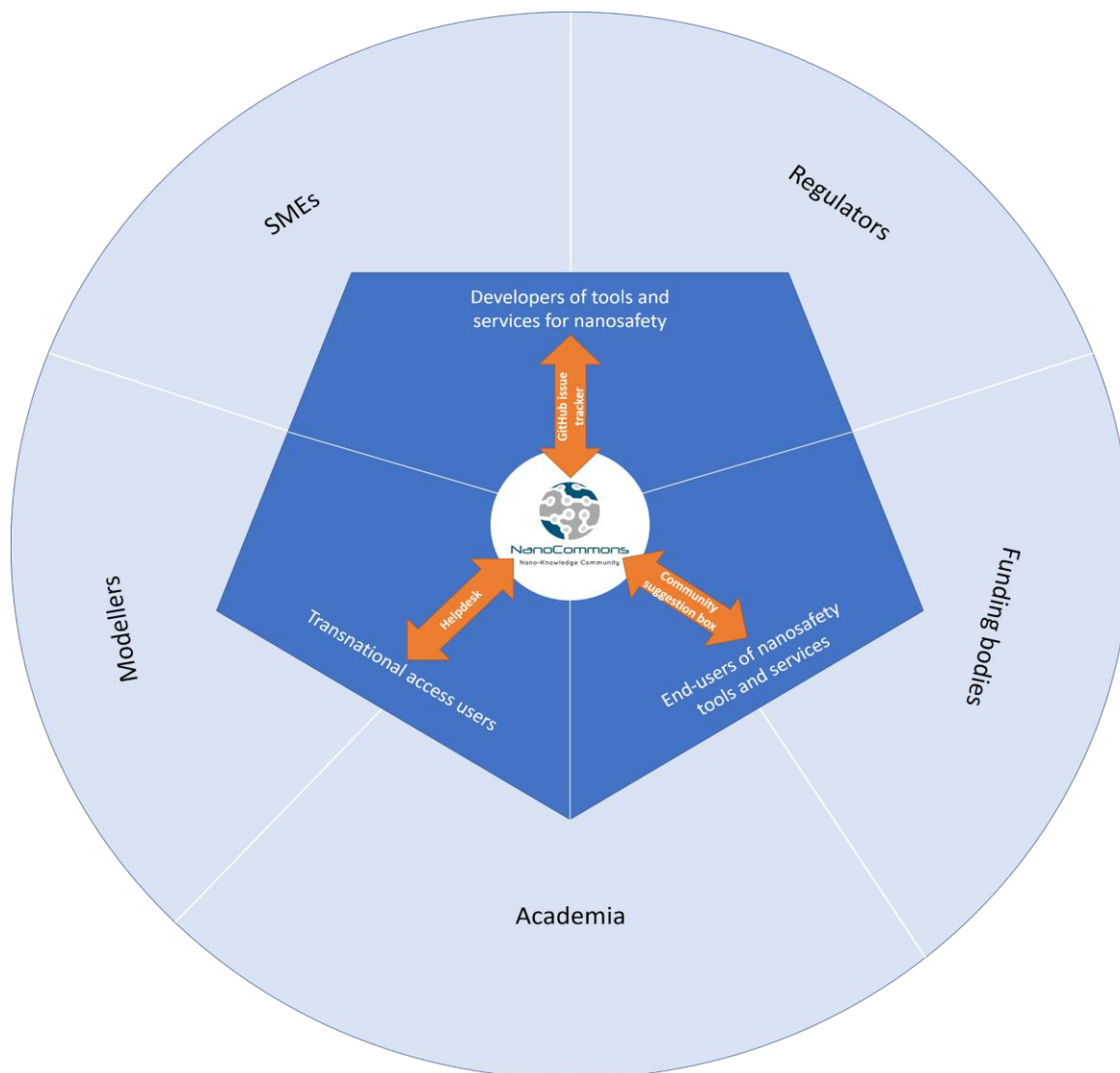


Figure 1. Schematic illustration of the various approaches for communication with developers, end-users and TA users, and the overall placement of these in the wider stakeholder context. [SME – Small to medium enterprise].

Action Plan for Year 3

A number of key activities, in which further developments are needed and which should be strengthened during year 3, have been identified and are described in the following sections.

1. Ontologies

Ontology extension is recognised as the most pressing roadblock in many projects collecting or curating data, and will be a strong focus of year 3 and 4. In the first two years of the project, the clear need to extend existing ontologies to cover the different fields of nanosafety research, as well as for ongoing and even increased training on ontology usage and semantic annotation of data and protocols, became visible. This has resulted in a number of TA applications on ontology related topics, which have been conducted or are ongoing. Therefore, the JRA work and training offers in year 3 and continued in year 4 will be structured to best support the TA projects, and all activities will be funneled into the overarching nanosafety ontology system to optimise the benefits for all by mutually sharing developments and exchanging knowledge.

NanoCommons will work together with different projects of the EU Nanosafety Cluster and international projects to collect the demands regarding additions to the existing ontologies, harmonization of ontologies and improvement of user guidelines supporting usage of ontologies as well as identifying any issues observed in the data structure encoded in the existing ontologies. Strong collaboration with the EU-funded projects [ACEnano²](#), [NanoFASE³](#), [NanoSolveIT⁴](#) and the [US NIKC⁵](#) project, which give valuable feedback on the terms needed in the areas of physicochemical characterization, exposure and fate and assay/methodology/model annotation, will be maintained. NanoCommons has identified the eNanoMapper ontology as being the most advanced ontology for nanosafety currently, and in order to avoid complex mapping schemes between ontologies, submission of the identified missing terms into the [eNanoMapper development pipeline⁶](#) for integration into future versions (NanoCommons-releases) of this ontology has been established as an ongoing activity. This has been successfully demonstrated for the ACEnano terminology and will be continued for all interested projects of the current TAs and future TAs of year 3 and 4. In addition new terms that were added to the eNanoMapper ontology will be migrated to upstream ontologies such as the NanoParticle Ontology (NPO) to make these as inclusive as possible also.

