DELIVERABLE REPORT D1.1

Requirements Analysis and System Design

GRANT AGREEMENT:	604134
ACRONYM:	eNanoMapper
NAME:	eNanoMapper - A Database and Ontology Framework for Nanomaterials Design and Safety Assessment
PROJECT COORDINATOR:	Douglas Connect GmbH
START DATE OF PROJECT; DURATION:	1 February 2014; 36 months
PARTNER(s) RESPONSIBLE FOR THIS DELIVERABLE:	DC Douglas Connect
DATE:	13.3.2015
VERSION:	V.2.1
LICENSE	CC BY 4.0

SEVENTH FRAMEWORK PROGRAMME

Call identifier	FP7-NMP-2013-SMALL-7
Document Type	Deliverable Report
WP/Task	WP1 / Community outreach
Document ID	eNanoMapper D1.1
Status	Final Draft

Partner Organisations	• Douglas Connect, GmbH (DC) • National Technical University of Athens (NTUA) • In Silico Toxicology (IST) • Ideaconsult (IDEA) • KarolinskaInstitutet (KI) • VTT Technical Research Centre of Finland (VTT) • European Bioinformatics Institute (EMBL-EBI) • Maastricht University (UM)
Authors	Markus Hegi (DC) Barry Hardy (DC) Egon Willighagen (UM) Review by all eNanoMapper Consortium
Purpose of the Document	To report on eNanoMapper Requirements Analysis and Systems Design
Document History	• Initial Draft: 18.9.2014 • Second Draft: 26.1.2015 • Final Draft 2.0: 23.2.2015 • Release Candidate 2.1 13.3.2015

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY 7
   1.1 ABOUT ENANOMAPPER 7
2. INTRODUCTION 9
   2.1 SCIENTIFIC IMMATURITY 9
3. CONTEXTUAL DESIGN PROCESS 11
4. INTERVIEWS 12
   4.1 47 INTERVIEWS ACROSS THE NANOSAFETY CLUSTER 12
   4.2 INTERVIEW METHODOLOGY 12
5. INTERPRETATION *DEBRIEFINGS* 14
6. CONSOLIDATION 15
   6.1 NOTES CONSOLIDATION 15
   6.2 IMMATURITY: "BAD SCIENCE" 15
   6.1 "SCIENTISTS DON'T SHARE" 16
   6.2 HARMONIZATION NEED 16
      6.2.1 The Language: Ontologies 16
      6.2.1 Processes: SOP and protocol 17
      6.2.2 Interfaces: API 17
      6.2.3 Data 18
   6.3 ENANOMAPPER - A DATABASE OR A STANDARD? 18
   6.4 SEARCH 19
7. USE CASES 20
   7.1 LIST AND SPECIFICATION OF USE CASES 20
   7.2 EXAMPLE USE CASE 23: SEARCH NANO-MATERIAL X 20
8. SYSTEM DESIGN 22
      8.1.1 Introduction 22
   8.2 SYSTEM ARCHITECTURE CONTEXT DIAGRAM 23
      8.2.1 The REST API 23
      8.2.2 The services 24
      8.2.3 Data Store 24
9. PAPER PROTOTYPING 26
   9.1 SCOPE OF PAPER PROTOTYPES 26
   9.2 REFINING AND TESTING THE PROTOTYPES 26
   9.3 EXAMPLE: PAPER PROTOTYPE FOR USE CASE 23 26

9.3.1 Use Cases covered by this prototype *26*
9.3.2 Use Cases, PARTLY covered by this prototype: *27*
9.3.3 Persona / Flow-Diagram *27*
9.3.4 Screen Flow *28*
9.4 TECHNICAL ARCHITECTURE 28
10. CONCLUSION 30
11. BIBLIOGRAPHY 32
12. ANNEXES 34
12.1 AFFINITY DIAGRAM 34

TABLE OF FIGURES

Figure 1 Contextual Design: The base of our development 7
Figure 2: The main project challenge: Scientific Immatuirity. Design Workshop 18-21.6.2014 Milano &Ispra 9
Figure 3 Contextual Design as the base of our development. 11
Figure 4: Use Cases 20
Figure 5: System Design 22
Figure 6: System Architecture Context Diagram 23
Figure 7: Persona / Flow-Diagram 27
Figure 8: Screen flow 28
Figure 9: Technical Architecture 28

