ProSafe

Grant Agreement Number 646325

Deliverable 3.1

Landscape of databases useful to EHS assessment of nanomaterials – Gaps and overlaps review

Due date of deliverable: **January 31st 2016**

Actual submission date: **Final Draft - Sept 22nd 2016**

Final for upload to EC – Sept 2016

Author(s) and company:	Peter Ritchie (IOM), Ioannis Basinas (IOM), Martie van Tongeren (IOM), Hugues Crutzen (JRC), Sara Totaro (JRC)
Work package/task:	WP3 / T3.1
Document status:	final
Confidentiality:	confidential / restricted / public
Key words:	

DOCUMENT HISTORY

Version	*Date*	*Reason of change*
1	March 31st 2016	Initial v1 Draft
2/3	April - May 2016	Intermediate working drafts in team iterated to v4
4	June 7th 2016	Final Draft for further partner circulation, ProSafe formatted
5	June 21st 2016	JRC comments to IOM
6	June 30th 2016	Final WP3 updates as discussed and agreed
7	July 6th 2016	Final updates implemented for submission to MC
8	Sept 22nd 2016	Final minor updates after feedback from MC / WP4 re D4.4 info
9	Juli 14, 2017	Project Office harmonized lay-out

Funded by The European Union

Lead beneficiary for this deliverable: IOM, 2

Owner(s) of this document	
Owner of the content	IOM (2)
Co-Owner 1	JRC (3)

Table of Contents

1 Description of task 3.1

(Extract from ProSafe DoW, 06-08-2015, after amendment no.1)

Task 3.1 - Mapping databases - Start M1, end M12
Task Leader: IOM Partners: JRC, TEMAS

This task will identify the locations and types of databases related to nanomaterial EHS, at EU level and, as far as possible, at global level. It shall collaborate within the EU Nanosfety Cluster, which has an active dedicated working group on databases and ontology and conducts a biannual survey on databases. Key EU projects will be contacted to understand their strategies in data(sets) management. Where possible, this task shall analyse the way ontologies / naming conventions are used or being developed in connection with those databases.
The COST initiative MODENA-COST, which looks into datasets and databases for toxicological modelling that would serve regulators, is to be contacted, too. The US-EU dialogue bridging nanoEHS research has a dedicated Community of Research on databases and ontology, which this task shall contact.
The end result of this task is an overview of the database landscape at EU level and beyond, related to nanomaterials EHS information. This overview complements works done by the Nanosafety Cluster and the FP7 eNanoMapper. It will add value by identifying gaps and overlaps in databases.

2 Description of work & main achievements

2.1 Executive Summary

This deliverable presents the results of ProSafe Task 3.1, which was been successfully achieved in the wider context of WP3– “Streamlining data acquisition, collection and data management”. The report describes the landscape of useful data resources by: 1) gathering information about and mapping the accessible databases generated by EU projects and other global initiatives of the last few years that are of relevance to the aims of ProSafe, and 2) providing a detailed description and appraisal of current activities, developments and resources in the overall nano-EHS database landscape, with gaps and overlaps.This work is linked to the Safe by Design (SbD) Synergy Scan research in WP2.1 (reported in D2.1), which is a similar mapping exercise.

A detailed database mapping exercise was undertaken through a rigorous information gathering exercise to examine recent (within the previous 3.5 years) and current nano-EHS data management activities and resources from the EU, the US and beyond. Suitable databases have been identified and their significant properties and attributes have been described. Information was recorded systematically in a project database. Data available across the reviewed initiatives or projects was very variable in quality. Not all of the data ideally required could be extracted from the limited information resources available. This yielded an inventory of 57 different projects with potential data resources of interest, covering the data domains of nano physico-chemical information (n=34); toxicology, (n=30); ecotox (n=20); exposure (n=17); other (n=8). A detailed breakdown of the results and their characteristics is provided. The inventory, particularly if supplemented with some currently missing data, may provide a useful resource to others in future. Projects with the potential for data linkage with ProSafe for subsequent tasks in WP3 are noted.

A literature search was carried out to compile an informative bibliography of peer-reviewed publications relevant to the contemporary nano-EHS state of the art. It helped to shape the mapping exercise and was extremely important in forming and providing wider context to the overall landscaping exercise, both in terms of particular developments, such as project databases, and for ongoing interactions and activities in the nano-EHS data management community. The bibliographic database produced, especially if it can be kept up to date, could provide a very useful resource for the community in future.

A comprehensive overview and examination of existing knowledge, contacts and recent and new information arising from the area was undertaken out to provide a detailed description of the nano-EHS database landscape. Besides information from the mapping exercise and bibliography, key sources examined were several recent key developments in the EU and US, especially the eNanoMapper and CEINT NKI initiatives respectively. These relevant areas are cited and elaborated, with key features, gaps and overlaps underlined. The crucial importance of the rapidly developing field of NanoInformatics and its influence on the landscape is highlighted.

Ongoing collaborative efforts to systematically identify and address problems, issues and gaps, are examined and summarised, providing an analysis of needs in the current landscape and, where relevant, suggestions or initiatives underway to address them. Significant interactions and collaborations of the WP3 team with other ongoing initiatives and activities in the field, as proposed by the DoA, have taken place and are noted in the report.

This deliverable and its annexes, with a summary of conclusions from the mapping and landscaping exercise, can: 1) feed into the next steps of WP3 and further areas of the ProSafe project, and; 2) provide substantial bases for useful resources and information (mapping and inventory database, bibliography, landscaping) for wider nano-EHS database developments and the advancing nanoinformatics field more generally.

.

2.2 Introduction

This deliverable presents the results of ProSafe Task 3.1, which has been carried out in the wider context of WP3– “Streamlining data acquisition, collection and data management”. WP3 is composed of the following tasks and their related deliverables:

- T 3.1 Mapping databases –
    - D3.1 "*Landscape of databases useful to EHS assessment of nanomaterials - Gaps and overlaps review*", M12
- T 3.2 NanoEHS community agreed database management system –
    - D3.2 "*ISA-TAB-NANO database system established and adopted within the Nanosafety Cluster*" – Initial report M12 (postponed to M14 with MC agreement), final report M23,
    - D3.3 "*Minimal ontology and naming convention for nanosafety data*", Initial report M12 (postponed to M14 with MC agreement), final report M23
- T 3.3 Linking databases – D3.4 Report on available database linking tools –
    - D3.4 "*Report on available database linking tools*", M23.

The original aims and objectives for T3.1 from the DoA were to:

- Identify the locations and types of databases related to nanomaterial (NM) environment, health and safety (nano-EHS) at EU level and, as far as possible, at global level,
- Collaborate within the EU NanoSafety Cluster (NSC), which has an active dedicated working group on databases and ontology and conduct a biannual survey on databases,
- Contact key EU projects to understand their strategies in data(sets) management,

- Where possible, analyse the way ontologies / naming conventions are used or being developed in connection with those databases,
- Contact the COST initiative MODENA-COST, which looks into datasets and databases for toxicological modelling that would serve regulators,
- Contact the US-EU dialogue bridging nano-EHS research with its dedicated Community of Research (CoR) on Databases & Computational Modeling for NanoEHS.

In other words, the remit of T3.1 was to describe the landscape of useful data resources by gathering information about and mapping the accessible databases generated by EU projects and other global initiatives of relevance to the aims of ProSafe. This in essence would provide an overview of the "state of the art" in this area, with an inventory and review of existing resources, and then feed into the subsequent tasks of WP3. This work is linked to the Safe by Design (SbD) Synergy Scan research in WP2.1 (reported in D2.1), which is also a mapping exercise.

2.2.1 Synergy with T2.1 / D2.1 – Synergy Scan

The work performed in T3.1 (and described in the present deliverable D3.1) is also to help provide information pertinent to the Synergy Scan (D2.1). Indeed, one of the crucial points in performing SbD is the availability of good quality data to allow the prediction of the behaviour of NMs[1], and assisting in the design of intelligent testing strategies. The possibility to have available data, structured in a harmonised way, should in principle also facilitate the ability to derive Structure-Activity relationship for the NMs, contributing to the vision of Safety-by-Design that has been embraced by ProSafe and NANoREG.

During the 3rd ProSafe Consortium Meeting (February 2016), the importance of having good data sources was restated, and it was suggested that the Synergy Scan included also information on "quality of data, harmonisation of data management, lack of continuity of data storage". However, it was also agreed that, since D3.1 was mainly related to the topic of "data", such discussion should be transfer in this latter deliverable. Nonetheless, in this context it is worth mentioning that a clear synergy is established between T2.1 and T3.1 (and their related deliverables).

Several projects that have been highlighted by T3.1 (Table 4 in present deliverable D3.1) are also included in the Synergy Scan (deliverable D2.1). This correspondence is shown in Table 1 below.

[1] ProSafe D4.4 – *Inventory of harmonised national regulation oriented tasks*

Table 1: Correspondence between projects addressed in D3.1 and D2.1

Project	D3.1 Mapping	D2.1 Synergy scan
BUONAPART-E	√	
CaNanoLab - Cancer Nanotechnology Laboratory	√	
CASCATBEL	√	
Dana	√	
DENANA		√
DUSTINANO		√
ECO-TEXNANO (program Life)		√
eNanoMapper (eNM)	√ *	√
ENPRA		√
EURONanoTox (Nano-HEALTH)	√	
FIBRALSPEC	√	√
FutureNanoNeeds	√	√
GLADIATOR	√	
GUIDEnano	√	√
HINAMOX	√	√
Inflammatory, genotoxic and tumorigenic effects of multi-walled carbon nanotubes with emphasis on the interleukin-1 family		√
ITS-Nano (Intelligent Testing Strategy for ENMs)	√	√
LICARA	√	
MARINA	√	√
MembraneNanoPart	√	
MOD-ENP-TOX	√	√
MODERN	√	
ModNanoTox	√	√
NanoCare	√	
nanoCOLT		√
NanoDefine	√	
Nanodetector	√	
NanoDevice	√	
NanoEIS	√	
NanoFASE (H2020)	√	√
NANOFATE	√	√
NANOfutures	√	
NanoGEM	√	
NanoGenotox		√
nanoGRAVUR		√
NANOHETER (funded under SIINN)	√	√
NanoHouse	√	
Nanohub (EU)	√	
Nanohub (US/NSF)	√	
NanoLyse	√	
Nanomaterial Biological Interactions Knowledgebase	√	
NanoMICEX	√	√
NanoMILE	√	√
NanoMiner (FP7 NANOMMUNE project)	√	
NANOMOBIL		√
Nanoparticle Information Library	√	
NanoPUZZLES	√	√
NaNoREG	√	√

Project	D3.1 Mapping	D2.1 Synergy scan
NanoReg2 (H2020)	√	√
NanoReTox		√
NanosafePACK	√	√
NANOSOLUTIONS	√	√
nanoSTAIR	√	
NanoSustain		√
NanoTOES	√	
NanoTransKinetics	√	
NanoUmwelt		√
NanoValid	√	√
nanOxiMet	√	√
NMs Registry	√	
NorNANoREG		√
OECD Database on Research into the Safety of Manufactured NMs	√	
PreNanoTox	√	
ProCycle		√
PROSPECT		√
qNANO		√
QualityNano	√	
SANOWORK	√	√
Scaffold	√	
SERENADE		√
SetNanoMetro	√	
SIINN	√	√
SIRENA (Life Proram)	√	√
SmartNano	√	
SUN	√	√
Turning forestry biomass into sustainable nanocellulose-based materials		√
*** given the particular role of eNanoMapper in the current landscape, it was treated as a special case in the T3.1 mapping exercise, rather than mapped as a potential project data resource, as further described in section 2.4.**		

As mentioned above, for a deeper discussion on the projects that have been highlighted as potentially influential with respect to SbD, we refer the reader to D2.1 – Synergy Scan.

2.2.2 *Synergy with T4.2 / D4.4 – Inventory of the harmonised national regulation oriented tasks and associated ProSafe search call*

Latterly in revising this deliverable our attention was drawn to related information arising in WP4. So in addition to the cross-linkage identified in 2.2.1 above there is also notable synergy between parts of the work of Task 3.1 presented here and aspects of that carried out under ProSafe Task 4.2. In order to harmonize national and other regulation oriented protocols, procedures and data, its ProSafe work has been focused on identifying and linking existing initiatives, promoting the acceptance, and uptake of safe-by-design approaches. Deliverable D4.4 "Inventory of the harmonised national regulation oriented tasks" (draft version 2 at the time of writing), addresses the harmonisation of national regulatory oriented tasks and resources, including a section on data and databases and ongoing requirements for their standardisation and harmonisation, as well as upcoming needs foreseen in relation to text and data mining for the extraction of information (D4.4, 3.1.4.6).

In addition a major output of that WP4 work has been the document "ProSafe search call for tools and Information on the implementation of Safe-by-Design" (version of December 1, 2015). For the preparation of the calls and the harmonisation in ProSafe, information is needed on already existing and available tools for the implementation of the NANoREG Safe-by-Design concept (SbD). This has an enumerated list of requirements as well as an overview of the state of the art for SbD, including databases and libraries, in section 1. Database requirements are summarised in recommendations R1 to R6 (section 1.1.1), and section 1.1.2 provides a summary of the state of the art for several databases that include Nano EHS related information resources of interest which are considered elsewhere in subsequent parts of section 2 of the current report (D3.1), and so they are not further described here.

2.3 Overall WP 3.1 task planning, timetable, and liaison

Given changes in the configuration of ProSafe and the reallocation of some tasks in the early months of the project this task was allocated to IOM and initiated in May 2015. A combined WP2/WP3 meeting was subsequently held in Brussels on 12 June 2015 giving the opportunity for JRC and IOM representatives to meet and discuss WP3, presenting proposals for the practical aims and objectives of T3.1. In addition several key stakeholders in the area of EHS data management were invited to participate (largely by teleconference) to obtain an overview and provide input to the further planning of WP3. As a result, it was agreed (in line with the key DoA objectives stated above) that the database mapping and landscaping exercise would encompass the following key tasks to enable ProSafe to liaise with the appropriate key players and participants in the field, and in particular to gather knowledge, information and results, on existing and ongoing nano-EHS database developments by:

- Investigating key EU projects to catalogue databases and strategies for data management, utilising in the first instance existing resources the NSC Project Compendium website information,
- Scanning recent publications and sites for information on other relevant database developments,
- Collaborating within the EU NSC WG4 on databases and ontology, and with respect to its biannual survey on nano-EHS databases,
- Where possible, analysing how ontologies/naming conventions are used or being developed for nano-EHS databases,
- Contacting and collaborating with the FP7 eNanoMapper (eNM) project to understand and complement work where appropriate,
- Contacting other initiatives, including MODENA-COST and the US-EU CoR, regarding datasets and databases used for toxicology modelling.

These areas are reported in the relevant sections below and followed by an overview analysis and discussion, bringing together the information into a detailed inventory or map of the nano-EHS database landscape, followed by a summary of conclusions and recommendations to feed in to the subsequent tasks of WP3 and, potentially, beyond. A schematic overview of the general workflow and information sources for the tasks in WP3.1 is shown in figure 1.

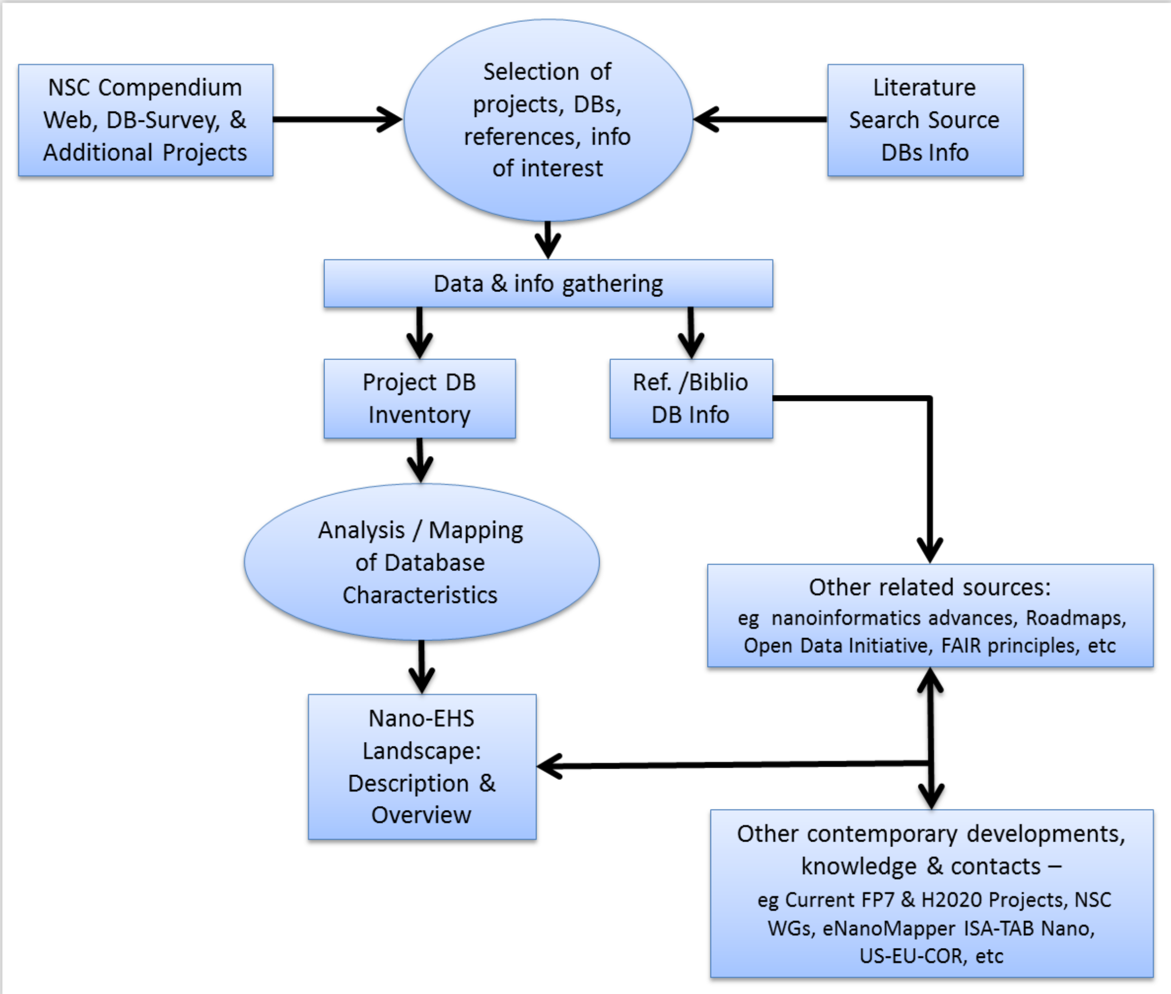


Figure 1: schematic overview of the general workflow and information sources for tasks of WP3.1

2.4 Main database inventory and mapping exercise

This task encapsulates information gathering exercises carried out to: investigate key EU projects (FP 7 and Horizon 2020); gather information from publications (papers and other articles) on other databases and related developments; and accumulate the information in a working catalogue (an inventory database for the project) that could form the basis to "map" properties, characteristics and features of the respective database information.