The interaction with the NIKC project team needs specific attention since this project developed the eNanoGrammar as a defined terminology for annotating the NIKC database as part of their data curation effort. Although loosely based on the eNanoMapper ontology, it largely extends the coverage to enable annotation of complex nanomaterial (NM) samples and the developers also saw the need to modify the relationships between some of the different terms compared to eNanoMapper. To make this interoperable with the annotations based on eNanoMapper in these databases, and indeed to update eNanoMapper to support the community needs, an integration of eNanoGrammar into

² <http://www.acenano-project.eu/>

³ <http://www.nanofase.eu/>

⁴ <https://nanosolveit.eu/>

⁵ <https://ceint.duke.edu/research/nikc>

⁶ <https://github.com/enanomapper/ontologies/issues>

eNanoMapper would be preferred or at least a mapping between equivalent terms is needed. Year 3 will be used to develop the concept for such a harmonization and to resolve the inconsistencies between the two encoded data models.

One issue slowing down the data annotation process and therefore the upload of data into the NanoCommons Knowledge Base (or indeed other databases) is that integration of new terminology into the eNanoMapper and underlying ontologies need to be based on community-agreed definitions. This process of consensus forming and aligning it with the release cycles of the ontologies takes time and the data curation process is blocked until the terms are officially released. To circumvent this, a mechanism to create provisional terms on-the-fly was developed and is undergoing implementation in year 3. In this way, the data provider can request a new ontology entry by providing a preferred term with a clear definition and synonyms if appropriate. The system will then generate a request to the ontology developers but also immediately a provisional term, which the data provider can use for annotation of the data allowing them to directly continue with data curation and upload. The development team will check the term, propose it to the community for integration into the ontology and, if consensus is obtained, provide it in the next release. At this point, the provisional term will be linked to the new official term allowing database systems to update the entries accordingly. Besides implementing this workflow, NanoCommons will also provide guidance to the database developers on how to integrate the generation of provisional terms into their ontology lookup services and how to obtain notifications to start the update of the provisional with the official ontology terms.

Supporting data exchange and upload, the Resource Description Framework (RDF) format that uses the eNanoMapper ontology depends not just on classes, but also on ontology properties. These have not traditionally been included in the eNanoMapper ontology. We plan for year 3 to explore an update of the ontology build system to support extraction of properties from third-party ontologies such as *npo:has_component_part* of the NPO.

To support this work on the ontology development, integration, and testing framework, NanoCommons will develop a more sustainable build system for ontology development. The [ROBOT Tool](#)⁷ will be explored as an alternative of the current ontology development and maintenance methods. Furthermore, the [OBO Dashboard](#)⁸ tool will be explored as a quality control tool to implement as part of the ontology development workflow.

All the ontology and annotation tool development will be complemented by training activities, continuing the series of hackathons and training events started in the first half of the project. This training will target two different groups. On the one hand, experimental researchers and especially early-stage researchers will be trained in using ontologies and ontology lookup services to annotate their datasets. This will be done by developing tutorials and other training materials, but even more importantly by providing hands-on sessions (virtual and face-to-face) in which the participants work on their own data (Excel spreadsheets) guided by a group of trainers. On the other hand, the hackathons and hands-on sessions will be continued focusing on the ontology developers but also database developers in the different projects. These will cover how to define and integrate terms in the existing ontology structures, how ontology terms can be combined to avoid the combinatorial

⁷ <https://robotframework.org/>

⁸ <http://dashboard.obofoundry.org/>

explosion of needed terms and how to resolve domain- or database-specific differences in the encoded data models leading to inconsistencies as seen with the example of NIKC.

An EU Curation prize was proposed in the action plan for year 2, which should be awarded to young researchers who have demonstrated excellence in curation of nanosafety datasets and a commitment to making their research data FAIR. Although there was consensus on the need for such an award to promote and emphasise the significance of data curation and management in everyday scientific research and to offer an official qualification to young scientists, technical issues prevented establishing this prize until now. Awarding of the first prize is planned for June/July 2020 in conjunction with hands-on training sessions to support experimentalists in curating their datasets.

2. Development of a structural identifier for nanomaterials - Nano-InChI

Central to the success of small molecule databases and chemoinformatics approaches has been the availability of unique structural representations for each chemical like InChI (Heller et al, 2015) and SMILES (Weininger, 1988). These distinguish one chemical compound from another and facilitate unambiguous mapping of chemical substances from different sources, thus facilitating automation of data retrieval and exploitation. NMs cannot be represented as a simple chemical graph and, as such, pose a conceptual challenge prohibiting the use or adaptation of existing representation to this class of materials. Their three-dimensional (3D) organisation, the close correlation between size and properties, the emergence of quantum properties at sub-30 nm sizes, and their very high surface to volume ratios relative to the bulk mean that any representation of NMs requires information on their chemical composition, physical, and structural features. Description and representation of NMs is further complicated by the fact that many of their properties are dynamic and influenced by their environments, changing with the medium composition (salts, biomolecules), temperature, time, or other external factors. Nevertheless, the demonstrated need for easy identification of relevant data based on an unambiguous encoding of the NM in order to (i) facilitate classification and grouping of nanoforms in regulatory applications and (ii) support the generation of starting structure in nanoinformatics modelling and simulations corresponding to some experimental conditions, renders the development of a structural identifier in form of a 1D line representation one of the major challenges to be addressed by NanoCommons in year 3.