GLOSSARY

Abbreviation / acronym	Description
API	Application Programming Interface
DB	Database
ENM	Engineered Nano Material
GUI	Graphical User Interface
ISA-TAB	ISA stands for Investigation, Study, Assay – ISA-TAB is a Tabular format to provide rich description of the experimental metadata - see http://isa-tools.org
QSAR	Quantitative Structure-Activity Relationship
REST	REpresentational State Transfer
SEO	Search Engine Optimization
SOP	Standard Operating Procedure
UI	User Interface
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
WP	Work Package

1. EXECUTIVE SUMMARY

The objective of WP 1 is, to enable the communication between the eNanoMapper project and the wider nano-safety community and to ensure the alignment of the objectives of eNanoMapper with the needs of the community. At the start of the project, an extensive requirements analysis was conducted and based on the results, an overall system design was worked out. The process was based on the methodology of Contextual Design and included five steps, each building on the previous one, as outlined in Figure 1. This process lays the foundation for the development of the various eNanoMapper components.

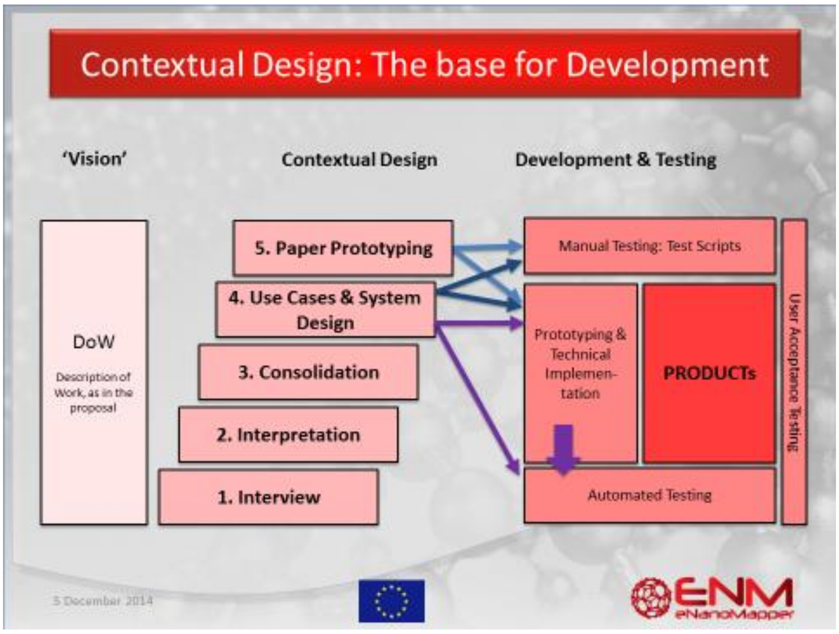

Figure 1 Contextual Design: The base of our development

During the initial interview phase, 47 interviews were conducted with scientists, manager, data analysts and IT professionals from 28 organizations. As the interviews were interpreted and consolidated, the main topics relevant for the eNanoMapper project were identified and discussed: The nano safety field is a relatively new scientific field with a lot of complexity. The main need expressed in the interviews is, to achieve a harmonization and standardization on all levels: Language (ontology), procedures (SOPs) as well as computing (APIs).

The interview were processed into notes, which were clustered in workshops into a hierarchy of topics (consolidation). Design ideas were added to the topics and used to identify the 55 main use cases, discussed in the deliverable D 1.2: "Use Cases and Test Suite".

Out of this, the high level system design was created as an overall framework of development for eNanoMapper. Recently, we started the creation of paper prototypes. This last phase is ongoing.

1.1 About eNanoMapper

The eNanoMapper project aims to provide standards as well as technologies for the computational infrastructure of the European NanoSafety Cluster. This includes the creation of an ontology, the definition of standards for API, templates and modeling, as well as the set-up of web platforms for storing data and running algorithms. With this, the project will support the collaborative safety assessment for ENMs by creating a modular, extensible infrastructure for transparent data sharing, data analysis, and the creation of computational toxicology models for ENMs.

Based on recent development work of consortium partners, eNanoMapper is building a flexible computational infrastructure, based on interoperable, standards-compliant and modular web services that maximise cross-talk and interaction between different databases. This includes key services for ontologies, data storage, data analysis and modelling as well as supporting services (e.g. for authentication and authorisation) and prototype user interfaces (GUIs and APIs) for data submission, data retrieval, computation, and analysis.

2. INTRODUCTION

Work package 1 establishes various mechanisms for collaboration with and feedback on our project with the European NanoSafety cluster, as well as the various nano-safety and nano-technology communities. Like this, WP 1 ensures the alignment of the objectives of the eNanoMapper project with all relevant scientific, commercial and regulatory communities.