2.4.1 Methods

2.4.1.1 Data collection planning, information sources and methodology

This task sets about building a basic inventory of FP7 and H2020 projects of interest to the mapping and landscaping exercise and descriptions. It was agreed that we would not execute an extensive survey that would laboriously contact each FP7 or H2020 project in turn for information, (via email, with reminders, follow-up, non-response chasing, telephone contact to seek further details and so on). Given available resources, the existing knowledge of WP3 team members and in the interests of efficiency, we agreed instead to make use as far as possible of existing information about the relevant projects.

It was initially agreed in the WP3 team that in order to focus on current or recent developments in the field of EHS databases (DBs), the time criteria for selecting projects to investigate were that the project was either currently ongoing or had been completed within the last two years. In practice, as work progressed, to allow some flexibility in this regard and since not all projects stated their end dates precisely, the completion period was relaxed to approximately 2 to 2.5 years, i.e. since January 2013. In addition, to keep the remit of the search appropriate to our purposes, it was initially agreed (in the June 2015 meeting) that the main data domains of interest for potential databases of concern for ProSafe were those with Physico-Chemical, Toxicological, and Ecotoxicological data. This remit was later extended to also include exposure data, following an agreement to assist on actions to identify exposure datasets in the NSC, with WG3 Exposure.

Initially, two main NSC sources (A and B below) were used to identify projects for inclusion in the inventory, with a third added later (C below):

A. Firstly there were details of FP7 and H2020 Projects, as published in the June 2015 NSC Project Compendium (http://www.nanosafetycluster.eu/home/european-nanosafety-cluster-compendium.html);

B. A was supplemented some other projects not cited in the Compendium, but described on the NSC website (http://www.nanosafetycluster.eu/eu-nanosafety-cluster-projects/seventh-framework-programme-projects, (as of August 2015);

C. This set of projects (A and B combined) was later supplemented with some other projects of significant interest not already included, arising mostly from the results of the NSC WG 4 Biannual Database Survey of 2014 [1] and from literature, and deemed to be useful non-EU-funded additions to the inventory.

The main data gathering task for A and B above took place over August to September 2015, with some follow up investigations and updates occurring in subsequent months. Selected information from C was added later in December 2015.

An objective and unbiased approach was taken in order to independently abstract and assess pertinent information available about each project. It was quickly established that the Project Overview information available from the Compendium (its Overview Matrix table), whilst a helpful summary of the areas of concern of the project, was not necessarily accurate nor up to date with regard to database aspects. Thus our scan of the detailed compendium contents was extended well beyond that to other related materials that included the project website, other articles, newsletters, deliverable reports and papers about the project. Indeed, to obtain sufficient information, the same methods were applied to the projects from sources B and C.

An experienced IOM researcher screened all of the selected projects to check basic eligibility criteria and then qualitatively assessed the information sources. Based upon the given aims, objectives and a description of each project, expert analysis, knowledge and judgement was applied to establish for the stated data domains of interest in the project: the clear use/presence, or the stated intention of developing/using a database for the project; or to assign the likelihood of presence/use of a database.

It should be noted that, realising that the term database could be perceived and used in different ways within different projects, some flexibility was required in our interpretation. Therefore to all intents and purposes for ProSafe our aim was to identify electronic/computerised databases, representing the deliberate collection and organised storage of a set or sets of data using a database management system (DBMS); or in some cases a spreadsheet or set of related spreadsheets (or possibly similar storage) files, that together could constitute a database for a project. Again, expert experience and judgement was applied to derive a realistic assessment of the actual or potential application of a database to the project's data domains.

Whilst for the longer term aims of ProSafe (regarding data linkage and exchange, etc) we were interested in identifying formalised databases using a DBMS, it was still considered useful to also record less formalised instances. In some cases the information from the Compendium makes no explicit mention of a database, although further examination of associated project materials revealed it as a specific development. A database will often not be acknowledged as a particular task or deliverable as a "product". Also undoubtedly some, potentially still useful, undocumented or "informal" (e.g. relatively simple/single user, single issue, etc) databases have been used as incidental tools for tasks in a project work package. Nonetheless, given the thorough examination and expertise of the Researcher reviewing the project it was in some cases apparent that a database would probably be used in meeting the aims and objectives of the project – for example when the aim is to elaborate models based upon empirical data and a combination of datasets to provide a feed into the model development is an inherent part of the process. Although such databases may not be readily accessible (and indeed, if of further interest, their existence would be need to be established absolutely through direct dialogue) it was considered worthwhile to still record the observation, as well as the likelihood of database use. Thus the existence of a database, rather than a simple binary *Database Present* / *Database Not Present*, was distinguished and recorded as follows:

1. Presence clearly described, database exists (i.e. database referenced and clearly described in the available project materials);
2. Not clearly described, but database likely to exist (i.e. no direct reference in the project materials on the presence of a database is made, but based on objectives and information available, it is deemed that a database is likely to exist – e.g. in order to feed data to a model development process);
3. Database unlikely to exist (i.e. no reference related to the presence of a database is made within the project materials and based on project objectives and information available it is unlikely that one will be required/created/used);
4. Unknown (i.e. available information from project materials not sufficient to evaluate the likelihood of database presence).

Figure 2 shows a schematic of the workflow for the collection of information from data sources A and B described above. In essence, knowledge initially garnered from the Compendium was supplemented or confirmed by further details from the project's website or other project-related articles, deliverables or papers. For source B an extra step to source initial basic information was required, since they did not have Compendium entries. Occasionally any of this information could be augmented with "inside knowledge" on certain projects from other WP3 members.

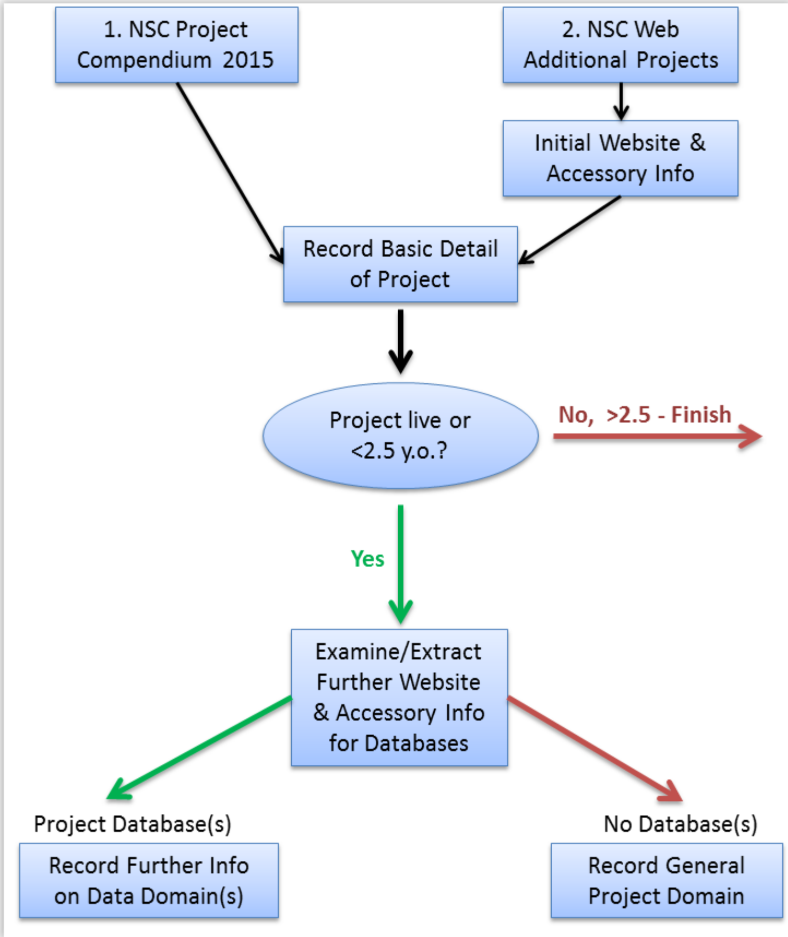


Figure 2: Schematic of data extraction and recording workflow.

A list of data to be recorded was initially developed, and augmented in the early stages, as the first few projects examined suggested additional parameters to record. To allow flexibility, the data was entered by the researcher into an Excel® spreadsheet with appropriate look-up lists for the coding of categorical information for several of the characteristics and properties of a project and its data management features. The schema for the data recording is given in Table 2 below, which shows the variables recorded as well as the labels used in the look-up lists in Table 3.

Table 2: Data collection items

Data Item (Excel® column) name	Item description (See also look-up lists for further detail and examples)
Project name	Name of the project
Time frame	Start and end year (include end month where given)
Project website	URL to Project website
Brief project aim/description	Summary project aim / description
• Main domain	Main data domain (e.g. Toxicology)
• Sub-domain 1	Sub domain of Main, e.g. in-vitro
• Sub-domain 2	E.g. In-vivo

• Sub-domain 3	E.g. -omics
• Assessment of presence of database	Expert assessment on the certain presence or likelihood of database presence / use
• Type of collected database	Qualitative, quantitative, combined
• Form of collected data	Original, third-party, combined
• Source of collected data	Own original experiments, collaborations, literature
• Exposure routes of concern	Inhalation, ingestion, dermal; added for WG3 exposure data investigations
• Data entry methods description	Description of general methods where given
• Entry format	Data collection format / templates
• Purpose of establishment	Summary of reasons / motives for database use
• Database size	Indication of scope / scale
• General database info description	Extracted specific database summary information where available
• Location of Database	E.g. tox lab, or online, etc
• Data curation	Information on specific curation activity, where stated
• Ontologies	Information on specific ontology use or development, where stated
• Main manager	The Main manager
• Contact info	Contact info for this database domain
• Comments	Further comments by the researcher on the search and database findings
Indented Items, beginning with a bullet point may be repeated for each different main data domain pertaining to a project.	

Table 3: Look-up lists for data collection items

Data Item	Look-up list
Main domain	Phys-Chem
	Toxicology
	Ecotox
	Exposure
	Other
Sub-domain 1, 2, 3	In-Vitro
	In-Vivo
	Geno
	Omics
	Human

	Environmental
	Occupational
	Consumer
	Unknown
	LCA
Exposure routes of concern	Inhalation
	Dermal
	Ingestion
	Inhalation / Dermal
	Inhalation/Ingestion
	Dermal / Ingestion
	Inhalation / Dermal / Ingestion
	Not Specified
	Not Relevant
Existence / likelihood of	Presence clearly described, database exists
	Not clearly described, but database likely to exist
	Database unlikely to exist
	Unknown
Type of collected database	Quantitative
	Qualitative – information in DB is mostly textual/descriptive in nature eg contextual info depicting processes involved
	Quantitative / Qualitative
	Unknown
Form of collected data	Original
	Third Party
	Original And Third Party
	Unknown
Source of collected data	Own Experiments
	Collaborations
	Own Experiments / Collaborations
	Literature
	Own Experiments / Collaborations / Literature
	Collaborations/Literature
	Unknown
Entry format	ISA-TAB / ISA-TAB-NANO
	OECD
	Unknown
	Use Own Templates
	Own Templates, possibly also ISA-TAB-NANO
	Other
	NA
	NR

The project information data was essentially recorded at two linked levels: firstly basic information about the project; and secondly for one or more of the four key data domains of physico-chemical, toxicology, ecotoxicology and exposure, (and potentially their more detailed sub-domain information), regarding use of databases. In Table 2 the indented items, (beginning with a bullet point) may be repeated rows, one for each different main data domain pertaining to a project, where required.

The data encountered was highly variable: depending upon the project and the quality and quantity of information available, and of course whether the project used databases or not, the items in the records would be more or less populated. For example a project that was closed and outside of the time limit criteria (prior to January 2013) would only have data for Project name, Time frame, Project Website, Brief project aim/description; even where a database was indicated, not infrequently there may be little more detailed information available. For present purposes the details of those older projects have been screened out of the inventory table (although the original records are retained elsewhere).

For some projects access to the information was efficient and straightforward, for others considerable time and effort were expended to find materials and track down the relevant facts. For example, where web links no longer worked and there was no other fall back, the search had to be curtailed, as it became too time consuming or a dead-end. Prudent expert judgement was applied by the Researcher to assess these factors. In general the circumstances are noted in the inventory data file. Further overall explanatory comments are given in the table in Annex 1.

The Excel® spreadsheet data was later uploaded to an Access® database that could be better used to arrange the two different levels of data into two separate relations and to select, retrieve and report on appropriate sets or sub-sets of data. This has not been worked upon to the extent of being suitable for public consumption at this time. However it is suggested that, given some time for updating, the database could form the basis of a simple on-line resource, with updated inventory contents in future. We have suggested that the data updates might be accomplished in association with the next Biannual WG4 Database survey. The Access® database also houses the relevant bibliographic information as described in section 2.3.1 below. The screen shot in figure 3 below illustrates the fields and layout of information for the toxicology data domain of the MARINA project in the inventory database.

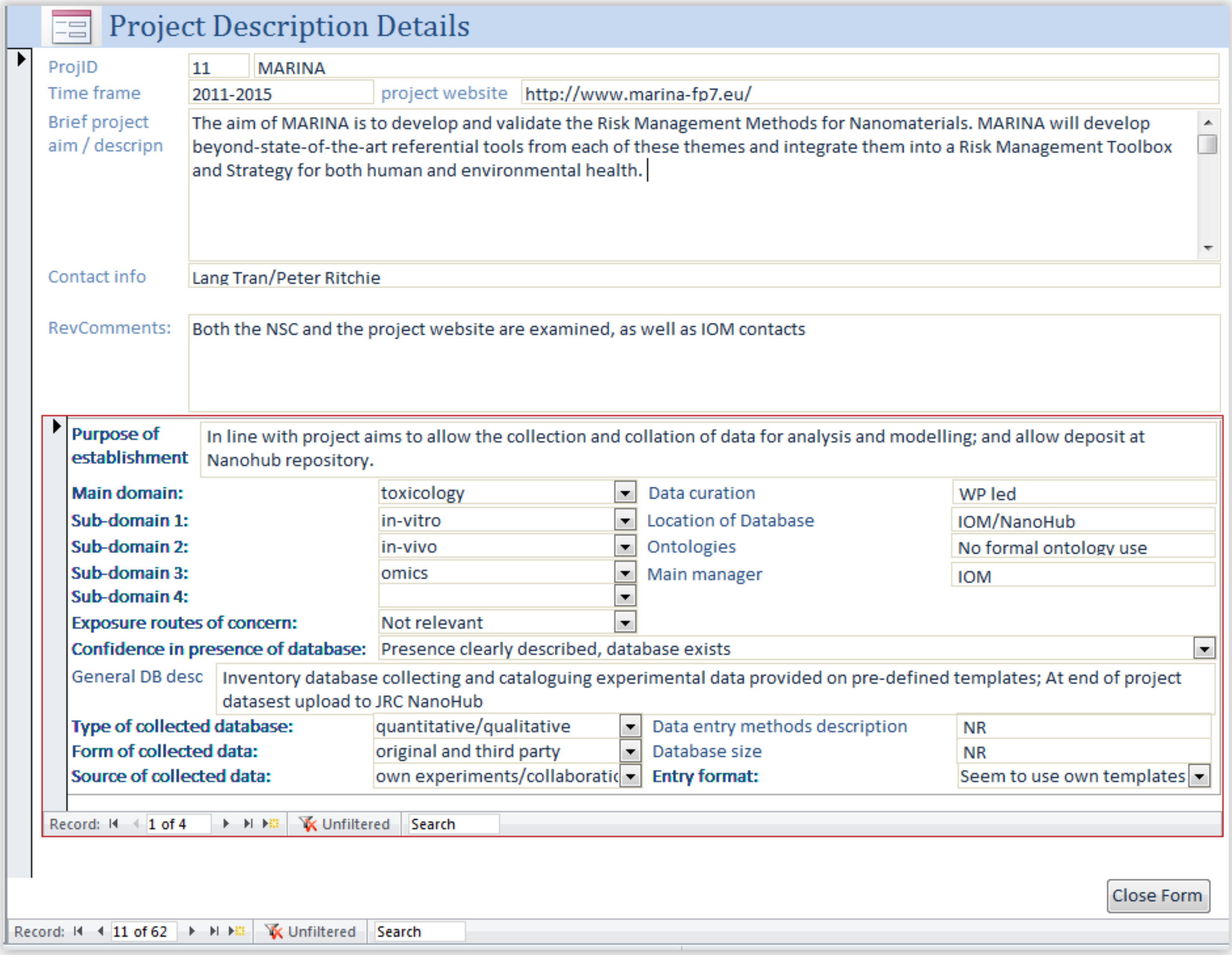


Figure 3: Access® database of inventory information example from MARINA

2.4.1.2 Addition of a sub-set of NSC WG 2014 survey information

At a later stage, after NSC information had been added to the results file and following the collation of information on available literature and papers in the field, it was determined that an appropriate sub-set of the results of the NSC WG4 2014 Database Survey [1] should also usefully be added to the database inventory. This was done as a means of recording, in a manner consistent with the earlier NSC data (A and B above), information on databases that might be of use to ProSafe activities for linking and exchanging data. The same Researcher firstly cross-checked the survey against the data items in the inventory of NSC projects, checking and resolving any anomalies or mismatches wherever possible. Where a new database relevant to our remit was encountered this was added to the inventory. Relevant databases included those with potentially accessible quantitative test results data from one or more of the four key data domains of interests. It did not include several databases concerned with protocols and procedures or reference literature, but without actual results data.