Even if standard InChIs are not able to represent NMs, the underlying concept of layers representing different aspects of the structures is well suited to be extended to NMs, similarly to the ongoing efforts to describe polymers, mixtures and reactions. Starting with a workshop organized by NanoCommons and NanoSolveIT in early 2020, case studies will be defined to collect requirements for an InChI for NMs (NInChI). This will then be further developed in a series of online sessions to define the layers needed to capture NM structures including: core composition, surface topography, surface coatings or functionalization, doping with other chemicals, representation of impurities, NM distributions (size, shape, composition, surface properties, etc.) and crystallographic form. Based on these requirements, a first version defining the NInChI will be proposed and disseminated to the community for refinement in year 4 and alignment with other InChI development activities organized by the [InChI Trust](https://www.inchi-trust.org/)⁹.

⁹ <https://www.inchi-trust.org/>

3. and 4. Electronic notebooks and Protocols - templates and workflows

Throughout year 3 the use of electronic lab notebooks (ELNs) for data collection and annotation will be further promoted both through the implementation of Demonstration Case Study #1 (See deliverable D9.2 for further details) and via training events in summer 2020. These activities will also promote the use of NanoCommons workflows which support users in aligning the data generated with the requirements of the European Commission Horizon 2020 (H2020) Open Data Pilot framework generally, and for nanosafety in particular, and will facilitate the integration of the data into the EU Observatory for Nanomaterials (EUON).

We will explore how the ELNs can integrate / align with the proposed standards for protocols, data and metadata, as well as interfacing with templates and web interfaces for data and metadata upload. In the previous years, four protocol formats (see Deliverable Report D3.2 - Guidance on protocols preparation and repository) have been analysed and evaluated by NanoCommons. Linear or text-based formats are probably more in line with the ELN concept but most often lack computer-readability, comparability and harmonisability. Therefore, options to integrate more structured, box-based workflow or hybrids of the two approaches will be investigated, which will also facilitate a more generic approach that is independent of a specific ELN format and will support the automatic interconversion between formats.

Additionally, activities will be continued to annotate nanosafety protocols including timings and trouble-shooting tips. This will be complemented with exploration of the possibility to directly use such protocol information as metadata attached to each dataset generated by experimental procedures following previously reported protocols or standard operating procedures but annotated with exact values e.g., on the timing of individual steps and specific parameter settings. This work will be performed in accordance with the guidelines established in WP3 and will be presented in Deliverable report 3.1 - *Initial guidance on emerging data and metadata standards*. To encourage the widespread use based on the consensus in the community, the discussion on current and ongoing work to facilitate and improve protocols and the accompanying data capture templates will be extended to external partners from the EU NanoSafety Cluster, the US-EU communities of research and applicants from the TA projects.

4. Data warehouse

The activities started during years 1 and 2 will be continued and expanded in year 3. This includes the integration of existing repositories into NanoCommons, supporting the upload of data not yet publicly available to the NanoCommons data warehouse or another data warehouse integrated into the NanoCommons infrastructure, and development of recommendations for users as to where and how to best store their datasets accompanied with the required metadata. With the public release of the NanoCommons [Knowledge base](#)¹⁰ and knowledge warehouse at the end of year 2, we expect that the data upload support activities will be even more intensified by improved disseminations options using live demonstrations of the benefits of the platform and the meta-search options across all integrated data warehouses. Additionally, user guidelines will be developed, extended and implemented to make data warehouses FAIRer based on clear FAIRness metrics. Papers on metadata standards and

¹⁰ https://ssl.biomax.de/nanocommons/cgi/login_bioxm_portal.cgi

FAIRification will be drafted and published in a special issue (co-edited by NanoCommons coordinator) to disseminate these guidelines to the community and foster their broad implementation.

Showcases for the integration and linking of existing data warehouses into the NanoCommons base (eNanoMapper, ACEnano) and the hosting of project data in the NanoCommons knowledge warehouse (NanoFase) were implemented in the first part of the projects. These will now be used as blueprints for the activities in years 3 and 4 to enlarge the data coverage and reduce data fragmentation. Different TA applications for integrations of data or complete data warehouses have already been received. Additional discussions have been started with project coordinators and the project partners responsible for data management to organise their data curation and FAIRification according to the NanoCommons best-practices workflows. These activities will follow the proposed workflows either for integration of complete databases (see NanoCommons [Deliverable Report D4.5¹¹](#)) or for individual datasets (see NanoCommons Deliverable Report D3.3), where the latter being specifically optimised and customised to address the large amount of data produced in supported research projects.

In parallel, the search/browsing and access options to expoure the datasets in the Knowledge Base will be further optimised to be able to provide these functionalities for data from all integrated databases. In this way, we will establish a central data access point for nanosafety data, which can then be integrated into other knowledge and information resources like the EUON operated by the European Chemicals Agency. EUON is meant to provide a full picture of all available data to their customers and accessible information to the general public, thereby supporting the confidence of citizens and reducing the need for the generation of new data especially from animal experiments.