The following report outlines the requirements gathering process that was performed in order to design a computational infrastructure for toxicological data management of engineered nanomaterials (ENMs) based on open standards, ontologies and an interoperable design to enable a more effective, integrated approach to European research in nanotechnology. This included collecting data from interviews conducted across the entire Nanotechnology community. These data was carefully organized, analyzed and used to define how the eNanoMapper project will operate conceptually.

2.1 SCIENTIFIC IMMATURITY

Nano safety is a relatively new scientific field. It inherits all the main challenges known from chemical toxicology, and adds new challenges, related to the physical properties of nanoparticles and their characterization and identity.

In most of the interviews, as well as throughout the analysis and design phase, topics around the ‘scientific immaturity’ of the nano safety field were central topics discussed, including their impact on computing standards and computing infrastructure, as envisioned to be developed by eNanoMapper. Figure 2 shows the outcome of an initial brainstorming at the Milan meeting held in June 2014.

Main challenges identified were:

- The specific **characterization and substance identity** is often not known exactly and / or very difficult to identify. In Nanoscience, there are no equivalents for the systems of molecular formulas in chemistry, like SMILES or the mol representation. ENM are characterized by size and weight distributions, light scattering or simply by the protocol to generate it. In most of the cases it is very difficult or impossible with today’s methodologies to determine, if two nanomaterials are exactly identical. Hence, exact reproducibility is not possible.
- The complexity in characterization leads to the **complexity in modelling** and a large gap between experiments and the models. In many areas, models are still missing or not verified. The common models, eg in pathway analysis, have limited applicability to nanomaterials, and predictions are impossible or very unreliable.
- The Scientific immaturity, the problems of characterization as well as the complexity in

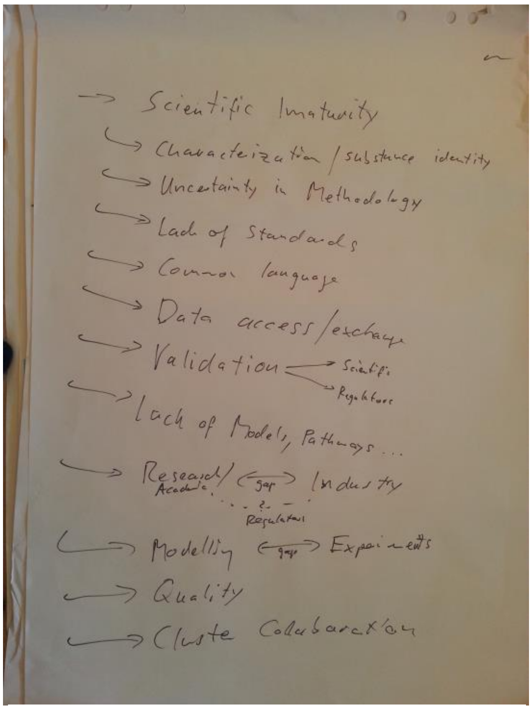

Figure 2: The main project challenge: Scientific Immatuirity. Design Workshop 18-21.6.2014 Milano &Ispra

models lead to **uncertainty in methodologies, a lack in standards and in common language**.

- The **challenges for regulators** are the complexities and uncertainties of **validations**. This results in risks associated with using nanomaterials and/or a high complexity in regulatory approval.

One of the main challenges of eNanoMapper is, to narrow this **gap between science, regulation and industry** in the NanoSafety field.

3. CONTEXTUAL DESIGN PROCESS

Prior to designing and building eNanoMapper, the consortium initiated an extensive requirements gathering exercise. To build an effective solution for scientists across the NanoSafety Cluster, it was essential to understand what work is being performed throughout the cluster and how it is performed. A methodology referred to as contextual inquiry/design can assist in the collection of information to use in developing the system requirements for the eNanoMapper project. Specifically the Focused Rapid Contextual Design approach as detailed in (Karen Holtzblatt, 2005) was chosen in this project.

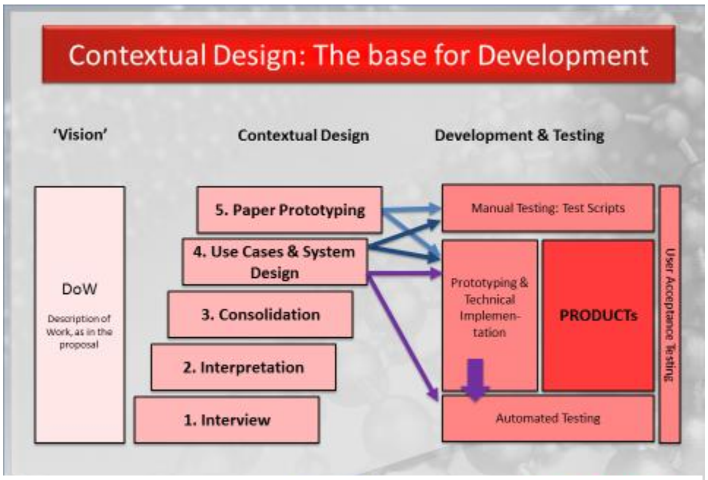

Figure 3 Contextual Design as the base of our development.