2.4.2 Breakdown of collected data and mapping results

The data was entered into a master spreadsheet as described above. A copy of the complete dataset is available as an Excel® spreadsheet file in Annex 1. After initial scanning to remove some superfluous entries or others that were discovered to be out with the timescale the overall results of this exercise are recorded in the table in Worksheet “Entry Form” in Annex 1.

Projects which referred to innovation, co-ordination and/or network building received a main domain which is “Other”. Whilst they have been retained for completeness, database existence of interest for our purposes for these projects was unlikely (and *currently* they are coloured in red in Annex 1).

For the removal of any doubt the ProSafe and eNanoMapper projects are *not* currently included in the inventory. As ProSafe currently is a co-ordinating and support action, for our purposes it does not (at least to date) have a database output of interest. For eNanoMapper there is not yet available generally accessible data as such, although test data is being loaded. However given its obviously huge strategic and developmental importance in this area, as well as its general alignment and overlap with many of the objectives of WP3 of the ProSafe project, eNanoMapper and its role in relation to the landscape is further discussed in section 2.4.1 below.

Given the scope and scale of the main table, and in particular because of the qualitative nature of much of the information collected, with many gaps in the coverage, our earlier concept of relatively simply being able to map different databases types and characteristics across the different domains and areas of interest via diagrams and infographics has, at least for the time being, been set aside. It is hoped that this may be enabled to a reasonable extent if we can gain significant missing data with information from a proposed NSC database follow-up survey. Instead, a simpler approach that breaks down the information into more easily digestible portions in separate tables has been followed and is summarised below.

The information from this data gathering exercise has derived from 57 projects from the three sources outlined above:

A) 26 from the NSC Compendium;

B) 15 additional NSC website projects;

C) 16 other additional known resources, from the NSC WG4 Database Survey 2014 [1].

Table 4 below is a summary of the complete inventory data in Annex 1, and shows the evaluation of the likelihood of the presence of a database of interest and data domain covered across all mapped projects together with an overview of the source used to identify them, the Project time frame and the relevant information link. Annex 1 contains the full commented entries. The table shows the likelihood of there being a suitable database of interest, and for those with a possible database, each of the key data domains covered, with counts for each as follows: phys-chem,(n=34); toxicology, (n=30); ecotox (n=20); exposure (n=17); Other (n=8). Note that beyond the occurrence of a database in a domain, its volume, availability and means and format of access may be highly variable. These factors have been summarised in the subsequent tables.

Table 4. Presence of a database of interest and main domains covered across all mapped projects together with an overview of the source used to identify them, their time frame and the relevant information link (further full description and overall explanatory comments are given in table 1, Annex 1).

Project name	Time frame	Website	Data domain					Database presence	Source*
			Phys-chem	Toxicology	Ecotox	Exposure	Other		
BUONAPART-E	2012-2016	http://www.buonapart-e.eu/					√	Unlikely	B
CASCATBEL	2013-2017	http://www.cascatbel.eu/					√	Unlikely	B
FIBRALSPEC	2014-2017	www.fibralspec.net/			√			Unlikely	A
FutureNanoNeeds	2013-2016	www.futurenanoneeds.eu/	√					Likely	A
GLADIATOR	2013-2017	http://graphene-gladiator.eu/					√	Unlikely	B
GUIDEnano	2013-2016	www.guidenano.eu	√		√	√		Likely	A
HINAMOX	2009-2013	http://www.hinamox.eu	√	√				Likely	B
LICARA	2012-2014	https://www.tno.nl/en/focus-area/healthy-living/prevention-work-health/innovations-for-workplace-health-and-safety/licara-assesses-risks-and-benefits-of-NMs/			√			Likely	B
MARINA	2011-2015	http://www.marina-fp7.eu/	√	√	√	√		Certain	A
MembraneNanoPart	2013-2015	www.membranenanopart.eu		√				Likely	A
MOD-ENP-TOX	2013-2015	http://fys.kuleuven.be/apps/modenptox	√	√				Certain	A
NanoDefine	2013-2017	www.nanodefine.eu	√					Unknown	A
Nanodetector	2012-2015	www.nanodetector.eu					√	Unlikely	A
NanoEIS	2012-2015	http://www.nanoeis.eu/					√	Unlikely	B
NanoFase	2015-2019	Not available	√		√			Likely	A
NANOFATE	2010-2014	https://wiki.ceh.ac.uk/display/nanofate/Home	√		√	√		Certain (likely for phys-chem)	A
NANOfutures	2010-	http://www.nanofutures.eu/					√	Unknown	B

Project name	Time frame	Website	Data domain					Database presence	Source*
			Phys-chem	Toxicology	Ecotox	Exposure	Other		
	Unknown								
NanoHeter	2013-2016	http://nanoheter.cerege.fr				√		Likely	A
NanoHouse	2010-2013	http://www-nanohouse.cea.fr				√		Unknown	B
NanoLyse	2010-2013	http://www.nanolyse.eu/default.aspx					√	Unlikely	B
NanoMICEX	2012-2015	http://www.nanomicex.eu	√		√	√		Likely	A
NanoMILE	2013-2017	http://www.nanomile.eu	√	√	√			Certain	A
NanoPUZZLES	2013-2015	http://www.nanopuzzles.eu/	√	√				Certain	A
NANoREG	2013-2016	http://www.nanoreg.eu/	√	√	√	√		Certain	A
NanoReg2	2015-2018	http://www.nanoreg2.eu/	√	√	√	√		Certain	A
NanosafePACK	2011-2014	http://www.nanosafepack.eu	√			√		Certain	A
NANOSOLUTIONS	2013-2017	www.nanosolutionsfp7.com	√	√	√			Certain	A
nanoSTAIR	2012- 2014 (potentially)	http://nanostair.eu-vri.eu/						Unlikely	B
NanoTOES	2010-02/2015	http://www.nanotoes.eu/						Unlikely	B
NanoTransKinetics	2011-2014	http://www.nanotranskinetics.eu		√				Certain	B
NanoValid	2011-2015	www.nanovalid.eu	√	√	√			Certain	A
nanOxiMet	2012-2015	http://www.nanoximet.eu	√	√				Certain	A
PreNanoTox	2013-2016	http://prenanotox.tau.ac.il/	√	√				Certain (likely for phys-chem)	A
QualityNano	2011-2015	http://www.qualitynano.eu/		√				Likely	A
SANOWORK	2012-2015	http://www.sanowork.eu	√	√		√		Certain	A
Scaffold	2012-2015	http://scaffold.eu-vri.eu/	√	√	√			Certain	B
SetNanoMetro	2013-2017	www.setnanometro.eu		√				Unlikely	B
SIIN	2013-2016	http://www.siinn.eu					√	Unlikely	A

Project name	Time frame	Website	Data domain					Database presence	Source*
			Phys-chem	Toxicology	Ecotox	Exposure	Other		
SIRENA	2013-2015	www.life-sirena.com	√		√	√		Certain	A
SmartNano	2012-2016	http://www.smartnano.org/						Unlikely	B
SUN	2013-2016	http://www.sun-fp7.eu	√	√	√	√		Certain	A
NanoGEM	Unknown	http://www.nanogem.de/cms/nanogem/front_content.php?idcat=123&lang=11		√		√		Likely	C
Nanomaterial Biological Interactions Knowledgebase	2010-Unknown	http://nbi.oregonstate.edu/	√		√			Certain	C
Dana	Unknown	http://www.nanoobjects.info/en/nanoinfo/knowledge-base	√	√	√	√		Certain	C
EURONanoTox (Nano-HEALTH)	2005-2012 (02)	http://www.nano-health.at/index.php?lang=english		√				Likely	C
CaNanoLab - Cancer Nanotechnology Laboratory	Unknown	https://cananolab.nci.nih.gov/caNanoLab/#/	√					Certain	C
Nanohub (EU)	2010-present	http://www.napira.eu/	√	√	√	√		Certain	C
Nanohub (US/NSF)	Uknown	https://nanohub.org/	√	√				Certain (likely for phys-chem)	C
NanoDevice	2009-2013	http://www.nano-device.eu/index.php?id=123	√	√				Likely	C
MODERN	2013-2016 (probably)	http://modern-fp7.biocenit.cat/index.html	√	√	√			Certain	C
NanoMiner (FP7 NANOMMUNE project)	2008-2011	http://compbio.uta.fi/estools/nanommune/index.php/		√				Certain	C

Project name	Time frame	Website	Data domain					Database presence	Source*
			Phys-chem	Toxicology	Ecotox	Exposure	Other		
NanoCare	2005-2009	http://nanopartikel.info/en/projects/completed-projects/nanocare	√	√		√		Likely	C
Nanoparticle Information Library	Unknown-Present	http://nanoparticlelibrary.net/	√					Certain	C
ITS-Nano (Intelligent Testing Strategy for ENMs)	2012-2014	http://www.safenano.org/research/our-projects/ec-supported-projects/our-core-knowledge-projects/its-nano/	√	√				Absent	C
ModNanoTox	2011-2013 (11)	http://www.birmingham.ac.uk/generic/modnanotox/index.aspx		√				Likely	C
NMs Registry	Unknown-Present	https://www.nanomaterialregistry.org/	√					Certain	C
OECD Database on Research into the Safety of Manufactured NMs	Unknown	http://www.oecd.org/chemicalsafety/nanosafety/testing-programme-manufactured-NMs.htm	√	√	√	√		Likely (Certain for ecotox)	C

Initial source of project information: A=NSC compendium, B=NSC Website list, C= Mustad et al. 2014

Table 5 shows for databases with physico-chemical data (n=34) as a main data domain their characteristics of type of database, form and sources of collected data. One can see here that in many cases there was insufficient information to identify the form of collected data or source of collected data.

Table 5. Basic characteristics of existing databases on physico-chemical attributes across projects.

Project name	Type of database	Form of collected data	Source of collected data
CaNanoLab - Cancer Nanotechnology Laboratory	Quantitative/ qualitative	Original and third party	Unknown
DaNa	Qualitative	Third party	Literature
FutureNanoNeeds	Qualitative	Original	Own experiments
GUIDEnano	Quantitative/ qualitative	Unknown	Own experiments/ collaborations
HINAMOX	Unknown	Unknown	Unknown
MARINA	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
MOD-ENP-TOX	Quantitative	Original	Own experiments
MODERN	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
NanoCare	Quantitative/ qualitative	Unknown	Unknown
NanoDefine	Unknown	Unknown	Unknown
NanoDevice	Unknown	Unknown	Unknown
NanoFase	Unknown	Unknown	Unknown
NANOFATE	Unknown	Unknown	Unknown
Nanohub (EU)	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
Nanohub (US/NSF)	Unknown	Unknown	Unknown
Nanomaterial Biological Interactions Knowledgebase	Quantitative/ qualitative	Original	Own experiments
Nanomaterials Registry	Quantitative/ qualitative	Unknown and third party	Unknown
NanoMICEX	Unknown	Unknown	Unknown
NanoMILE	Unknown	Unknown	Unknown
Nanoparticle Information Library	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
NanoPUZZLES	Quantitative/ qualitative	Third party	Literature
NANoREG	Quantitative	Original and third party	Own experiments/ collaborations/ literature
NanoReg2	Unknown	Original and third party	Own experiments/ collaborations/ literature
NanosafePACK	Quantitative/ qualitative	Third party	Literature
NANOSOLUTIONS	Quantitative/ qualitative	Original	Own experiments/ collaborations

NanoValid	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations
nanOxiMet	Quantitative	Original	Own experiments
OECD Database on Research into the Safety of Manufactured NMs	Quantitative/ qualitative	Unknown	Unknown
PreNanoTox	Quantitative/ qualitative	Third party	Literature
SANOWORK	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
Scaffold	Quantitative/ qualitative	Third party	Literature
SIRENA	Quantitative/ qualitative	Third party	Literature
SUN	Quantitative	Original	Own experiments

Table 6 shows the same characteristics of existing databases for the domains of toxicology (n=30) and ecotoxicology (n=20) across projects.

Table 6. Basic characteristics of existing databases on the domains of toxicology and ecotoxicology across projects.

Project name	Domain: Toxicology				Domain: Ecotoxicology		
	Sub-domain	Type of collected database	Form of collected data	Source of collected data	Type of collected database	Form of collected data	Source of collected data
CaNanoLab - Cancer Nanotechnology Laboratory	In-vitro/ in vivo	Quantitative/ qualitative	Original and third party	Unknown			
DaNa		Qualitative	Third party	Literature	Qualitative	Third party	Literature
EURONanoTox (or Nano-HEALTH as called in website)		Unknown	Original	Own experiments			
GUIDEnano					Quantitative/ qualitative	Original and third party	Own experiments/ collaborations
HINAMOX	In-vitro/ in vivo	Unknown	Unknown	Unknown			
ITS-Nano (Intelligent Testing Strategy for Engineered Nanomaterials)		Qualitative	Third party	Literature			
LICARA					Qualitative	Unknown	Unknown
MARINA	In-vitro/ in vivo	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
MembraneNanoPart	Unknown	Qualitative	Unknown	Collaborations			
MOD-ENP-TOX	In-vitro/ in vivo	Qualitative	Original and third party	Own experiments/ collaborations/ literature			
MODERN	In-vitro/ in vivo	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
ModNanoTox	Unknown	Unknown	Unknown	Unknown			

Project name	Domain: Toxicology				Domain: Ecotoxicology		
	Sub-domain	**Type of collected database**	**Form of collected data**	**Source of collected data**	**Type of collected database**	**Form of collected data**	**Source of collected data**
NanoCare	In-vitro/ in vivo	Quantitative/ qualitative	Unknown	Unknown			
NanoDevice		Qualitative	Unknown	Unknown			
NanoFase					Unknown	Unknown	Unknown
NANOFATE					Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
Nanohub (EU)	In-vitro/ in vivo	Quantitative/ qualitative	Original and third party	Own experiments/collaborations/literature	Quantitative/ qualitative	Original and third party	Own experiments/collaborations/literature
Nanohub (US/NSF)	In-vitro/ in vivo	Quantitative/ qualitative	Third party	Literature			
NanoGEM	Unknown	Unknown	Unknown	Unknown			
Nanomaterial Biological Interactions Knowledgebase					Quantitative/ qualitative	Original	Own experiments
NanoMICEX					Unknown	Unknown	Unknown
NanoMILE	In-vitro/ in vivo	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
NanoMiner (FP7 NANOMMUNE project)		Quantitative/ qualitative	Third party	Collaborations/literature			
NanoPUZZLES	Unknown	Quantitative/ qualitative	Third party	Literature			
NANoREG	In-vitro/ in vivo	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
NanoReg2	In-vitro/ in vivo	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature

Project name	Domain: Toxicology				Domain: Ecotoxicology		
	Sub-domain	**Type of collected database**	**Form of collected data**	**Source of collected data**	**Type of collected database**	**Form of collected data**	**Source of collected data**
				literature			
NANOSOLUTIONS	In-vitro/ in vivo	Quantitative/ qualitative	Original	Own experiments/ collaborations	Quantitative/ qualitative	Original	Own experiments/ collaborations
NanoTransKinetics	In-vitro/ in vivo	Quantitative/ qualitative	Third party	Collaborations			
NanoValid	In-vitro/ in vivo	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations
nanOxiMet	In-vivo	quantitative	Original	Own experiments			
OECD Database on Research into the Safety of Manufactured Nanomaterials	In-vitro/ in vivo	Quantitative/ qualitative	Unknown	In-vitro/ in vivo	Quantitative/ qualitative	Unknown	Unknown
PreNanoTox	In-vitro/ in vivo	Quantitative/ qualitative	third party	Literature			
QualityNano	Unknown	Unknown	Unknown	Unknown			
SANOWORK	In-vitro	Quantitative/ qualitative	Original and third party	Collaborations/ literature			
Scaffold	In-vitro/ in vivo	Quantitative/ qualitative	Third party	Literature	Qualitative	Unknown	Unknown
SIRENA					Quantitative/ qualitative	Third party	Literature
SUN	In-vitro/ in vivo	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/literature	Quantitative/ qualitative	Original	Own experiments

Table 7 summarises the same basic characteristics of databases on exposure (n=17) across projects, also where possible identifying the type of exposure covered, i.e. Occupational, Consumer and Environmental.