5. Models and modelling tools

Efforts in year 3 will be invested into development of a generic comprehensive biokinetics R library, which will allow the implementation of any physiological based pharmacokinetic (PBPK) model that can be expressed as a system of ordinary differential equations, and the automatic deployment of the model into the Jaqpot modelling platform. Using existing functionalities, the produced PBPK models will be accessible through both Application Programming Interface calls and user-friendly graphical user interfaces.

The process for integration across models used for adverse outcome pathway-based risk assessment (exposure models, read-across, nanoQSARs, biokinetics and dose-response models) will also be completed, which is presented in Figure 2 in NanoCommons [Deliverable report D6.1¹²](#) in the section entitled “Demonstration of this workflow through a case study”.

The third area of effort will be to continue the development of novel read-across approaches by completing the Apellis workflow, which has been briefly introduced in Deliverable report D6.2 (not yet publicly available). These models will also be presented through custom-made web applications. Extension of the mathematical programming read-across approach to classification prediction problems will be performed. A k-Nearest Neighbours tool will be developed that allows users to upload some minimal information about their specific NM and identify similar materials for which

¹¹ <https://zenodo.org/record/375616>

¹² <https://zenodo.org/record/3603100>

there is already data in the NanoCommons Knowledge Base or its federated databases. This will be used as a means for determining groups (sets) of nanoforms or confirming that the NM is a distinct nanoform for which additional toxicity testing is required.

6. Training

At the beginning of year 3 a number of virtual training tools had been made available through the NanoCommons infrastructure web space at <https://infrastructure.nanocommons.eu/>. To facilitate community access to the tools and the support materials accompanying them, work is underway to build a “resources” library linked to services offered through our TA actions, as shown in Figure 2.

TA service category	Nanoinformatics services	Training formats	Level
Experimental workflow	PC characterization protocols	Written tutorial	Basic
	Biomax data templates	Recorded webinar	Basic
Data processing & analysis	NIKC data templates	Written tutorial	Advanced
	Jaqpot platforms 4 & 5	Demo video	Advanced
		Demo video	Expert
	Enalos NanoXtract for TEM image analysis	Demo video	Basic
		Online tutorial	Basic
	Biocorona <i>in silico</i> modelling	Recorded webinar	Advanced
	OpenRiskNet e-infrastructure	Recorded webinars incl. demo videos	Basic
		Recorded webinars incl. videos and documentations	Advanced
Data visualization & toxicity prediction	Enalos cloud for zeta potential	Demo video	Basic
		Recorded webinar	Basic
	Enalos cloud for Safe-by-Design	Demo video	Basic
		Online tutorial	Basic
GUIDEnano	Recorded webinar	Basic	
Data storage	SciNote	Demo video	Basic
		Online tutorial	Basic
	ACE nano Knowledge Infrastructure	Online tutorial	Basic

Figure 2. List of training materials developed to support the NanoCommons services (tools) which are now available online *via* the NanoCommons e-infrastructure website.

Based on this list of online available tools a series of virtual training sessions has been decided upon to be conducted throughout year 3 (2020). The intention here is to make potential users aware of the offers and to promote their application by researchers at the bench. Therefore, an initiative is being targeted particularly towards early-stage researchers, such as, for instance, the *Young NanoScientists* (YNS) group of the European NanoSafety Cluster. The decided action plan includes the following items:

- 27 March 2020: Martin Himly (PLUS) will introduce NanoCommons and its tools and services at the Venice School (see Figure 3 below) and establish a remote interactive training session in collaboration with Antreas Afantitis (NovaMechanics) on the Enalos Cloud platform;
- 2 April 2020: Martin Himly (PLUS), Beatriz Alfaro Serrano (BNN) and Lee Walker (UKCEH) will introduce the NanoCommons e-infrastructure to the YNS in a virtual session and survey their

expressed needs in order to create a “hit list” of the most-wanted services and tools; Anastasios Papadiamantis (UoB) will provide training on the use of ELNs including an introductory virtual hackathon on the SciNote tool;

- 28 April 2020: Egon Willighagen and Laurent Winckers (UM) will conduct an ontology hackathon for the YNS;
- May 2020: Anastasios Papadiamantis (UoB) will conduct a more in depth interactive training session for the YNS, on the use and versatile applicability of ELNs for nanosafety assessment workflows including environmental and human health experimental approaches;
- June 2020: Antreas Afantitis (NovaMechanics) and Dieter Maier (Biomax) will conduct an interactive online training session for the YNS on the NanoCommons Knowledge Base and KNIME workflows;
- After the summer break: Philip Doganis (NTUA) will conduct an online training session for the YNS on the Jaqpot 4 & 5 suite of tools.

All event details and resources from the workshops will be available in the NanoCommons events log at <https://infrastructure.nanocommons.eu/>.

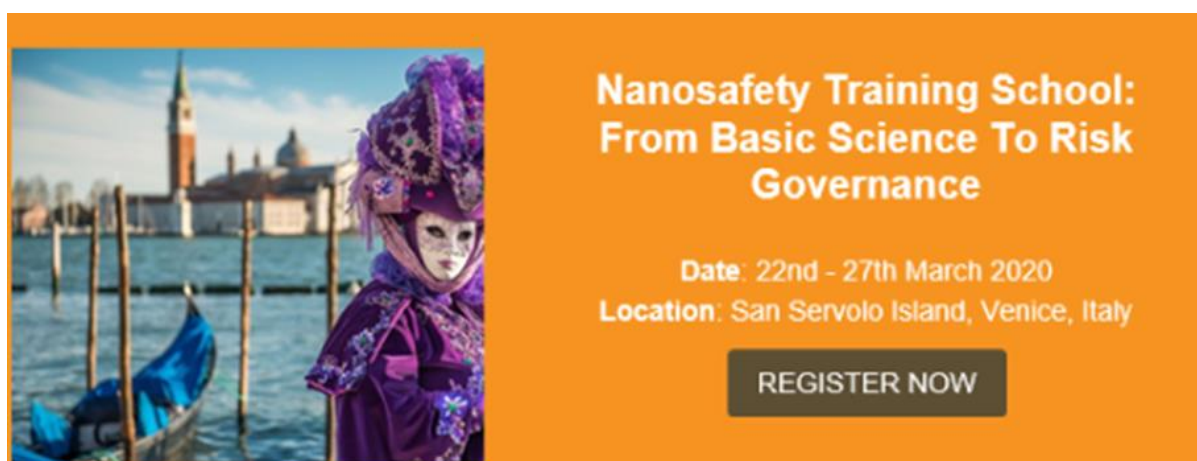


Figure 3: Online notification of the upcoming Nanosafety School in Venice, March 2020.

DEVIATION from action plan: Due to the worldwide CoVid-19 crisis, the *Venice Training School* will be shifted to fall 2020. However, a date has not been announced yet. It might be necessary to conduct the entire event remotely.

Representatives from EUON, namely Abdelqader Sumrein and colleagues, have been invited to join all these interactive online training sessions for the YNS. Additionally, a list of EU projects (NMBP15) and H2020 Marie Curie Innovative Training Networks in the fields of nanosafety and similar topics like advanced/novel/innovative materials and microplastics will be established to develop a wider audience for the webinars and training sessions. Contacts will be made by NanoCommons representatives of WP2 and WP10 with the coordinators initially, and invitations to join these online training sessions will be extended to their experimentalists and modellers. Finally, similar training sessions will be conducted for the US National Cancer Institute working group on nanoinformatics (organised by Mervi Heiskanen, NCI with NanoCommons coordinator Iseult Lynch as the Co-Chair) and for the US National Nanotechnology Coordination Office (organised by Stacey Harper, Oregon Univ.).

As mentioned in the section on ontology developments, the 1st EU Nanosafety Data Curation prize is planned to be awarded to a young researcher who has demonstrated excellence in curation of nanosafety datasets and commitment to making his/her nanosafety research data FAIR. To make a larger audience aware of this new prize and to give an additional incentive to participate in the virtual ontology hackathon in April, this prize will be directly linked to the hackathon and awarded to one of the participants. Every participant will be asked to submit his/her dataset to the NanoCommons data warehouse or any other FAIR database and to inform the NanoCommons team about this. The NanoCommons awards committee (5 members including the NanoCommons coordinator, a FAIR data champion, a database developer, a data curator and our training lead partner) will then award the best of the uploaded nanosafety datasets. Participants will have 4 weeks from the date of the Hackathon to finalize data curation and upload the annotated datasets. Remote support will be provided to participants following the Hackathon to address any additional queries or challenges that may arise.

7. Interactions with the EOSC and the ESFRI roadmap

The European Union expressed its full commitment to open data, open science and open scholarship to make science more efficient, reliable, and responsive to societal challenges (Burgelman, 2019). This has resulted in the Open Research Data Pilot in Horizon 2020 as a first step, which is now continued and intensified with open science becoming the modus operandi of Horizon Europe. Two of the major elements for achieving this goal are:

- 1) the establishment of the European Open Science Cloud (EOSC) bringing together all existing pan-European (e-)infrastructures, harmonizing the procedures for access to the tools provided by those and making them more interoperable; and
- 2) defining the long-term strategy for the development and funding of research infrastructures (including e-infrastructures) via the European Strategy Forum on Research Infrastructures (ESFRI) roadmap, which provides a 10-20 year plan for investment in new and existing infrastructures is updated on an approximately 3-yearly cycle through solicitation of inputs from member states (<https://www.esfri.eu/esfri-roadmap>¹³).

Nanosafety and chemical safety in general are of central importance in order to guarantee the wellbeing of citizens, animals and the environment and, as such, are a major player in the European research arena but also a major beneficiary of cross-disciplinary, interoperable open science approaches. Therefore, the OpenRiskNet e-infrastructure project (coordinated by NanoCommons partner EwC) has started intense interaction with EOSC and the development of a proposal to become included in one of the next ESFRI roadmaps. After the end of OpenRiskNet, NanoCommons has taken over the responsibility for these two activities. In the third year of NanoCommons, further interactions with EOSC and especially the early-adopter programme of the [EOSC-hub](#)¹⁴ and [EOSC-life](#)¹⁵ projects are

¹³ <https://www.esfri.eu/esfri-roadmap>

¹⁴ <https://www.eosc-hub.eu/>

¹⁵ <https://www.eosc-life.eu/>

planned. This will, on the one hand, allow the NanoCommons community to profit from services provided by EOSC, but also enable NanoCommons to establish itself as a service provider on EOSC. On the other hand, it will make the central EOSC projects and other infrastructures providing services on EOSC more aware of NanoCommons as a starting community and, in this way, foster the collaboration and result in better interoperability and harmonization between research from the safety, medical, NMs development, environmental impact assessment and social sciences e-infrastructures, with the latter looking e.g., at life-style influences on exposures to pollutants. Discussions are underway to include the (nano)safety community as consumer, or even as a service provider, in a consortium addressing the challenges of the remaining H2020 infrastructure calls.