Table 7. Basic characteristics of existing databases on exposure across projects.

Project name	Type of exposure covered			Type of collected database	Form of collected data	Source of collected data
	Occupational	Consumer	Environmental			
DaNa				Qualitative	Third party	Literature
GUIDEnano	√		√	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations
MARINA	√	√	√	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
NanoCare	√			Quantitative/ qualitative	Unknown	Unknown
NANOFATE			√	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
Nanohub (EU)	√	√	√	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
NanoGEM	√			Unknown	Unknown	Unknown
NanoHeter			√	Unknown	Unknown	Unknown
NanoHouse			√	Unknown	Unknown	Unknown
NanoMICEX	√			Unknown	Unknown	Unknown
NANoREG	√	√	√	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
NanoReg2	√	√	√	Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
NanosafePACK	√	√		Quantitative/ qualitative	Third party	Literature
OECD Database on Research into the Safety of Manufactured Nanomaterials			√	Quantitative/ qualitative	Unknown	Unknown
SANOWORK	√			Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature
SIRENA	√	√	√	Quantitative/ qualitative	Third party	Literature
SUN	√	√		Quantitative/ qualitative	Original and third party	Own experiments/ collaborations/ literature

Table 8 summarises projects where the use of Ontologies is stated positively (n=3), or is Possible (n=5). Where Possible, this is either as an ambition as stated in materials, or through a potential collaboration and testing of systems, possibly in association with ISA-TAB-Nano pilot tests, e.g. in SUN and NANOSOLUTIONS, most probably with the eNanoMapper project.

Table 8. Projects indicating use of ontologies in their database developments / data management

Project name	Data domain					Presence of ontology	Comment
	Phys-chem	Toxicology	Ecotox	Exposure	Other		
FutureNanoNeeds	√					Yes	Projects aims, among others, to support naming initiatives for next generation nanomaterials
NanoMILE	√	√	√			Possibly	Consortium acts as an NDCI stakeholder liaisons for the data curation, on databases and ontologies; will collaborate with eNanoMapper, at least for NM substance phys-chem data
NanoPUZZLES	√	√				Yes	Use of terminology at ISA-TAB-NANO
NANoREG	√	√	√	√		Yes	Use of terminology in the International Uniform Chemical Information Database (IUCLID). Ontology mapping in collaboration with eNano Mapper.
NanoReg2	√	√	√	√		Possibly	Use of terminology in IUCLID
NanosafePACK	√			√		Yes	CAS numbers, chemical names for ENP are provided.
SUN	√	√	√	√		Possibly	Not employed to date. Exploring possibilities and to test subset with eNanoMapper
Nanomaterial Biological Interactions Knowledgebase	√		√			Possibly	Fields for entry are available but no data are entered in all the records examined
MODERN	√	√	√			Yes	Includes 1903 classes and 81 properties, encompasses knowledge underlying the preparation, chemical composition, physicochemical characterization and in vitro/in vivo characterization of nanomaterials. Intended for top-level modelling of nanomedicine and NM safety (nanosafety) concepts.

Table 9 summarises information extracted on the format of data entry and storage where indicated on identified projects that contained databases. In summary, we can see that some indirect attempt of standardisation is taking place across projects, primarily as result of data exchange and project build up (i.e. when projects are designed in a fashion that allows use of previously collected data). For example, several closely related projects utilise templates based on the MARINA (i.e.

MARINA, GUIDEnano, SANOWORK, SUN) or the ISA-TAB-Nano logic (NANoREG, NanoReg2, MODERN, NanoPUZZLES).

Table 9. Format of data entry and storage on identified projects that contained databases.

Project name	Data domain					Data Format	Comment
	Phys-chem	Toxicology	Ecotox	Exposure	Other		
GUIDEnano	√		√	√		MARINA Excel® Templates	MARINA Project derived Exposure Scenario recording Spreadsheets. NECID may also be used for some exposure collection
MARINA	√	√	√	√		MARINA Excel® Templates	Own adapted templates (Xls test data record and Word Test Method Description Form)
MOD-ENP-TOX	√	√				Own templates	Genedata internal format called GDA/GDC used to store toxicity assay endpoint data; . 3rd party data on in vitro tox of silica NP from 60 peer reviewed papers and the S2Nano
NANOFATE	√		√	√		Own templates	Entry manual on own formats
NanoPUZZLES	√	√				ISA-TAB/ISA-TAB-NANO	Standardised data collection templates, based upon ISA-TAB-Nano. Python code enables conversion to a format required for submission to other online databases e.g. MODERN project
NANoREG	√	√	√	√		Excel templates built following the ISA-TAB-Nano logic	Entry is performed in Excel®-based templates and in a dedicated NANoREG DB.
NanoReg2	√	√	√	√		Potentially, the eNM database model using ISA-TAB-Nano as basis.	Proposed: Entry is performed in Excel®-based templates in liaison with eNM and making use of NANoREG derived ISA-TAB-Nano related templates
NanosafePACK	√			√		Other / Own scenario templates	Review results were entered in a references database using Microsoft Access® 2010. exposure scenarios in Excel®
NANOSOLU–TIONS	√	√	√			MARINA Derived and own project templates; maybe also ISA-TAB-Nano for some tox data	Marina type derived for cell tox; Proposed collaboration with eNanoMapper to also test out ISA-TAB-Nano based approach.
NanoValid	√	√	√			Own templates	
PreNanoTox	√	√				Data extraction system	An automatic text and graph extraction system is used. The system is based on tools developed for a previous project called NHECD that was

Project name	Data domain					Data Format	Comment
	Phys-chem	Toxicology	Ecotox	Exposure	Other		
							part of FP7.
SANOWORK	√	√		√		MARINA Derived and own templates;	For tox some templates were adapted from Marina ones; Some partners own templates/xls spreadsheets
SIRENA	√		√	√		Own templates	A Technological Surveillance System is used. Relevant literature material seems to be captured automatically and then critically reviewed and included on the database
SUN	√	√	√	√		MARINA and own templates; poss test ISA-Tab-Nano with eNanoMapper collaboration	Spreadsheets, Marina exposure scenario templates & NECID data collection as well as own ones depending on the domain
Nanomaterial Biological Interactions Knowledgebase	√		√			Own templates	Data appear to be entered manually in either Excel® or Access® sheets developed internally for this purpose. Downloads are provided in Excel®.
OECD Database on Research into the Safety of Manufactured NMs	√	√	√	√		OECD	Data Providers enter and revise information through a password protected online web portal. The respective representatives to the OECD and the OECD Secretariat will facilitate individual access to the database.
CaNanoLab - Cancer Nanotechnology Laboratory	√					Own templates & ISA-TAB-Nano	To submit data into caNanoLab, a user must belong to data curator group. Then can submit samples and associated characterizations, protocols, and publications.
Nanohub (EU)	√	√	√	√		OECD and IUCLID structure	Need to register to JRC for EU and OECD projects. Projects can upload their own results. Structure based on OECD protocols and IUCLID structure.
Nanohub (US/NSF)	√	√				OECD and IUCLID structure; submitters own format	Only study specific datasets seem to be available
MODERN	√	√	√			ISA-TAB/ISA-TAB-Nano	Data are entered electronically through an online platform developed within the project.
NanoMiner (FP7 NANOMMUNE project)		√				Data extracted from standard GEO Omics formats with	The datasets in NanoMiner were downloaded as raw data files from Gene Expression Omnibus (GEO,

Project name	Data domain					Data Format	Comment
	Phys-chem	Toxicology	Ecotox	Exposure	Other		
						customised preprocessing	http://www.ncbi.nlm.nih.gov/geo/) [7] and ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) [8].
NanoMaterials Registry	√					Seem to use own templates	Manual entry on standard forms (own ones)

As all the relevant information was not always readily available from the NSC Compendium or the project's website and it was not feasible to follow up for all projects, there are gaps in our detailed knowledge for several projects. In a few cases we were able to supplement the compendium information with "local knowledge" as projects were those that WP3 members, or colleagues in their institutes, had direct experience - e.g. the MARINA and SUN at IOM, and the NANoREG at JRC. Besides the NSC Compendium information, the additional provision of an easily accessible inventory or catalogue of nano-EHS data to allow at least the discovery of potentially useful datasets will be very useful, pending an improved future scenario with the adoption of more standardised and harmonised data formats and database systems such as the eNanoMapper model. However, the take-up of this across a range of projects will take several years, so it will be helpful to develop an accessible inventory in the interim.

It has been suggested that information from this mapping exercise may later be supplemented with data from an augmented NSC WG4 Biannual Database Survey. The latter will be taking place in 2016, and one or more ProSafe WP3 members are likely to participate in designing and running the survey. Another suggestion to help improve information gathering regarding databases available from NSC projects is that the requirement for projects to contribute to the NSC Compendium is supplemented by a more specific and rigorous section on data and database outputs for a project (updated annually).

Scanning the results for active projects that could be contenders for interaction with ProSafe in terms of data exchange and linking, as foreseen in T3.3 (concentrating at least in the first instance on those with both physico-chemical and toxicological data) we observe the following list of possibilities in table 10.

Table 10. Active projects with data-linkage potential for ProSafe.

Project Name	Time frame	Comments on potential for interaction with ProSafe
NANOFASE	2015 – 2019	Not predominately Tox but potentially some; will have Ecotox data
NANOMILE	2013 – Mar 2017	Database plans not clear. Likely to collaborate with eNanoMapper (eNM) on Phys-chem and Toxdata loading to eNM database. (Will also have omics)
NANoREG	2013 – 2017	NANoREG database and JRC ISA-TAB-Nano-based templates used. Ongoing collaboration and synergy with eNM. Public release of NANoREG data planned for 2017; transfer into eNM DB instance under evaluation.
NanoReg 2	2015 – 2018	Will use databases from NANoREG, supplemented with additional aspects; will also use the eNanoMapper data model and data management tools to consume/upload collected datasets that are likely based upon the NANoREG adapted ISA-TAB-Nano and possible older data (ENPRA etc) using MARINA type template.
NANOSOLUTIONS	2013 – Mar 1017	Data being assembled. Possible collaboration with eNM on testing the parsing of phys-chem and tox data and ISA-TAB-Nano transfers (also will have omics data that may be tested in collaboration).
SUN	2013 – Mar 2017	Data being assembled. Possible collaboration with eNM on testing the parsing of phys-chem and tox data and ISA-TAB-Nano transfers (also will have omics data that may be tested in collaboration)
DaNa	Ongoing non EU	A German government sponsored database, not FP7 or H2020. Appears potentially to have some datasets of interest for possible exploitation.

In addition, the MARINA and SANOWORK projects, although recently closed, may yet offer some potential for limited data exchange, given the sharing with WP3 of mutual contacts.

Each of these would of course need further investigation to establish the actual possibilities and confirmation of real data availability for linking or sharing that may be carried out via ProSafe. We have noted where collaborations with eNanoMapper are established or suggested, as this could be a significant factor in assessing the likelihood or ease of progressing data linkage and sharing trials. It makes sense to evaluate the alignment of these objectives in continuing dialogue between eNanoMapper and ProSafe in order to enhance efficiency and mutual benefits.

2.4.3 Other Nano-EHS related databases of potential interest

Over and above the information on databases gathered via the three NSC resources we are also aware of other nano-EHS related databases that are of interest to ProSafe. Several other publicly available databases have been noted from our literature search (described in more detail in section 2.3) and more generally through our ongoing experience and liaison with the NSC, contact with FP7 and H2020 Projects and others active in the field. To date those with potentially accessible ENM Physico-Chemical properties and toxicological test data have predominated. More recent project databases have expanded to include more exposure related data. Some are more product-centred, with information on products containing ENM, whilst others are more non-expert or consumer level introductions to nanoscience and technology, covering the use, production and marketing of ENMs. Good general examples of the range of this type are NanoWerk (http://www.nanowerk.com/nanotechnology_databases.php) or SAFENANO (http://www.safenano.org/), which are both very substantial resources, but limited in terms of readily available results data to serve the ProSafe objectives. Rather than repeat such work that overviews resources of this nature the reader is referred to the following publications for further information on potential resources:

- Table A1 in Maojo et al., [2] which is an extensive list of initiatives and databases spanning the collection, curation, and provision of nanotechnology related information,
- Table 1 of Paneerselvan et al. [3] which gives an overview of nanomaterial-related databases.
- The review work undertaken in the eNanoMapper project. Deliverable 3.1, section 2.1, cites some additional commonly used chemical and toxicogenomics databases of relevance[2].

2.5 Literature search and collation of relevant publications for T3.1 bibliography

It was agreed that another important element for informing Task 3.1 and to help in building up knowledge and information on the nano-EHS database landscape as a whole was via the examination of relevant recent publications in this field. Whilst we would not carry out an exhaustive or ”full systematic review”, gathering information pertinent to the current database landscape and “state of the art” from recent literature would help to provide relevant contextual information and a bibliographic resource for the work in similar fashion to the inventory database. As work progressed it was suggested that this material could also usefully be made available to other users, within and beyond the remit of WP3 in future.

2.5.1 Methods

The IOM’s Scientific Information Officer (SIO) was tasked with scanning for relevant publications. It was agreed that only more recent relevant database publications were desirable, so the time period for searches was confined to publications from January 2012.

Searching was carried out in August and September 2015. An initial set of key words was used to select a first list of potentially relevant publications, and this was examined and refined. Thus, after some judicious initial filtering, and as the IOM researchers had agreed to expand our remit to cover also *exposure-related* nano-EHS data (to assist NSC WG3, Exposure - see above), the area was searched using the following keywords of interest, suitably combined where appropriate:

Nanoinformatics

nanotechnology & database*

nano database / nanodatabase*

nanoinformatics & database*

nanotoxicology & database*

nano / nanomaterials & exposure & database*

nanotechnology & exposure & database*

nanomaterials & exposure & database*

nano / nanotechnology & ontolog*

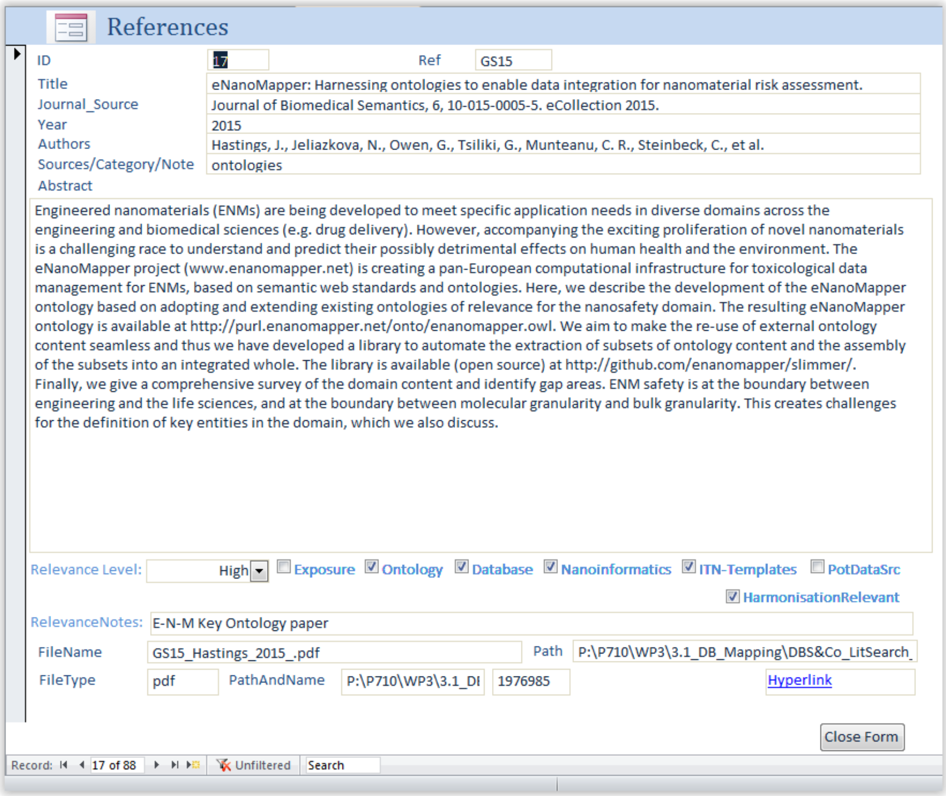
[2] https://www.enanomapper.net/deliverables/d3/150131eNanoMapper-D3.1-IDEA-2015012701.pdf

With regard to “Safe by Design” in relation to the keywords, it was quickly realised that in terms of databases of interest to ProSafe, any such occurrences of interest were encapsulated in the above keyword scheme anyway.

Two main branches of search were done, and then combined into one source: a) using the Proquest Dialog and the Stanford University Highwire database; and b) by using Google Scholar. The details of the publication and its abstract were saved, for further examination by the IOM T3.1 researcher who scanned the materials, selecting the results deemed appropriate to this task, and saved into a temporary bibliography file for processing. All titles and abstracts were read, and the majority of the full publications were downloaded where available, either free or through existing IOM subscriptions, although some, deemed less immediately relevant or important and that were behind pay walls to which we did not immediately have rights, were not accessed or read in full. A few additional more recent references (into 2016) were subsequently obtained and added to the bibliography later in the task. These were items that came to the attention of the writers whilst the mapping and landscaping descriptions were being composed for this report. This was done where it was judged salient to add freshly available information of relevance to the results of T3.1.

2.5.2 Literature search outcomes

In total, 73 relevant references were recorded with the full text of 60 assessed as meriting download to date. The downloaded bibliography has been managed using the Mendeley system, but in addition the abstracts and details of the publications were stored in a small Access® database, alongside the existing project inventory results, to allow local searching and reporting. In this regard, as there are not copyright issues around abstracts this information could potentially (again, suitably updated before release) be made available as an on line resource to the nano-EHS database community in future, especially if there were some resources available to allow it to be kept up to date periodically. An example entry in the database is shown in Figure 4.



References

ID 17 Ref GS15

Title: eNanoMapper: Harnessing ontologies to enable data integration for nanomaterial risk assessment.

Journal_Source: Journal of Biomedical Semantics, 6, 10-015-0005-5. eCollection 2015.

Year: 2015

Authors: Hastings, J., Jeliazkova, N., Owen, G., Tsiliki, G., Munteanu, C. R., Steinbeck, C., et al.

Sources/Category/Note: ontologies

Abstract

Engineered nanomaterials (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 exciting proliferation of novel 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. Here, we describe the development of the eNanoMapper ontology based on adopting and extending existing ontologies of relevance for the nanosafety domain. The resulting eNanoMapper ontology is available at http://purl.enanomapper.net/onto/enanomapper.owl. We aim to make the re-use of external ontology content seamless and thus we have developed a library to automate the extraction of subsets of ontology content and the assembly of the subsets into an integrated whole. The library is available (open source) at http://github.com/enanomapper/slimmer/. Finally, we give a comprehensive survey of the domain content and identify gap areas. ENM safety is at the boundary between engineering and the life sciences, and at the boundary between molecular granularity and bulk granularity. This creates challenges for the definition of key entities in the domain, which we also discuss.

Relevance Level: High Exposure Ontology Database Nanoinformatics ITN-Templates PotDataSrc HarmonisationRelevant

RelevanceNotes: E-N-M Key Ontology paper

FileName: GS15_Hastings_2015_.pdf Path: P:\P710\WP3\3.1_DB_Mapping\DBS&Co_LitSearch_

FileType: pdf PathAndName: P:\P710\WP3\3.1_DI 1976985 Hyperlink

Close Form

Record: 17 of 88 Unfiltered Search

Figure 4: Bibliography in Access® database

The full bibliography that has been built to date is listed in Annex 2.

Publications were assessed and marked up for general relevance in relation to the aims of WP3, as High, Medium, or Low. They were also marked in relation to several key sections of interest to the ProSafe project (besides the Authors own keywording) these being:

i) Field of investigation, namely Exposure; Ontology; Database; Nanoinformatics; ISA-TAB(-Nano) or other templates , (again regarding database interests, encompassing the Safe by Design theme of ProSafe);

ii) The reference/source might be a potential a data source (in terms for example of being able to link to or make use of their existing data);

iii) Coverage of general data harmonisation.

Marking up such areas was not an absolute measure and was done through the judgement of the Researcher, indicating whether this area was addressed in the paper, but it does not include any weighting as to its emphasis within the reference as a whole, nor to its relative importance across the different area. For example, whilst "database" would be expected for all instances given our interests, there were some records where database in itself was of far less concern than was the harmonisation of exposure measurement and collection techniques. Irrespective of particular database technology aspects "Potential data sources of interest" covered three "general purpose" public nano-data repositories that may provide phys-chem data on application, and four exposure related papers whose sources could possibly be approached for data.

Table 11. Data areas of interest included in references.

Data area	Number of references including area
Exposure	20
Ontology	23
Database	45
Nanoinformatics	34
ISA-TAB-(nano) or other templates	12
Potential data source of interest	7
General data harmonisation	20

In general the references in the bibliography have been very useful in informing and giving context to the database inventory mapping. They have also been very helpful to the general landscaping exercise regarding the "state of the art" for nano-EHS databases, and our assessment of gaps and overlaps, which, whilst quite generally known of before T3.1, can now be far more readily highlighted. It is worth noting that several papers in the field of NanoInformatics as well as the output from the eNanoMapper project (with two key papers and two project deliverables to date) were particularly significant for this. This is described in further detail in the Landscape background in section 5 below.

2.6 Other nano-EHS database related activities and developments relevant to the nano-EHS database landscape and ProSafe actions

This section briefly describes several other significant recent or ongoing nano-EHS database related developments of relevance to T3.1 and the database landscape. These may be specific projects, related working groups in the field, other activities of which WP3 members are aware, have been Involved in, or have been in contact with. Several of these were specified in the DoA as important contacts and interactions to have: they are outlined below with appropriate commentary. In fairly general terms for these areas significant factors and issues arising, and useful information and knowledge gained from them are taken into account in the database mapping and landscaping tasks.

2.6.1 The eNanoMapper project and its particular importance for ProSafe WP3

Currently the eNanoMapper project (https://www.enanomapper.net/) is of tremendous importance to the whole field of nano-EHS database development at European level, and beyond. It has been funded by FP7 as the foremost project specifically undertaking fundamental developments in the field. It also has clear alignments and overlaps with the aims and objectives of ProSafe WP3, and other data-centric projects. There has for many years been a growing awareness of the lack of standardisation in data management and shortcomings in data harmonisation and exchange abilities in many FP6, FP7 and now H2020 nano-EHS projects. These have seriously hampered the ability to share and better exploit large quantities of research data funded by the EU. Individual projects alone have been unable to develop solutions to this problem.

The eNanoMapper project started on 1 February 2014 to thoroughly analyse and address the problems in the area of nano-EHS data management and modelling by developing new computational and technical facilities. This includes formats for data collection; ontologies; database systems, architecture and processes; data exchange and linking; and tools for analysis and modelling. Its work packages encompass Community Outreach, Ontology Development, Database Development, Analysis & Modelling, Application Development, Dissemination and Training. It has close interaction with the NSC as it connects with its activities on several fronts. The key aims and objectives for eNanoMapper are given here in full: https://www.enanomapper.net/enm.

There is considerable overlap and synergy between several of the aims and objectives of the ProSafe and eNanoMapper projects. Members of ProSafe WP3, (initially under the auspices of some other FP7 projects) have been liaising in general terms with the eNanoMapper project over the last 12 to 24 months as part of eNanoMapper's Community Outreach WP (participating in user requirements gathering and analysis), through the NSC generally, and, in some cases, more specifically through collaboration and testing of eNanoMapper developments and functionalities. To date this has included working on the exchange and further development of phys-chem and toxicological data collection templates and datasets in relation to the NANoREG and MARINA projects (potentially also in NANOSOLUTIONS and SUN) as well as aiming to test the developing ontology in relation to data definition and collection. In particular, in liaison with eNanoMapper the JRC have undertaken considerable development of phys-chem and in-vitro/in-vivo cell tox data collection templates based upon the ISA-TAB-Nano logic within the NANoREG project.

More recently in the ProSafe project itself, we (JRC) have been discussing eNanoMapper's current developments (i.e. ISA-TAB-Nano templates, data formatting and database linkage) and eNanoMapper personnel participated in the most recent ProSafe consortium meeting (February 2016) contributing to the WP3 database strategizing and brainstorming session and plans for continuing interaction and collaboration via practical data exchange and linkage aspects that are being addressed over the remaining months of the ProSafe (Tasks 3.2 and 3.3) and eNanoMapper projects.

With regard to database landscaping there are overlaps in our exercises to identify existing nano-EHS databases and their features. Indeed, the Spring 2014 Database Survey, which we have incorporated into our Database mapping inventory, was undertaken largely by eNanoMapper staff. We have also benefitted more generally from some of the outputs of eNanoMapper, D3.1 (Technical specification and initial implementation of the protocol and data management web service) as well as information from their published papers [4,5]. More generally, as discussed at the February 2016 ProSafe Consortium meeting, given our close relationship with eNanoMapper we acknowledge any overlaps and make reference to eNanoMapper specific outputs where appropriate in this report. Furthermore, it is recognised that the ongoing liaison between the two projects will aim, where appropriate, to align and mutually benefit from overlapping developments and outputs.

2.6.2 EU NanoSafety Cluster Database Working Group (WG4)

Prior to and during the ProSafe project, both IOM and JRC WP3 staff have actively and positively participated in the EU NanoSafety Cluster Database Working Group (WG4)[3]. This group communicates via a monthly teleconference, e-mails, and periodic (6 monthly or so) face to face Cluster meetings of all the WGs. Whilst previously attending as database practitioners on some other FP7 projects, we have also over the last year consciously represented ProSafe as a Project, updating the group on its progress and contribution to the field, as well as aiming to help progress advances in the nano-EHS field via cooperation and liaison with others active in the field. This has been beneficial to ProSafe T3.1 with connections here facilitating better liaison on information exchange, the sharing of knowledge and data management resources and materials.

2.6.3 Other related NanoSafety Cluster developments

Highly relevant to the database landscape, we have also been active in the wider consideration of issues in relation to data/database standardisation and harmonisation and in particular the longer-term “sustainability” of accessible data related resources, so as to enable the ongoing provision of a nano-EHS Knowledge Infrastructure and Framework. The strategy, planning and particularly resourcing required for real sustainability is currently a much debated matter. WP3 members have participated in the discussions, in particular attending a meeting addressing this area in Brussels in January 2016[4]. They are now involved in follow up, discussing the possible development of a separate Data Sustainability Working Group for the NSC.

2.6.4 US-EU Community of Research – Databases & Computational Modelling for NanoEHS

In a similar vein to 2.6.3, in the interests of ongoing liaison and knowledge exchange, one or more representatives of T3.1 attended several of a series of fortnightly teleconferences, held over the second half of 2015 between nano-EHS domain experts and in the context of the US-EU Community of Research (CoR) '*Databases & Computational Modelling for NanoEHS*'[5]. Including several members of the NSC WGs, and their US counterparts, this series was developing case studies around the sharing of data, project information, knowledge and other data, analysis and modelling resources. Liaison in this area is still active, and indeed has helped to build links with US developments of great interest to ProSafe (see for instance 2.6.5 below). The programme will continue with a further workshop by the group in Arlington in June 2016 (http://us-eu.org/2016-us-eu-nanoehs-workshop/). There were no databased resources directly arising from this liaison currently available for use by WP3.

2.6.5 US CEINT - Center for the Environmental Implications of Nanotechnology and the NDCI

Connected to the CoR, but as a separate initiative, a potentially very significant liaison has been established with CEINT (Centre for Environmental Implication of NanoTechnology, headquartered at Duke University) by the JRC WP3 representatives. In ProSafe links to Duke have also been established in WP1, which has assisted in liaison on database issues. This is in an area that is

---

3 http://www.nanosafetycluster.eu/working-groups/4-database-wg.html
4 http://www.enanomapper.net/data-ontology-and-harmonisation
5 http://us-eu.org/communities-of-research/search-communities-of-research/databases-ontologies/

closely connected to T3.3 (database linking tools), but also to WP3 more generally. CIENT also aligns very closely with parallel eNanoMapper developments, as well as sharing similar database designs. There is a clear synergy that may be built on with significant mutual benefits. Christine Hendren (from Duke University, as representative of CEINT) attended the 3rd ProSafe Consortium meeting to present the CEINT approach to nanoinformatics and data management, and participated in WP3 discussions and brainstorming. In addition, she leads the initiative on the Nanomaterial Data Curation Initiative (NDCI), a collaborative approach to assessing, evaluating, and advancing the state of the field, and promotes further developments of the ISA-TAB-Nano, format and the NanoParticle Ontology (NPO)[6]. Recent developments have also seen CEINT investigating the further adaptation and use of ISA-TAB-Nano so that it can also be used to capture and exchange exposure data. In particular, CEINT and other US counterparts are actively looking into the further exploitation of the open domain 'NANoREG templates' recently published by JRC (see 2.6.10) developed to facilitate the data recording in a well-structured manner during lab assays (as opposed to *a posteriori* from published literature).

2.6.6 Exposure data and the NECID exposure database system

In the FP7 NANEX project (http://www.nanex-project.eu/index.html) Excel® templates were developed for the recording of Exposure Scenario information with much contextual information, with appreciable amounts of qualitative descriptive information on the conditions prevailing during the potential or actual exposure event. Collectively these were compiled into a Nano Exposure Scenario Library. The concept and the associated materials were further developed and used also in the MARINA and SUN projects. However, these templates did not include a standardised definition or specification of the appropriate quantitative data. More recently (in the last 12 months) within the NSC a greater impetus has been placed on enabling the harmonisation of Exposure data, principally under the auspices of NSC Exposure Working Group (WG3)[6]. Whilst separate from ProSafe itself, several of the ProSafe WP3 personnel have been involved in the discussions, which are relevant to ProSafe's interests and the database landscape. This exchange has been stimulated partly through consideration of the pros and cons of use of the Nano Exposure & Contextual Information Database (NECID) [7] in recent FP7 projects, and in part by the eNanoMapper project. This has also helped to further highlight to exposure researchers some tricky generic issues regarding data incompatibility and barriers to data sharing, initially seen in the Phys-Chem and Toxicology fields. T3.1 members have participated in NSC WG3 discussions to help identify existing exposure datasets in the NSC, and this data domain was added to broaden our search remit in T3.1.

NECID has been developed by the PEROSH group in order to harmonise and standardise the collection and recording of occupational exposure data[7]. Whilst it is still being trialled by some researchers and undergoing further development, some recent projects have been using NECID for exposure data capture, for example SUN, GuideNano and NANoREG. The use of NECID is not completely open and currently needs to be approved and licenced (cost free) by the PEROSH group. Whilst the technical development (i.e. the code base of the system) is currently restricted to its development team, significantly eNanoMapper is now working with the NECID team on the development of a unified ontology for exposure data that can be shared between databases and users. This promises to be a major advance. IOM have tested and explored the use of the NECID database for data collection and attended a workshop on its use. IOM will continue to assist in its evaluation. Currently, there are no actual data resources arising or available that could be directly exploited by WP3.

[6] http://www.nanosafetycluster.eu/working-groups/3-exposure-wg.html
[7] https://www.perosh.eu/research-projects/perosh-projects/necid/

2.6.7 MODENA Cost

As project leader of MODENA Cost, Professor Lang Tran is based at IOM. ProSafe T3.1 has thus been able to have direct contact with that project. Mostly concerned with modelling and the development of a 'modelling teaching dataset' for younger scientists, minor updates were made to in-vitro data collection templates. These are closely based upon those used in the MARINA project, modified to allow the addition of in-situ characterisation results. Based upon a subset of the (suitably anonymised) data from the NANOMMUNE and ENPRA projects, and some other anonymised data from several MODENA associates, an Excel®-based teaching dataset has been developed for training toxicology modelling. Whilst in essence currently a “stand-alone” dataset, rather than a formal database, it is unlikely to be immediately exploited for ProSafe WP3 linkage. However in collaboration with eNanoMapper, it is possible that a subset of this dataset could be utilised, depending on permission from the various data “donors”. In addition, it is hoped that further resources may be raised in future and used to develop a larger combined database with in-vitro data from NanoTEST, NANOMMUNE, and ENPRA, (and perhaps others), potentially under the auspices of NanoReg2.

2.6.8 Ontology developments

It is undoubtedly generally accepted in the community that there is an urgent requirement for the use of a standardised ontology in the management of data and information in the nano-EHS data domain. Benefitting from earlier work on the NPO, recent major developments of relevance in this area have been undertaken within the eNanoMapper project, to which the reader is referred [5]. These developments continue to evolve and, as mentioned in relation to the NECID-eNanoMapper liaison, have now grown to encompass nano-EHS exposure data. eNanoMapper has organised webinars and a training session to introduce the community and potential users to the theory and use of the ontology, and its intimate relationship to the eNanoMapper database. These ontology developments and WP3’s ongoing collaboration with eNanoMapper will clearly be of major importance in T3.2.2 “Minimal ontology and naming conventions”. It is also important to note that ontology development also intimately relates to ISA-TAB-Nano developments, as the use of this data storage scheme will crucially depend upon the ontology for its standardised terms and definitions.