[ELIXIR](#)¹⁶, an intergovernmental organisation that brings together European life science resources including databases, software tools, training materials, cloud storage and supercomputers, supports communities of practice in specific research domains. The goal of ELIXIR is to coordinate resources so that they form a single infrastructure. Establishing NanoCommons as a major player in the European infrastructure landscape started with the establishment of a toxicology ELIXIR community and continued with the EOSC activities described above, will further show the need for long-term sustainability of (nano)safety research. Even if NanoCommons was funded to establish and organise a starting community and the infrastructure is not mature enough to be listed in the next ESFRI roadmap (2021 roadmap), years 3 and 4 will be intensively used to finalise the concept development and parts of the design phase, as well as strengthening the support from the member states and becoming integrated into national roadmaps to be prepared for inclusion in the ESFRI roadmap in 2024/2025.

8. Long-term sustainability of NanoCommons outputs

Sustainability of NanoCommons tools and services can only be achieved if the infrastructure is adopted by as many users as possible demonstrating a demand. To promote the tools and services available, EU and nationally funded nanosafety projects, as well as EU and national funding programmes such as Marie Curie Initial Training Networks (ITNs), were identified as potential customers, and evaluated with respect to their relevance for NanoCommons within WP10. Deliverable [D10.2](#)¹⁷ defined the possible ways of interaction with those projects and programmes. As the first action, contacting specific projects and identification of possible collaborations was started. Over the next year, we also plan to start the interactions with the funding programmes as a second step.

Due to the large number of identified projects, a prioritisation process took place based on the partial overlap between the identified project's partners and the NanoCommons consortium, as this increased the likelihood of a meaningful engagement. This resulted in a selection of:

- 29 from a total of 85 identified EU funded projects and
- 16 from a total of 106 identified nationally funded projects.

All NanoCommons partners contributed to the identification of specific topics for the interaction with those projects, which were described in deliverable D10.2. *Annex 1* gives more details on the possible types of interactions with these projects, while *Annex 2* provides an updated list of the projects that

¹⁶ <https://elixir-europe.org/about-us>

¹⁷ <https://zenodo.org/record/3552697#.XhY1CEf7TIW>

have been approached to date and the outcomes of these interactions. The interactions with the following 19 EU-funded projects are prioritised for dedicated support during year 3:

- FP7 project NanoFATE: raw data to be uploaded into NanoCommons Knowledge Warehouse
- FP7 eNanoMapper: external database integration into the NanoCommons Knowledge Base
- FP7 NanoFARM: covered by a TA with NanoCommons involving the University of Aveiro
- FP7 NanoMILE: direct integration of the datasets (data transfer and not just linking)
- FP7 NANoREG and H2020 NanoReg2: linking of the original database without data transfer based on the interfaces developed for eNanoMapper
- H2020 NanoFASE: direct integration of datasets into NanoCommons Knowledge warehouse
- H2020 caLIBRAte: linked via eNanoMapper into the NanoCommons Knowledge Base
- H2020 EC4SafeNano: collection of services/tools and curation & intake of case study data
- H2020 OpenRiskNet: collection of tools (i.e. OpenRiskNet Services Catalog) integrated
- H2020 ACEnano: integration of the ACEnano protocols and data warehouse in the NanoCommons Knowledge Base as well as collection of additional tools
- H2020 GRACIOUS: eNanoMapper based database for integration with NanoCommons
- H2020 PATROLS: eNanoMapper based database for integration with NanoCommons
- H2020 Gov4Nano: discussion on data management and sharing – ELN Case study collaboration
- H2020 NanoInformaTIX: Consensus modelling and integration of tools and models
- H2020 NANORIGO: discussion on data management and best practice in data re-use
- H2020 NanoSolveIT: collection of tools and integration into IATA
- H2020 PANDORA (ITN): datasets to be included in NanoCommons knowledge warehouse
- H2020 RiskGONE: Collaboration on data management and integration of models & tools.

The way a user searches for, and selects, data in the NanoCommons Knowledge Base is independent of where the data is stored as long as the various data warehouses are integrated, which is a key goal of NanoCommons. This can be realised by users (partner projects) uploading their data into the NanoCommons knowledge warehouse (e.g., NanoMILE and NanoFASE) or by linking existing data warehouses (e.g. eNanoMapper, ACEnano) into the Knowledge Base. Once a user makes a search for a specific NM in NanoCommons, they will get a list of all datasets, independently of how the integration of the datasets has been achieved. In this way, NanoCommons can offer all projects optimal support without the need for data duplication. As seen in the list above, many projects use the eNanoMapper database or a variant as their data management solution. At the moment, the original database is being integrated. Once this has happened, the project-specific database extensions could also be integrated with much less effort (as their structure is similar to that of the “mother” database) and can be achieved once the projects release the data under a public license.

The remaining 16 EU-funded projects will be approached in Q2-Q3 2020 via the already identified “NanoCommons Consortium Contacts” (see Annex 2). Once an “ongoing Project” is interested in interacting with NanoCommons (as users of services, data providers, service provider...), a collaboration can be implemented by means of a TA project.

NanoCommons has also started to interact with some of the identified national funded projects. The future interactions with these projects will be reported within the scope of deliverables D2.8 (Action Plan for Year 4) and D7.2 (First TA Resource Access and Use Report).

Conclusions

In the first two years, NanoCommons has established the core infrastructure concentrating on:

- 1) A global solution for data sharing in the nanosafety community by providing a data storage solution, the NanoCommons knowledge warehouse, for research data not yet publicly available and a solution to link this warehouse with other nanosafety data repositories into the NanoCommons Knowledge Base;
- 2) A nanoinformatics toolkit for processing, analysis, modelling and risk assessment;
- 3) Guidelines and standards especially in the area of minimal reporting, scientific metadata and FAIR (Findable, Accessible, Interoperable and Re-usable) data sharing.