2.6.9 Data capture and exchange templates, further development of the ISA-TAB-Nano logic for literature data capturing – The NanoPUZZLES and MODERN projects

The use of many different data capture facilities and formats has been an ongoing problem, creating a barrier to data exchange. Many projects have developed their own data capture solutions of variable quality and, whilst some projects have shared data collection templates with others, these have been patchy at best, lacking standards, common vocabulary or ontology, and have often diverged over time. Hence good data compatibility has been rare. Again, devising solutions to these problems is a component of the work of eNanoMapper. Pre-existing data templates, e.g. from MARINA and NANoREG, made available in collaboration with ProSafe IOM and JRC personnel respectively, have been shared with eNanoMapper to assist in the development and testing of new data capture solutions. To help accommodate a variety of data providers eNanoMapper’s approach has been to not restrict data capture at the front end to one particular format, but instead to allow flexibility, providing processes to parse and extract data from a range of established collection templates. At the same time it is supporting the ISA-TAB-Nano default data capture and exchange logic. ISA-TAB-Nano will also be the standard container into which data from its relational database will be exported when required.

Further ISA-TAB-Nano development and use is intimately connected to the nano-EHS ontology development and use for data definition and collection. Whilst ISA-TAB-Nano use for ProSafe is more a concern of T3.2.1, and is highly significant in the context of the eNanoMapper project, a summary is itemised here in relation to the database landscape.

For several years the ISA-TAB and ISA-TAB-Nano format [8–10] has been advocated as a solution to many of the problems regarding the harmonisation of data collection and exchange using standardised common spreadsheet files as containers for the format, aiming to provide a general and flexible framework to record and integrate nanomaterial study data and other necessary metadata and protocol information. It supports the use of ontology terms to promote standardised vocabulary and descriptions and to facilitate search and integration of data. Whilst conceptually very well defined and robust, it is reasonable to say that overall the take up of this for general experimental data capture has been very slow. Whilst it is mediated through the use of Excel® spreadsheet files the format and the underlying rules for its use are complex and not easy to use for the average end user unfamiliar with the concepts of ontologies and the many other vagaries of arranging data (let alone meta-data) in its idealised, disciplined and uniform manner. It is thus presently considered neither sufficiently intuitive, or user-friendly for the general user.

In a further proof of concept ISA-TAB-Nano has been used to capture data *from the literature* and submit it to a suitably designed database within the NanoPUZZLES (http://www.nanopuzzles.eu/) and MODERN (http://modern-fp7.biocenit.cat/) projects [10]. It was used to successfully record extracted phys-chem and cytotoxicity data in what appears to have been a very complex and labour intensive process. The resulting ISA-TAB-Nano files from this process are available for download and testing [11]. Whilst the resulting files are of much interest to those further developing this format and nano-informaticians, in more general terms understanding their arrangement and use is still quite arcane and challenging. Issues highlighted by this work have prompted several of the updates to the new version (1.3) of ISA-TAB-Nano which is now under development. The generated datasets were submitted to the nanoDMS (Nano Data Management System, http://nanodms.biocenit.cat/), a product of the MODERN project implemented to be compliant with ISA-TAB-Nano. From the above it is evident that the findings of the NanoPUZZLES research and the datasets themselves will be useful data resources for the community. However, to enable access to its structures and data, further useful dialogue between the MODERN database, ProSafe T3.2, and eNanoMapper should take place.

The NanoPUZZLES investigations also emphasise the need for more intuitive and easier to use front ends to the ISA-TAB-Nano format for *literature data* capture. Whilst new version developments propose to change the underlying data file format to the more lightweight, portable and efficient JSON format, as this an underlying file change rather than end-user interface enhancement it is likely to provide only limited advantage, if any, to the (potential) average end user.

2.6.10 Data capture and exchange templates, further development of the ISA-TAB-Nano logic for experimental data capturing – The NANoREG and eNanoMapper projects collaboration

As highlighted in 2.6.9 above, reworking of the “front end” of an ISA-TAB-Nano data exchange system into a more usable tool is still an outstanding requirement. It is certainly so for recording directly data generated during experiments. Experimentalists do need user-friendly systems to *capture experimental data* (raw or processed) and have, by far, almost always resorted to their own-purpose built Excel® spreadsheets of very limited compatibility across laboratories of different organisations, or even within their own. Towards this aim, ProSafe WP3 personnel (JRC), working from within the FP7 NANoREG project, have developed simplified versions of Excel® templates based on the ISA-TAB-Nano logic, constantly following the 'Keep It Simple' leitmotiv[8].

The data management strategy followed in NANoREG aimed at finding a common ground for those different data uses and recording needs: on the one side enable a user-friendly way of logging data in a harmonised and structured manner that, on the other side, can generate datasets sufficiently robust to support, for instance, causal correlation analyses, QSAR modelling or Safe-by-Design (SbD) applications. Importantly, NANoREG made the choice not to record the 'raw data'

[8] http://www.nanoreg.eu/media-and-downloads/templates/269-templates-for-experimental-data-logging with more on the NSC newsletter no.7 Spring 2016, p.14: http://www.nanosafetycluster.eu/newsletter.html .

from the experimental devices into the templates. Only the processed data (sometimes called 'curated data'), resulting from the application of the SOP (standard operating procedure) directions to process the raw data, is logged into a template. The project entrusts its partner organisations with the accurate preservation of the raw datasets in accordance with their own best practices.

NANoREG knows that this translation is an approximation of the actual complete ISA-structure. Nevertheless, this can be the common basis for different communities (e.g. experimental scientists, modellers and even regulators) to speak the same language for several aspects of nano-EHS, when it comes to *experimental data*.

NANoREG has been collaborating with eNanoMapper on translating the files further into the actual ISA-structure. With the support of that project, the NANoREG and future dataset(s) contained in the templates may be easily transferred (using parsers that are being developed by eNanoMapper) into an eNM-database instance (e.g https://data.enanomapper.net/) and also translated into the proper ISA-structure. This also overlaps with ongoing T3.2 work.

2.6.11 EU-NanoHub

Another significant platform, the NANOhub database (http://www.nanohub.eu/) (also known as NAPIRAhub) maintained by the JRC was released in 2009 as a comprehensive IT platform dedicated to the management of information on nanomaterials which are relevant for safety and risk assessment. Over the years it has been accumulating data from such areas as cosmetics, food and medical, as well as hosting a number of collections of research project data including NanoTEST, ENPRA, and the OECD Working Party on Manufactured Nanomaterials (WPMN). Principally now a nanomaterial information repository for OECD-WPMN projects, it was earlier foreseen as a key long term repository for FP7 project data. However uptake and use for the latter purpose has been very modest in recent years.

The database structure and principle of operation is based on that first used in the IUCLID system for REACH purposes. The structure is centred upon the nanomaterial, from which derive branches of results arranged according to standard OECD tests. However, although results data from several FP7 projects (eg ENPRA, NANOmmune, MARINA) have indeed been stored in the system, the database is not presented as a standard relational database, and the architecture is not optimised for such storage purposes. Furthermore, it is neither simple nor easy to use to add, search, link or retrieve data and secondary data users have found it cumbersome to operate for the purposes of accessing data for further analysis or modelling purposes. Projects are secured and partitioned separately, with permission (sought from the Project Leader) to access a particular project’s data being granted on a case by case basis. This also seriously hampers the ability to combine data through searches (for instance by type of nanomaterial) across different projects in the database.

It is recently reported that the role of this system will be subsumed into an updated and expanded EU IUCLID 5.6 / 6.0 system. It seems very unlikely that this system will be suitable to fulfil a role as a long-term repository for nano-EHS results data that can be directly stored and accessed for research purposes. However it may still be that some useful datasets could be made available for ProSafe data linkage purposes, although the effort needed to retrieve it may well outweigh the return.

2.7 The nano-EHS Landscape– An Overview and Observations

This section firstly outlines some additional information and knowledge sources that help to frame the current nano-EHS database Landscape. Then it aims to summarise and comment upon the view, attempting to outline some of the current outstanding needs and requirements, thereby highlighting the “gaps”. Whilst not so readily identifiable, it also presents some of the overlaps, joins or alignments between currently ongoing developments.

2.7.1 Key additional drivers and actors in the evolution of the nano-EHS database Landscape

Besides the specific projects or development activities described in the Mapping section 2.4 and in section 2.6 above, there are other important contributing factors that helped to produce and shape the contemporary nano-EHS database landscape. It is likely that these will continue to heavily influence the contours of the landscape. To help describe the overarching situation, or "bigger picture", besides examining specific project needs, T3.1 sought to answer the question: '*What are the more generic drivers and stimuli for continuing development and improvement in this domain?*'

Limiting this reasonably to the ambit of ProSafe WP3, we have in particular to take on board the key influences of nanoinformatics and its related technologies on nano-EHS research. Recent suggestions and proposed roadmaps (policy and strategy recommendations) from the nano-EHS community regarding the future of nano-EHS research and development, including the promulgation of Safe(r)-by-Design (SbD) as a core concept, also need to be taken into account. These and other ongoing discussions in the community have helped to highlight some of the perceived gaps and shortcomings in the nano-EHS database landscape, such as current data availability, system and format incompatibility, accessibility, future sustainability of DBs and more. Summary information about these influential areas is given below.

2.7.2 Nanoinformatics and related developments in the nano-EHS database field

The rapidly advancing field of nanoinformatics is obviously of tremendous importance in relation to any examination of the nano-EHS database landscape. It is after all in this context that the work of T3.1 has taken place and in which further tasks of WP3 are embedded. Whilst it is beyond the remit of this deliverable to provide a comprehensive review, nanoinformatics is intimately entwined with the mapping exercise and the ongoing developmental areas and issues covered in this report. Furthermore, it is more generally connected to the WP3 landscape and the ambition to promote and promulgate SbD through ProSafe and beyond. In the following paragraphs we very briefly summarise the nanoinformatics current role with reference to some of the most recent papers relevant to landscaping as identified via our literature search, described in section 2.5.

"*Nanoinformatics is the science and practice of determining which information is relevant to the nanoscale science and engineering community, and then developing and implementing effective mechanisms for collecting, validating, storing, sharing, analysing, modelling, and applying this information. Nanoinformatics also involves the utilization of networked communication tools to launch and support efficient communities of practice. Nanoinformatics is necessary for intelligent development and comparative characterization of nanomaterials, for design and use of optimized nanodevices and nanosystems, and for development of advanced instrumentation and manufacturing processes. Furthermore it fosters efficient scientific discovery and learning through data mining and machine learning techniques.*"[9]

The *Nanoinformatics 2020 Roadmap* [12] from 2011 describes the developing field at length describing its great potential and the way ahead. It sets out the many needs for nanomaterials data management including minimum information requirements, data collection, curation, integration, and sharing and accessibility.

Our relatively limited literature search yielded several papers in our bibliography (Annex 2) from the last 3-4 years wherein a core subject was the consideration of recent developments in Nanoinformatics, as well as several others more specifically addressing one or more nano-EHS-related data domains, which were also key-worded via "nanoInformatics" (or "nano-informatics"). The bibliography includes 17 publications author-key worded for nanoinformatics. This term is also cross linked with, or occasionally regarded as a branch of, bioinformatics, biomedical informatics, cheminformatics, computational biology, health informatics, systems biology and omics, amongst others.

[9] http://nanoinformatics.org/nanoinformatics/index.php/Main_Page

Several of the identified publications give insight to the steps that led to the contemporary database landscape and by discussing the related developments also highlight some of the perceived outstanding development requirements or gaps. Some of the most helpful historical perspectives of the development and evolution of nanoinformatics, including the involved challenges and perceived gaps, are provided in two papers by Maojo and colleagues [2,13]. These themes are continued and advanced in several other more recent papers in the two years prior to the main literature search (September 2015, but with some recent additions since then). For example, Ostraat et al. [14] describes the diversity and complexity of nanotechnology research data, outlines gaps in the published data and in standardized protocols and measurement schemes, and highlights the consequent need for minimal information standards, controlled vocabularies, standard data formats and publicly available analysis and modelling tools. Related discussions, though from different angles and including some particular solutions for some of the requirements within nanoinformatics, can also be found in other relevant publications [15–19].

2.7.3 Other recent nanoinformatics and nano-EHS related publications and information relevant to the landscape

Powers et al. [20] focused on critical connections for designing and implementing future nanomaterial research. The study describes how those in the ENM community have increasingly recognized two related, but distinct concerns:

a) Discordant data due to differences in experimental design or reporting, and

b) Lack of data to inform decisions about nano-EHS.

The authors are far from concerned solely with the technical database aspects of data and information exchange, and discuss in detail the barriers and opportunities for data exchange and sharing.

Hendren et al. [21], focus on the critical central role of data curation in facilitating nano-EHS data collection, data workflows, data combination and integration within the nanomaterial community's efforts to progress towards the functional interoperability of datasets. US Initiatives highlighted here include:

a) The RTI International Nanomaterial Registry (http://www.nanomaterialregistry.org);
b) The National Cancer Institute (NCI) Nanotechnology Characterization Lab (http://ncl.cancer.gov);
c) The Nanotechnology Knowledge Infrastructure (http://www.nano.gov/NSINKI); and
d) The Materials Genome Initiative (http://materialsinnovation.tms.org/genome.aspx).

Other important topical updates in the database landscape can be found in a more recent publication on nano-EHS informatics [22], which presents a thematic series of papers addressing the related subjects of nano-EHS databases, i.e. the eNanoMapper project [8] and others on nano-data curation, ISA-TAB-nano and the NanoPUZZLES project, nano-data mining and machine learning, and simulation and risk assessment tools. In addition, the emergence of "Big Data" combined with "Data Science" has also now become a significant element in this milieu, with the ability to organise and analyse masses of data from discrete and potentially disparate sources. This offers developmental, commercial and many other possible benefits that could be exploited in nano-EHS. Further, the ability to source and extract suitable data for analysis and modelling (i.e. scientific results from papers and other publications for nano-EHS and other data) is being rapidly enhanced by increasingly sophisticated textual analysis and data recognition algorithms with artificial intelligence techniques. This will continue to advance rapidly, opening up far greater opportunities for the curation and possible combination of related data from diverse sources, which can be exploited by members of the nano-EHS community.

2.7.4 The nano-EHS Community and EU NanoSafety Cluster

Whilst the role of the EU-promoted NanoSafety Cluster (NSC) was introduced in section 2.6.2 above in relation to its specific WG4, it also has provided substantial policy direction in terms of suggestions and roadmaps for future nano-EHS research. Of particular topical significance here is the recent publication "*Nanosafety in Europe 2015-2025: Towards Safe and Sustainable Nanomaterials and Nanotechnology Innovations*" by Savolainen et al. [23] It presents a Strategic Research Agenda (SRA) for future research on the safe use and safe applications of ENMs in four main thematic areas:

a) Nanomaterial identification and classification;
b) Nanomaterial exposure and transformation;
c) Hazard mechanisms related to effects on human health and the environment; and
d) Tools for the predictive risk assessment and management, including data standardisation, databases and ontologies.

In highlighting what is needed to build NM-based products that are safer by design, and propagate best practice in NMs safety (nanosafety) assessment, it is vital to bridge the current gap between nanotechnology developments and nanosafety assessment, for which quality information and data is crucial. The document suggests that the most practical solution requires a distributed set of facilities, linked by best practice. These facilities must share an underlying focus on quality and quality assurance, data quality and data sharing protocols, common databases to enable modelling and development of quantitative structure-property and/or structure-activity relationships (QPARs / QSARs). They should also work hand-in-glove with the regulatory authorities and relevant industry platforms, such as the European Technology Platforms and industry organisations.

2.7.5 Recent nano-EHS database related forums and initiatives relevant to the landscape

Apart from the previous mentioned actions, the NSC has organised several collaborative meetings to address the identified problems and issues around nano-EHS data and databases. Over the past few years, considerable concern has been voiced within the community over the lack of standardisation in data from EU funded nano-EHS research. The general inability to share and further exploit the large data volumes is a worry reflected in the requirements or shortcomings evidenced in the preceding sections of this deliverable. These have been re-emphasised by the NSC, given the increasing need to enhance the ability to re-use and exploit EU-generated scientific research data, along with recent requirements for the results of completed H2020 projects to be provided as "open data". Furthermore, a growing awareness of the role and objectives of the NSC, in particularl the stimuli (direct or indirect) provided by the work of the eNanoMapper project, have significantly nudged many in the community to give greater attention and practical input to furthering data sharing and collaboration.

In January two closely related workshops were organised:

a) An NSC workshop titled "*Data, Ontology and Harmonisation Needs for Nano Safety Cluster & Projects*" organised with the eNanoMapper project (Brussels, 25/01/16). This brought together many of the key players in the field. In the first part of the meeting the eNanoMapper project presented its ideas and progress in terms of ontology, database (design, procedures and APIs) and data collection (via ISA-TAB-nano) developments to date. The meeting then addressed some of the more generic shortcomings, or gaps, in current approaches to managing nano-EHS information and data through presentations, discussions and debate. ProSafe WP3 representatives attended the meeting.

b) A second follow-up workshop, entitled "*Knowledge Infrastructure And Framework For Nano Safety*" (Brussels, 26/01/16) expanded several of the themes of the previous day to:

i. Form an integrated perspective for a Nano Safety Knowledge Infrastructure (NSKI) supporting the work activities of all stakeholders involved in the research, assessment and regulation of the safety of nanotechnology;
ii. Define the critical components of such an NSKI, including stakeholders, expertise, frameworks, roadmaps, methods, resources, models and databases;
iii. Agree on how a common harmonisation approach to the NSKI, including consensus on open standards and ontology, could be best advanced;
iv. Formulate a plan for refining the NSKI both for NSC needs and for supporting international engagement, refinement, collaboration and EU-US cooperation (CoR).