These early successes made it obvious that NanoCommons is addressing important needs of the community, especially with respect to metadata standards, semantic annotation, ELNs, data storage solutions and data management and sharing as well as nanoinformatics and risk assessment workflows. These activities will therefore be continued in year 3 and probably even year 4.

Many collaborating projects and TA applications requested support for ontology development in different areas like physicochemical characterisation, modelling and simulation, risk governance and safe-by-design and NanoCommons is reacting to these demands by strengthening the corresponding activities in year 3 and supporting them with the development of new tools.

Sharing and reuse of data is also facilitated by unique and persistent identifiers and clear representations of the NM under investigation. In contrast to small molecules, NMs with all their characteristics to differentiate them from the bulk material but also other nanoforms currently cannot be represented in a simple line notation easy to search and compare. This gap will be filled in year 3 by introducing an InChI for nano.

Additional areas, which were less developed in the first years, are training and community building. These were mainly addressed within the TA projects by customising services for specific projects and training staff of these projects on how to use the NanoCommons tools. The results and learnings from these TAs can now be exploited to generate training offers for larger audiences and to connect and collaborate with a larger number of projects to get more data into the NanoCommons Knowledge warehouse and the federated databases, provide the services as part of additional TAs and also for independent usage, as well as to attract other projects to provide additional services via the NanoCommons infrastructure either from the nanosafety community (nanoinformatics, governance and safe-by-design projects) or from neighbouring communities made available via the EOSC.

References

Burgelman, J.; Pascu, C.; Szkuta, K.; Schomberg, R.; Karalopoulos, A.; Repanas, K.; Schoupe, M.. (2019). Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. *Frontiers in Big Data*. 2. 43. 10.3389/fdata.2019.00043.

Heller, S.R., McNaught, A., Pletnev, I. Stein, S., Tchekhovskoi, D. InChI, the IUPAC International Chemical Identifier. *J Cheminform* **7**, 23 (2015). <https://doi.org/10.1186/s13321-015-0068-4>

Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chemical Information and Computer Sciences*. **28** (1): 31–6 (1988). [doi:10.1021/ci00057a005](https://doi.org/10.1021/ci00057a005).

Annex 1 – Topics for interaction with identified EU/national projects

All NanoCommons partners have contributed to the identification of the best ways of interaction with the identified EU and national projects. As already described in D10.2, possible options are:

- Actions on behalf of NanoCommons:
 - Collection of Needs: NanoCommons will gather information on the projects' needs with specific focus on requirements for development and improvement of nano-enabled products;
 - Collection of Data: NanoCommons will identify the potential for data integration from the specific project. This will include either integrating any project-specific databases or curating and hosting the data within the NanoCommons Knowledge Warehouse;
 - Collection of Services/Tools: NanoCommons will work with projects to identify the available tools/services that fit their specific needs and any gaps that need to be filled. NanoCommons will also consider whether projects have developed tools that would be of interest to the wider community and that should be integrated into the NanoCommons infrastructure;
 - Support for ongoing and future projects: NanoCommons partners are providing their expertise, contacts and networks to assist ongoing projects to achieve their goals. The NanoCommons community will also work as a hub for bringing different parties together to establish new collaborations and networks.
- Similarly, NanoCommons is encouraging equivalent actions from interested projects to facilitate establishment of this two-way relationship:
 - Communicate their Needs: The project has needs they would like to be addressed by the NanoCommons infrastructure and thus will communicate them actively to NanoCommons;
 - Provide Data: The project provides data (re-)sources for incorporation into the NanoCommons Knowledge Base following the FAIR principles and will work actively with NanoCommons data curators to implement the data integration workflow;
 - Provide Services/Tools: The project will be a provider of open access tools/services for the NanoCommons infrastructure and in particular to support TA services;
 - User of Data: The project will use data from the NanoCommons Knowledge Base;
 - User of Services/Tools: The project will be a user of the tools/services offered by the NanoCommons infrastructure eventually within a TA. This includes proposing tools to be integrated in the NanoCommons infrastructure in currently underrepresented areas;
 - Support for future projects: NanoCommons partners provide contacts and networks to bring the right people to the right projects. By providing this community building facility, it is intended to collectively achieve our aim of sustainable use of nanotechnology and NMs.

Annex 2 - List of EU-funded projects interacting with NanoCommons

List of the EU-funded projects relevant for NanoCommons, with which an interaction is planned/has been started (Status April 2020).

If the interaction with a project is already started, it is marked with a 1 in the last column of the table. The corresponding interaction is explained in section “8. Long-term sustainability” of this deliverable.

#	Acronym	Project Coordinator (Company / Organisation)	NanoCommons Consortium Contact (Organization)	Kind of Project	Category (NSC / EPPN / FET / other)	Type of Interaction (NC is active)	Type of Interaction (Project is active)	Interested in which NC Service/Tool	Interaction started ?
1	NanoReTox	University of Birmingham (UoB)	UoB	EU FP7	NSC	Collect Data	Provide Data	Tbd	
2	NanoPoly-Tox	LEITAT	LEITAT	EU FP7	NSC	Collect Data	N/A - project expired	N/A - project expired	
3	NanoFATE	UKCEH	NERC	EU FP7	NSC	Collect Data	Provide Data	Raw data uploaded into NanoCommons version of NIKC	1
4	eNano- Mapper	Edelweiss Connect (EwC)	EwC UM NTUA	EU FP7	NSC	Collect Data	Provide Data	Integration of eNanoMapper database into NanoCommons infrastructure	1
5	NanoMILE	University of Birmingham (UoB)	UoB	EU FP7	NSC	Collect Data	Provide Data	Tbd	1
6	NANoREG	Ministerie van Infrastructuur en Waterstaat	LEITAT BfR	EU FP7	NSC	Collect Data	Provide Data	Integration of NANoREG data (via eNanoMapper database)	1
7	Nano-Solutions	Finnish Institute of Occupational Health (FIOH)	LEITAT UCD	EU FP7	NSC	Collect data	N/A - project expired	N/A - project expired	
8	Nano- GRAVUR	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)	BfR	EU FP7	NSC ??	Collect Data	Provide Data	Tbd	