The meeting included presentations of encountered nano-EHS data-related issues, suggestions, techniques and solutions from various projects and developments that could contribute towards standardisation and shared solutions across the community. There were further discussions on agreeing and pursuing shared goals, and thoughts on how to establish a future sharing and support network (http://www.enanomapper.net/events/knowledge-infrastructure-and-framework).

The subjects addressed and outcomes of these workshops are not repeated in detail here. However, the workshops, especially the first, exemplify a great deal about the current landscape, and several of the issues discussed are distilled and presented in section 2.8 below.

In another collaborative effort including discussion of nano-EHS-related items, NanoDefine organised the 2nd Nanosafety Cluster (NSC) Synergy Workshop (2nd February 2016) aiming to identify overlaps and synergies between different projects to enhance cooperation opportunities. A central issue was the building of a common ontology and a European framework for data management and analysis, as planned within eNanoMapper, to facilitate closer interdisciplinary collaboration between NSC projects and to better address the need for proper data storage, analysis and sharing.

These requirements have also been recently highlighted by the new Open Data Pilots and Open Access requirements for H2020 projects (https://www.openaire.eu/opendatapilot). These principles are further promoted in the recent "*Amsterdam Call for Action on Open Science*". Its action plan includes specific objectives to accelerate the transition to open science in Europe and contains four main goals:

1. Full open access for all publicly funded scientific publications by 2020;
2. Open data as the standard for all publicly funded research;
3. New assessment, reward and evaluation systems;
4. More open science to maximise its effectiveness and impact on society.

(http://www.eua.be/activities-services/news/newsitem/2016/04/14/dutch-presidency-issues-amsterdam-call-for-action-on-open-science).

Another development related to the Amsterdam Call and of relevance to the landscape and its future contours are "*The FAIR Guiding Principles for scientific data management and stewardship*", which have recently been fully described in detail [24]. These have been under discussion recently by a diverse set of stakeholders from academia, industry, funding agencies and publishers. The FAIR guiding principles are: *Findable*, *Accessible*, *Interoperable*, and *Reuseable*. They act as guidelines for those seeking to enhance the reusability of their data holdings. The principles put emphasis on improving the ability of machines to automatically find and use the data and in supporting its reuse by individuals. This is clearly of relevance to the developing nano-EHS landscape.

2.8 Current database issues and gaps – A summary gathering of items and commentary

From the above it is clear that the fairly generic problem issues in this area (information standards, controlled vocabularies, standard data formats and sharing, publicly available analysis and modelling tools) have been raised frequently and recurrently. The earlier Mapping exercise also shows that whilst these issues have been acknowledged for several years, there are few projects to date that have adopted unified or shared approaches to nano-EHS data management. We are now seeing steps in that direction in some projects collaborations with eNanoMapper. This is not to suggest that individual projects hitherto have been beset by such problems or have necessarily suffered as a result. However, efforts to meet good data management needs have been duplicated across (many) different projects, where ideally they could have been shared, with mutual benefits.

There are many possible factors which could have contributed to this situation, and whilst knowledge of them may well help in devising solutions, we do not elaborate them in detail here. The scale of data management tasks required and undertaken has varied widely according to the subject and size of a project. The same applies also to the priorities and resources devoted to "data management" on a project wide basis, whether by design from the outset, or carried out on a more or less ad-hoc basis within separate (sometimes somewhat isolated) work packages. This used to be sufficient to meet the needs of a Project's DoW. Many past projects have progressed well enough in their own territories, but data has frequently been left in distinct separate silos, hampering the ability to share data across the project's own disciplines, let alone between different projects or, following completion of the project, with the wider scientific or public community.

To help in providing some more discrete "metrics" on the current landscape, key issues or problems pertaining to the nano-EHS Landscape that have been covered above and many of which were also addressed at the NSC meeting (2.7.5.a above) are summarised below in Table 12, first column. Some commentary on current initiatives to address the need, or on outstanding action required, is given in the second column.

Table 12. Summary of key issues and requirements and related current initiatives in the nano-EHS Landscape, as evidenced through the NSC.

Nano-EHS Landscape issue or development need	Current initiatives or *outstanding action*
No standardised language, i.e. ontology or controlled vocabulary, to classify and describe data yet available to be applied as standard	Developments ongoing in both US and EU. eNanoMapper draft ontology under continuing development and promotion/training to FP7 / H2020 projects and beyond. Linked intimately with use of ISA-TAB-Nano format. Development of ontology extensions for nano-exposure data with NSC WG3 & NECID. *Not necessarily easy to use for average data generator: help and guidance and very good integration with end user tools will be required.*
A lack of standardised data across the nano-EHS data domains. Diverse and poorly described formats and layouts. None or poorly described meta-data / documentation; needs expansion for further data types.	To be pursued by consistent use of ontology in data definitions and ISA-TAB-nano format as basis of data capture format. *Extend to other data types or develop analogous alternatives*
Need for data capture front-ends and data collection templates. For the end user needs to be lightweight, intuitive and easy to use. ISA-TAB-nano format alone not easy to use and a potential barrier to uptake.	eNanoMapper providing for parsing of other templates into a NSC-wide DB instance (at present at eNM), and underlying ISA-TAB-nano format. See for instance data from MARINA templates. Follow the NANoREG example and build on its open access and free templates, based on a simplified, user-friendly adaptation of the ISA-TAB-Nano format to record *experimental* data.

	Need to develop other front ends with shared standards for a variety of templates for different data types and situations. See NANoREG to date. *For occupational exposure, pursue the eNanoMapper-NECID work on ontology mapping.* *Explore the appropriateness of connecting to ontology and possibly front end data capture via eNanoMapper APIs* *Disconnected remote/off-line data capture modes needed.* *End user training and education requirements. Resources, materials, guidance, data templates and templates for data planning and workflows.*
Wide diversity in database construction and use: structurally and functionally. Database being re-invented.	eNanoMapper database instance to provide a consistent NSC-level model; to date for phys-chem and (eco)tox data. *Possible extensions for other data types domains.* *Making this available and understood. Mediating updates and consistency if DBs localised / distributed.* *Make DBs/data models, workflows descriptions and models, technology-agnostic, i.e. not dependant on particular DB brand or technology.*
No consistent, well-organised, shared approaches or accepted discipline applied widely for planning and managing data collection, processing and management of the data. No coherent approach to good nano-EHS Data Management Planning (DMP) and practice. This is needed for Open Data standards.	Conceptually being addressed, in examining / providing resources that can be shared, by the NSKI (EU), the CEINT Nanoinformatics Knowledge Commons (US) and the US-EU CoR. *Need to develop standardised approaches, resources, (templates, design guides etc) with training and guidance the can be adopted for use consistently throughout the community.*
Difficulty in discovering, obtaining (permission), accessing and retrieving data, even if it was "properly" databased. Inability to exploit or combine valuable EU-funded dataset for mutual and public benefits. Need to meet new "Open Data" requirements. Currently no accepted route or resources for long-term data availability and sustainability. Data not submitted (recorded) for IPR reasons.	NSC charged with examining sustainability and funding. *Possibly establish new dedicated Sustainability WG to investigate and consult on strategy and resourcing.* *Adoption of standards and consistency to be enhanced, as well as the ability to locate data in a well-resourced (sustainable) repository.* *Need for a data sharing charter, possibly in '*Research Data Alliance (RDA)*'[10]-style.* *Enhancement of implementation of data as part of peer-review process.*
A lack of interoperability and re-usability of ontologies, datasets, DBs, IT tools and data management procedures.	eNanoMapper to help standardise. *Consistency and compatibility required in*

[10] https://rd-alliance.org/about-rda/who-rda.html

	databases implementation and data collection formats. Involve ontologies, APIs and possible exchange formats. *Data managers and end user (data collectors, curators, generators) training and education requirements. Resources, materials, guidance across the whole lifecycle for each and every data domain*
A wide diversity in (nano-EHS) modelling approaches and analysis requirements across the disciplines, which could be better catered for, such as in statistics, risk assessment, decision support systems, …	Discussions ongoing with modellers and eNanoMapper on facilitating data access for various different analysis. Some QSAR examples established

2.9 Summary of conclusions and next steps in ProSafe WP3

This section provides a general summary of the key conclusion from the preceding sections and also provides some suggestions on the next steps to be taken for WP3 and beyond.

T3.1 successfully carried out a comprehensive mapping exercise of contemporary nano-EHS database developments and related implementations at both EU level and beyond. It includes a detailed breakdown of the results and supporting observations as presented in sections 2.4, 2.7 and 2.8. The underlying data is available in Annexes 1 and 2 (section 5). Key points arising from this exercise are:

i. The information describing the availability of databases produced/generated in projects and useful to the mapping exercise was highly variable in both quality and quantity. This highlights the needs for better general information about projects, and their data management aspects that will allow them to be identified as viable sources for data exchange and linking.

ii. It is suggested that this information may be enhanced in the short term by updating some of the mapping data through an enhanced NSC WG4 database survey. Also, in the future the project information collection exercise used by the NSC should be improved to include expanded and more structured information about data management and database aspects. An revised database survey is now under discussion within NSC WG4.

iii. Enhanced database mapping information would be a very useful resource and it is further suggested that this could be provided as a relatively simple online resource, if it could be periodically maintained to continue to provide current state-of-the-art information on data availability, ontology use, formatting, potential for access, data exchange and other details of interest.

iv. Even for databases that could be readily identified, it is clear that very few are currently sufficiently compatible in formats to promptly allow data availability and exchange in a harmonised way. Some are exploring new approaches and collaborations. Whilst these issues have been acknowledged for several years, there are few projects to date that have adopted unified or shared approaches to nano-EHS data management.

v. Of the ongoing NSC-related projects just seven, not including eNanoMapper, are identified as having some potential for data exchange that might be developed in ProSafe (table 10). eNM indeed may be the “bridge” to exchange or link data in the short term via ongoing collaborations, and this is also highlighted in that project DoW.

vi. In addition, other nano-EHS database-related activities and advances, which are relevant to the nano-EHS database landscape and ProSafe actions and may have some potential for further interaction and possible data linkage or exchange, are cited in section 2.6.

T3.1 successfully carried out a comprehensive examination and description of the contemporary nano-EHS database landscape by examining the wider context of this domain. This covered not only ongoing theoretical and practical developments in this field, but also related data management and informatics areas with shared and overlapping issues and requirements, and similar aims and ambitions for their resolution.

i. The landscape description was informed initially by the results of the DB Mapping exercise, and substantially supplemented through a far wider examination of the literature and related information on recent progress in the field, including contact and interaction with other activities and initiatives.

ii. It is suggested that the bibliography could be enhanced and maintained for the benefit of the community if some resource for occasional maintenance could be obtained. This might be accomplished under the auspices of NSC WG4, potentially merging similar resources obtained by research in the eNanoMapper, or other projects.

iii. The Landscape reveals shortcomings and gaps in the basic requirements for an improved approach to nano-EHS data management, most of which are typified by many of the results of the Mapping exercise.

iv. Table 12 in section 2.8 provides a summary of these contemporary gaps and outstanding issues, but also shows several positive actions towards harmonisation that are well underway and making significant inroads: the new requirements for DMP and Open Data for EU projects and related harmonisation initiatives, such as FAIR, the development of an improved Data Publication and Citation framework, crucial work by eNanoMapper and collaborating projects and, last but not least, the EU-US collaboration in the CoR.

These developments will address many of the issues over the coming period, but there will remain an ongoing requirement for continuing maintenance, review and improvements of data and sharing systems. Whilst good data management practice in this field ought to be essential, it is not an end in itself. It needs to be supported and sustained in real projects by adequate resourcing and buy-in by all stakeholders. The mutual benefits of enhanced data quality, reliability, accessibility and reusability continue to need to be demonstrated and promulgated by the scientists, administrators and other secondary data users on an ongoing basis.

At the latest ProSafe project meeting (February 2016) WP3 chaired an open discussion on data management issues and continuing liaison with eNanoMapper and other interested parties. Notes of this are under preparation and together with the results of this deliverable will feed directly into the next steps in ProSafe WP3, and into the bigger framework of ProSafe deliverables and their wider impact through dissemination and or publication.

3 Deviations from the work plan

Task 3.1 as described in the DoA has been comprehensively executed with all proposed activities pursued and achieved.

The timetable for the work was altered and delivery of this report has been delayed. Following a three months delay in commencing the task at the start of the project after the reallocation of tasks amongst partners, then a further unforeseen delay caused by the two month sickness absence of the IOM Task leader in autumn 2015, the MC agreed to postpone D3.1 delivery until 25th April; and again until 3rd June.

It was agreed by partners in WP3 that the delay does not materially affect the progress of other tasks in WP3 or the project.

Heavy partner workloads delayed final review and revision until early July 2016. The final agreed report is submitted to MC 7th July 2016.

After feedback from the MC revisions made (adding section 2.2.2) and final report is re-submitted to MC 22nd September 2016.

4 List of abbreviations

CEINT	Centre for Environmental Implication of NanoTechnology
CoR	US-EU Community of Research
DB	Database
DBMS	Database Management System
DMP	Data Management Planning
DoA	Description of Activitiy
ENM	Engineered Nanomaterial
eNM	The FP7 eNanoMapper project
FP7	The EU's Seventh Framework Programme for Research and Technological Development 2007-2013
H2020	Horizon 2020 - The EU's Framework Programme for Research and Innovation 2014-2020
IOM	Institute of Occupational Medicine (United Kingdom)
JRC	The European Commission's Joint Research Centre, Ispra site (Italy)
NECID	Nano Exposure & Contextual Information Database
NM	Nanomaterial
NSC	The EU-promoted NanoSafety Cluster
NSKI	Nano Safety Knowledge Infrastructure
SbD	Safe(r) by Design
SRA	Strategic Research Agenda

5 Annexes

ANNEX 1: Spreadsheet of complete inventory of projects investigated for presence of nano-EHS databases of interest.

An Excel® spreadsheet file of the complete inventory of projects investigated for presence of nano-EHS databases of interest is embedded in this report and available by clicking the icon below.

Prosafe-T3.1-Database mapping_01_2016

ANNEX 2: T3.1 bibliography compiled via the literature search

1	Alexis F, Rhee J-W, Richie JP, *et al.* New frontiers in nanotechnology for cancer treatment. *Urol Oncol Semin Orig Investig* 2008;**26**:74–85. doi:10.1016/j.urolonc.2007.03.017
2	Alonso VL, Cambero JG. Informatic resources useful for the development of biomedical applications of nanomaterials and their toxicological evaluation. Rev. Toxicol. 2014;**31**:23–30.http://ojs.diffundit.com/index.php/revtoxicol/article/view/509 (accessed 17 May2016).
3	Artemieva IL, Ryabchenko N V. Nanomaterials Ontology Model. *Adv Mater Res* 2014;**905**:65–9. doi:10.4028/www.scientific.net/AMR.905.65
4	Bates ME, Larkin S, Keisler JM, *et al.* How decision analysis can further nanoinformatics. *Beilstein J Nanotechnol* 2015;**6**:1594–600. doi:10.3762/bjnano.6.162
5	Bekker C, Kuijpers E, Brouwer DH, *et al.* Occupational Exposure to Nano-Objects and Their Agglomerates and Aggregates Across Various Life Cycle Stages; A Broad-Scale Exposure Study. *Ann Occup Hyg* 2015;**59**:681–704. doi:10.1093/annhyg/mev023
6	Brouwer D, Berges M, Virji MA, *et al.* Harmonization of Measurement Strategies for Exposure to Manufactured Nano-Objects; Report of a Workshop. *Ann Occup Hyg* 2012;**56**:1–9. doi:10.1093/annhyg/mer099
7	Commission E, Initiative USNN. 2015 EU-U.S.: Bridging NanoEHS Research Efforts Joint Workshop Proceedings. 2015.
8	de la Calle G, García-Remesal M, Nkumu-Mbomio N, *et al.* e-MIR2: a public online inventory of medical informatics resources. *BMC Med Inform Decis Mak* 2012;**12**:82. doi:10.1186/1472-6947-12-82
9	de la Iglesia D, Cachau RE, García-Remesal M, *et al.* Nanoinformatics knowledge infrastructures: bringing efficient information management to nanomedical research. *Comput Sci Discov* 2013;**6**:014011. doi:10.1088/1749-4699/6/1/014011
10	de la Iglesia D, Harper S, Hoover M. Nanoinformatics 2020 Roadmap. Amherst, MA 01003, USA: 2011. doi:10.4053/rp001-110413
11	de la Iglesia D, Cachau RE, García-Remesal M, *et al.* Nanoinformatics knowledge infrastructures: bringing efficient information management to nanomedical research. *Comput Sci Discov* 2013;**6**:014011. doi:10.1088/1749-4699/6/1/014011
12	García-Remesal M, García-Ruiz A, Pérez-Rey D, *et al.* Using Nanoinformatics Methods for Automatically Identifying Relevant Nanotoxicology Entities from the Literature. *Biomed Res Int* 2013;**2013**:1–9. doi:10.1155/2013/410294
13	Gorman Ng M, Semple S, W. Cherrie J, *et al.* The Relationship Between Inadvertent Ingestion and Dermal Exposure Pathways: A New Integrated Conceptual Model and a Database of Dermal and Oral Transfer Efficiencies. *Ann Occup Hyg* 2012;**56**:1000–12. doi:10.1093/annhyg/mes041
14	Grieger KD, Redmon JH, Money ES, *et al.* A relative ranking approach for nano-enabled applications to improve risk-based decision making: a case study of Army materiel. *Environ Syst Decis* 2015;**35**:42–53. doi:10.1007/s10669-014-9531-4
15	Grulke CM, Goldsmith M-R, Vallero DA. Toward a Blended Ontology: Applying Knowledge Systems to Compare Therapeutic and Toxicological Nanoscale Domains. *J Biomed Biotechnol* 2012;**2012**:1–9. doi:10.1155/2012/308381
16	Harper SL, Hutchison JE, Baker N, *et al.* Nanoinformatics workshop report: current resources, community needs and the proposal of a collaborative framework for data sharing and information integration. *Comput Sci Discov* 2013;**6**:014008. doi:10.1088/1749-4699/6/1/014008
17	Hastings J, Jeliazkova N, Owen G, *et al.* eNanoMapper: harnessing ontologies to enable data

	integration for nanomaterial risk assessment. *J Biomed Semantics* 2015;**6**:10. doi:10.1186/s13326-015-0005-5
18	He L. *Building a database for nanomaterial exposure*. Durham, North Carolina: : Duke University 2015.
19	Hendren CO, Lowry G V., Unrine JM, *et al.* A functional assay-based strategy for nanomaterial risk forecasting. *Sci Total Environ* 2015;**536**:1029–37. doi:10.1016/j.scitotenv.2015.06.100
20	Hendren CO, Powers CM, Hoover MD, *et al.* The Nanomaterial Data Curation Initiative: A collaborative approach to assessing, evaluating, and advancing the state of the field. *Beilstein J Nanotechnol* 2015;**6**:1752–62. doi:10.3762/bjnano.6.179
21	Jeliazkova N, Chomenidis C, Doganis P, *et al.* The eNanoMapper database for nanomaterial safety information. *Beilstein J Nanotechnol* 2015;**6**:1609–34. doi:10.3762/bjnano.6.165
22	Jeliazkova N, Doganis P, Fadeel B, *et al.* The first eNanoMapper prototype: A substance database to support safe-by-design. In: *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE 2014. 1–9. doi:10.1109/BIBM.2014.6999367
23	Juganson K, Ivask A, Blinova I, *et al.* NanoE-Tox: New and in-depth database concerning ecotoxicity of nanomaterials. *Beilstein J Nanotechnol* 2015;**6**:1788–804. doi:10.3762/bjnano.6.183
24	Jung S-K, Qu X, Aleman-Meza B, *et al.* Multi-endpoint, High-Throughput Study of Nanomaterial Toxicity in Caenorhabditis elegans. *Environ Sci Technol* 2015;**49**:2477–85. doi:10.1021/es5056462
25	Klaessig FC. Developing Official Practices for nanoEHS Data Compilation , Curation and Compliance. In: Coenen C, Dijkstra A, Fautz C, *et al.*, eds. *Innovation and responsibility: engaging with new and emerging technologies*. Berlin, Germany: : Akademische Verlagsgesellschaft AKA GmbH 2014. 121–33.
26	Kuempel ED, Castranova V, Geraci CL, *et al.* Development of risk-based nanomaterial groups for occupational exposure control. *J Nanoparticle Res* 2012;**14**:1029. doi:10.1007/s11051-012-1029-8
27	Lavoue J, Friesen MC, Burstyn I. Workplace Measurements by the US Occupational Safety and Health Administration since 1979: Descriptive Analysis and Potential Uses for Exposure Assessment. *Ann Occup Hyg* 2013;**57**:77–97. doi:10.1093/annhyg/mes055
28	Lewinski NA, McInnes BT. Using natural language processing techniques to inform research on nanotechnology. *Beilstein J Nanotechnol* 2015;**6**:1439–49. doi:10.3762/bjnano.6.149
29	Li X, Hu D, Dang Y, *et al.* Nano Mapper: an Internet knowledge mapping system for nanotechnology development. *J Nanoparticle Res* 2009;**11**:529–52. doi:10.1007/s11051-008-9491-z
30	Liu R, Cohen Y. Nanoinformatics for environmental health and biomedicine. *Beilstein J Nanotechnol* 2015;**6**:2449–51. doi:10.3762/bjnano.6.253
31	Liu X, Webster T. Nanoinformatics for biomedicine: emerging approaches and applications. *Int J Nanomedicine* 2013;**8**:1. doi:10.2147/IJN.S41253
32	López-Alonso C, García-Cambero J. Informatic resources useful for the development of biomedical applications of nanomaterials and their toxicological evaluation. *Rev Toxicol* 2014;**31**:23–30.
33	Manonmani V, Anitha U. Nanoinformatics: A conjecture and a review. *Res J Pharm Biol Chem Sci* 2015;**6**:631–5.http://www.researchgate.net/publication/258507801_Acute_Hemorrhagic_Cystitis_in_Children_A_Case_Series

34	Maojo V, Fritts M, Martin-Sanchez F, *et al.* Nanoinformatics: developing new computing applications for nanomedicine. *Computing* 2012;**94**:521–39. doi:10.1007/s00607-012-0191-2
35	Maojo V, Fritts, de la Iglesia D, *et al.* Nanoinformatics: a new area of research in nanomedicine. *Int J Nanomedicine* 2012;**7**:3867. doi:10.2147/IJN.S24582
36	Marchese Robinson RL, Cronin MTD, Richarz A-N, *et al.* An ISA-TAB-Nano based data collection framework to support data-driven modelling of nanotoxicology. *Beilstein J Nanotechnol* 2015;**6**:1978–99. doi:10.3762/bjnano.6.202
37	Marchese Robinson RL, Cassano A. All NanoPUZZLES ISA-TAB-Nano datasets. Published Online First: 2015. doi:10.5281/zenodo.35493
38	Mattingly CJ, McKone TE, Callahan MA, *et al.* Providing the Missing Link: the Exposure Science Ontology ExO. *Environ Sci Technol* 2012;**46**:3046–53. doi:10.1021/es2033857
39	Mills KC, Murry D, Guzan KA, *et al.* Nanomaterial registry: database that captures the minimal information about nanomaterial physico-chemical characteristics. *J Nanoparticle Res* 2014;**16**:2219. doi:10.1007/s11051-013-2219-8
40	Morris SA, Gaheen S, Lijowski M, *et al.* caNanoLab: A nanomaterial data repository for biomedical research. In: *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE 2014. 29–33. doi:10.1109/BIBM.2014.6999371
41	Morris SA, Gaheen S, Lijowski M, *et al.* Experiences in supporting the structured collection of cancer nanotechnology data using caNanoLab. *Beilstein J Nanotechnol* 2015;**6**:1580–93. doi:10.3762/bjnano.6.161
42	Muñoz-Mármol M, Crespo J, Fritts MJ, *et al.* Towards the taxonomic categorization and recognition of nanoparticle shapes. *Nanomedicine Nanotechnology, Biol Med* 2015;**11**:457–65. doi:10.1016/j.nano.2014.07.006
43	Mustad AP, Smeets B, Jeliazkova N, *et al.* Summary of the Spring 2014 NSC Database Survey. 2014. doi:10.6084/m9.figshare.1195888
44	Orlowski E, Audignon-Durand S, Goldberg M, *et al.* EV@LUTIL: An open access database on occupational exposures to asbestos and man-made mineral fibres. *Am J Ind Med* 2015;**58**:1059–74. doi:10.1002/ajim.22498
45	Ostraat ML, Mills KC, Guzan K, *et al.* The Nanomaterial Registry: facilitating the sharing and analysis of data in the diverse nanomaterial community. *Int J Nanomedicine* 2013;**8**:7. doi:10.2147/IJN.S40722
46	Panneerselvam S, Choi S. Nanoinformatics: Emerging Databases and Available Tools. *Int J Mol Sci* 2014;**15**:7158–82. doi:10.3390/ijms15057158
47	Pelzer J. Struktur und Funktionalität der Nano-Expositions- datenbank NECID (Nano and Contextual Information Database). *Gefahrstoffe – Reinhaltung der Luft* 2013;**73**:302–4.
48	Pelzer J, Fransman W, Stöppelmann W, *et al.* Development of a Nano Exposure and Contextual Information Database (NECID). 2014;**2014**:7.
49	Powers CM, Grieger KD, Beaudrie C, *et al.* Data dialogues: critical connections for designing and implementing future nanomaterial research. *Environ Syst Decis* 2015;**35**:76–87. doi:10.1007/s10669-014-9518-1
50	Powers CM, Mills KA, Morris SA, *et al.* Nanocuration workflows: Establishing best practices for identifying, inputting, and sharing data to inform decisions on nanomaterials. *Beilstein J Nanotechnol* 2015;**6**:1860–71. doi:10.3762/bjnano.6.189
51	Richarz A-N, Madden JC, Marchese Robinson RL, *et al.* Development of computational models for the prediction of the toxicity of nanomaterials. *Perspect Sci* 2015;**3**:27–9. doi:10.1016/j.pisc.2014.11.015
52	Rumble J, Freiman S. Describing Nanomaterials: Meeting the Needs of Diverse Data

	Communities. *Data Sci J* 2012;**11**:ASMD1–6. doi:10.2481/dsj.AMDS2012_Rumble-001
53	Savolainen K, Backman U, Brouwer D, *et al. Nanosafety in Europe 2015-2025: Towards Safe and Sustainable Nanomaterials and Nanotechnology Innovations Nanosafety in Europe Towards Safe and Sustainable Nanomaterials and Nanotechnology Innovations*. Helsinki, Finland: : Finnish Institute of Occupational Health 2013. www.ttl.fi/en/publications/electronic_publications/pages/default.aspx\nThis
54	Sayes C, Smith, Ivanov. A framework for grouping nanoparticles based on their measurable characteristics. *Int J Nanomedicine* 2013;**8**:45. doi:10.2147/IJN.S40521
55	Tang K, Liu X, Harper SL, *et al.* NEIMiner: Nanomaterial environmental impact data miner. *Int J Nanomedicine* 2013;**8**:15–29. doi:10.2147/IJN.S40974
56	Thomas DG, Pappu R V., Baker NA. NanoParticle Ontology for cancer nanotechnology research. *J Biomed Inform* 2011;**44**:59–74. doi:10.1016/j.jbi.2010.03.001
57	Thomas DG, Gaheen S, Harper SL, *et al.* ISA-TAB-Nano: A Specification for Sharing Nanomaterial Research Data in Spreadsheet-based Format. *BMC Biotechnol* 2013;**13**:2. doi:10.1186/1472-6750-13-2
58	Vance ME, Kuiken T, Vejerano EP, *et al.* Nanotechnology in the real world: Redeveloping the nanomaterial consumer products inventory. *Beilstein J Nanotechnol* 2015;**6**:1769–80. doi:10.3762/bjnano.6.181
59	Wijnhoven SWP, Oomen AG, Bourgeois FC, *et al.* Development of an inventory for consumer products containing nanomaterials. 2011. http://ec.europa.eu/environment/chemicals/nanotech/pdf/study_inventory.pdf
60	Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018. doi:10.1038/sdata.2016.18

ANNEX 3: List of references cited in text of D3.1

1 Mustad AP, Smeets B, Jeliazkova N, *et al.* Summary of the Spring 2014 NSC Database Survey. 2014. doi:10.6084/m9.figshare.1195888

2 Maojo V, Fritts, de la Iglesia D, *et al.* Nanoinformatics: a new area of research in nanomedicine. *Int J Nanomedicine* 2012;**7**:3867. doi:10.2147/IJN.S24582

3 Panneerselvam S, Choi S. Nanoinformatics: Emerging Databases and Available Tools. *Int J Mol Sci* 2014;**15**:7158–82. doi:10.3390/ijms15057158

4 Jeliazkova N, Chomenidis C, Doganis P, *et al.* The eNanoMapper database for nanomaterial safety information. *Beilstein J Nanotechnol* 2015;**6**:1609–34. doi:10.3762/bjnano.6.165

5 Hastings J, Jeliazkova N, Owen G, *et al.* eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment. *J Biomed Semantics* 2015;**6**:10. doi:10.1186/s13326-015-0005-5

6 Thomas DG, Pappu R V., Baker NA. NanoParticle Ontology for cancer nanotechnology research. *J Biomed Inform* 2011;**44**:59–74. doi:10.1016/j.jbi.2010.03.001

7 Pelzer J. Struktur und Funktionalität der Nano-Expositions- datenbank NECID (Nano and Contextual Information Database). *Gefahrstoffe – Reinhaltung der Luft* 2013;**73**:302–4.

8 Jeliazkova N, Doganis P, Fadeel B, *et al.* The first eNanoMapper prototype: A substance database to support safe-by-design. In: *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE 2014. 1–9. doi:10.1109/BIBM.2014.6999367

9 Thomas DG, Gaheen S, Harper SL, *et al.* ISA-TAB-Nano: A Specification for Sharing Nanomaterial Research Data in Spreadsheet-based Format. *BMC Biotechnol* 2013;**13**:2. doi:10.1186/1472-6750-13-2

10 Marchese Robinson RL, Cronin MTD, Richarz A-N, *et al.* An ISA-TAB-Nano based data collection framework to support data-driven modelling of nanotoxicology. *Beilstein J Nanotechnol* 2015;**6**:1978–99. doi:10.3762/bjnano.6.202

11 Marchese Robinson RL, Cassano A. All NanoPUZZLES ISA-TAB-Nano datasets. Published Online First: 2015. doi:10.5281/zenodo.35493

12 D De La Iglesia, S Harper MH. Nanoinformatics 2020 Roadmap. 2011. 63. doi:10.4053/rp001-110413

13 Maojo V, Fritts M, Martin-Sanchez F, *et al.* Nanoinformatics: developing new computing applications for nanomedicine. *Comput Sci Eng* 2012;**94**:521–39. doi:10.1007/s00607-012-0191-2

14 Ostraat ML, Mills KC, Guzan K, *et al.* The Nanomaterial Registry: facilitating the sharing and analysis of data in the diverse nanomaterial community. *Int J Nanomedicine* 2013;**8**:7. doi:10.2147/IJN.S40722

15 Liu X, Webster T. Nanoinformatics for biomedicine: emerging approaches and applications. *Int J Nanomedicine* 2013;**8**:1. doi:10.2147/IJN.S41253

16 Harper SL, Hutchison JE, Baker N, *et al.* Nanoinformatics workshop report: current resources, community needs and the proposal of a collaborative framework for data sharing and information integration. *Comput Sci Discov* 2013;**6**:14008. doi:10.1088/1749-4699/6/1/014008

17 de la Iglesia D, Cachau RE, García-Remesal M, *et al.* Nanoinformatics knowledge infrastructures: bringing efficient information management to nanomedical research. *Comput Sci Discov* 2013;**6**:14011. doi:10.1088/1749-4699/6/1/014011

18 de la Iglesia D, Cachau RE, García-Remesal M, *et al.* Nanoinformatics knowledge infrastructures: bringing efficient information management to nanomedical research. *Comput Sci Discov* 2013;**6**:14011. doi:10.1088/1749-4699/6/1/014011

19 López-Alonso C, García-Cambero JP. Informatic resources useful for the development of biomedical applications of nanomaterials and their toxicological evaluation. *Rev Toxicol* 2014;**31**:23–30.

20 Powers CM, Grieger KD, Beaudrie C, *et al.* Data dialogues: critical connections for designing and implementing future nanomaterial research. *Environ Syst Decis* 2015;**35**:76–87. doi:10.1007/s10669-014-9518-1

21 Hendren CO, Powers CM, Hoover MD, *et al.* The Nanomaterial Data Curation Initiative: A collaborative approach to assessing, evaluating, and advancing the state of the field. *Beilstein J Nanotechnol* 2015;**6**:1752–62. doi:10.3762/bjnano.6.179

22 Liu R, Cohen Y. Nanoinformatics for environmental health and biomedicine. *Beilstein J Nanotechnol* 2015;**6**:2449–51. doi:10.3762/bjnano.6.253

23 Savolainen K, Backman U, Brouwer D, *et al. Nanosafety in Europe 2015-2025: Towards Safe and Sustainable Nanomaterials and Nanotechnology Innovations Nanosafety in Europe Towards Safe and Sustainable Nanomaterials and Nanotechnology Innovations*. 2013. www.ttl.fi/en/publications/electronic_publications/pages/default.aspx\nThis

24 Wilkinson MD, Dumontier M, Aalbersberg IjJ, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;**3**:160018. doi:10.1038/sdata.2016.18