9	NanoReg2	Institut National de l'Environnement et des Risques (INERIS)	BfR	EU H2020	NSC	Collect Data	Provide Data	Integration of NanoReg2 data (via eNanoMapper database)	1
10	DIMAP	Profactor GmbH	BNN	EU H2020	EPPN	Collect Data	Provide Data	Tbd	
11	Hi-RESPONSE	Precision Varionic International Limited	BNN	EU H2020	EPPN	Collect Data	Provide Data	Tbd	
12	INSPIRED	Joanneum Research Forschungsgesellschaft mbH	BNN	EU H2020	EPPN	Collect Data	Provide Data	Tbd	
13	NanoFARM	Universidade de Aveiro	UKCEH	EU H2020	NSC	Collect Data	Provide Data	TA involving University of Aveiro	
14	HISENTS	University of Leeds	UKCEH	EU H2020	NSC	Collect Data	Provide Data	Tbd	
15	Serenade	CEREGE CNRS-Aix-Marseille Univ	Tbd	EU H2020	NSC	Collect Data	Provide Data	Tbd	
16	R2R Biofluidics - NextGenMicrofluidics	Joanneum Research Forschungsgesellschaft mbH	BNN	EU H2020	EPPN	Collect Data	Provide Data	Possible TA - Tbd	1
17	NanoFASE	UKCEH	UKCEH	EU H2020	NSC	Needs; Data source; Provider; Support of future projects	Needs; Data source; Provider; Support of future projects	Data source and need for Lab notebook; Provider of Data processing and Analysis	1
18	caLIBRAte	National Research Centre for the Working Environment	LEITAT UKCEH Duke	EU H2020	NSC	Collect Data; Collect Services / Tools	Provide Data; Provide Services / Tools	Integration of caLIBRAte data (via eNanoMapper database)	1
19	EC4Safe-Nano	INERIS	PLUS UoB	EU H2020	NSC	Collect Services	Service Provider	Service source: Incorporation of its risk management tools and strategies into NanoCommons Sustainability of its platform. Its network would also be interesting for NC	1

20	OpenRisk-Net	Edelweiss Connect (EwC)	EwC, NTUA, UM, UoB	EU H2020	EINFRA	Collecting Services/Tools	Provide Services/Tools	Sustainability of its e-infrastructure	1
21	Pandora	Consiglio Nazionale delle Ricerche, Institute of Protein Biochemistry (CNR-IBP)	PLUS	EU H2020	NSC	Collect data	N/A - project nearly expired	N/A - practical work has ceased, project only ongoing for amendment and reporting	1
22	POROUS-4APP	Acondicionamiento Terrasense Asociacion (LEITAT)	LEITAT	EU H2020	EPPN	Collect data	Tbd	Tbd	
23	Lorcenis	SINTEF (Stiftelsen for industriell og teknisk forskning))	NTUA	EU H2020	NSC	Tbd	Use of Service/Tool: possible risk assessment	Risk assessment	
24	MIDES	FCC Aqualia S.A.	LEITAT	EU H2020	EPPN	Collect Data	Tbd	Tbd	
25	ModCOMP	NTUA (National Technology University Athens)	NTUA	EU H2020	NSC	Tbd	Use of Service/Tool: possible risk assessment	Risk assessment	
26	ACEnano	University of Birmingham (UoB)	UoB EwC NERC	EU H2020	NSC	Collect Data (including methods / protocols and experimental results)	Provide Data, Services and Tools	Integration of ACEnano knowledge infrastructure (KI) into NanoCommons infrastructure; sustainability of its KI	1
27	GRACIOUS	Heriot-Watt University	LEITAT Duke Uni BfR	EU H2020	NSC	Data Source	Data Source	Integration of GRACIOUS data (via eNanoMapper database)	1
28	PATROLS	Swansea University	UKCEH	EU H2020	NSC	Collect Data	Provide Data	Integration of PATROLS data (via eNanoMapper!)	1

20	Gov4Nano	Netherlands National Institute for Public Health and the Environment (RIVM)	UoB	EU H2020	NSC	All categories	All categories	Tbd	1
30	NanoInformaTIX	CSIC	BfR UCD NTUA	EU H2020	NSC	Collect Data	Provide Data	Integration of NanoInformaTIX data (via eNanoMapper database)	1
31	NANORIGO	ENAS	PLUS	EU H2020	NSC	Integrate NC activity into future NRGF	Ensure coherence with existing nanoinformatics platforms	Tbd	1
32	NanoSolveIT	NovaM	NovaM UoB NTUA UCD UM Biomax Duke	EU H2020	NSC	All categories	All categories	Tbd	1
33	RiskGONE	NILU	UoB NovaM UM	EU H2020	NSC	All categories	All categories	Integration of NanoInformaTIX data (via eNanoMapper database)	1