The Process can be divided into 5 stages, each of which is based on the previous step:

Step 1: Interview

Step 2: Interpretation

Step 3: Consolidation

Step 4: System design

Step 5: Paper prototyping

4. INTERVIEWS

4.1 47 INTERVIEWS ACROSS THE NANOSAFETY CLUSTER

From March to June 2014, the WP1 partners conducted 47 Interviews with scientists and managers across all main projects of the NanoSafety cluster and beyond. The interviewees were from 28 European, four US and one Indian organizations. Some of the interviews took place at the NanoTox 2014 conference in Antalya, others during on-site visits. In this process, we conducted 11 visits to organizations and their laboratories.

4.2 INTERVIEW METHODOLOGY

The main objective of the Requirements Analysis Interviews was, to find unbiased information about the interviewee, his/her organization, work profile and activities. A special focus was put on finding all information related to how nano-safety data and information is acquired, used, stored and distributed.

The interviews were aimed at understanding the present systems and processes for mapping nanomaterials to data and information. Answers were sought on the deficiencies in the present system and the expectations of the scientists from new systems and tools like the ones planned for eNanoMapper.

Detailed notes were taken along with the names of internal software systems in use, branded software systems in use, most popular sources of information including print media, publication sources, targeted search engines and any new and innovative platforms that have been used. This information was collected to ensure that any new system that is designed, meets the needs of scientists across the entire community and at the same time that it also fits within current workflows. The interviews, along with other requirements gathering exercises, resulted in over 500 separate notes. (See appendix 12)

The interviews were usually done in 4 parts:

1) **Context / Background:** Introduction about the interviewee and his/her organization. Focus areas of scientific and commercial work. Participation in projects and role within the European NanoSafety Cluster.
2) **Handling of NanoSafety Data**: Objective of the second part was, to understand the full process of NanoSafety data: Where does the organization receive and/or generate the data? How is the data processed? And where is it stored or published?
3) **Typical Process**: One typical process involving NanoSafety data was identified in this third part of the interview, and the interviewee took us step-by-step through all the details of this process. Answers were sought for questions like: What is the trigger of the process? Where does the interviewee get the data from? What does the interviewee do with the data? What / how does s/he produce new data? And where and how is the data published or passed on? – If possible, and when the security policy of the interviewed organization allowed, artifacts were collected, e.g. photos, graphics, forms, SOPs, reports, data files etc.

In these first 3 interview parts we avoided any questions or discussions related to the eNanoMapper project. This was done to prevent that any intensions of the interviewer or expectations of interviewees would bias the interview outcome.

4) In a last interview part, we asked questions related to, how the interviewee sees the **objectives of the eNanoMapper project**? How could we simplify his/her work, and in which areas of work the best potential existed to be simplified or improved through eNanoMapper. Finally, we laid out the different work packages of eNanoMapper and requested the interviewee to make a statement about where they thought eNanoMapper should put most focus on.

5. INTERPRETATION

5.1 DEBRIEFINGS

Shortly after each interview, a **debriefing session** was held with the interviewers, as well as with at least two more eNanoMapper team members, which did not participate in the interview. During these sessions, the interviewer told the story of the field interview and discussed its implications. Notes were taken by the eNanoMapper team members, who did not attend the interview: One member recorded the users role and characteristics, a profile of the organization, the physical environment, and wrote a series of notes. The notes were viewed by the entire group and recorded observations, issues, breakdowns in the work, questions, insights, and design ideas. This usually resulted in between 15-50 notes per interview and we generated approximately 500 notes all-in-all (see 12 Annexes

Affinity diagram).

The second eNanoMapper note taker captured the steps for each task (*sequences*), which included: The ordered steps necessary to complete a task, the triggers (how the event was initiated), and the intent (reason for the steps). These steps were annotated with artifacts used by the scientists in the interviews.

At the end of the debriefing, key terms used were extracted for the ontology work package, and open follow up questions identified for follow-up emails, calls and virtual meetings.

6. CONSOLIDATION

6.1 NOTES CONSOLIDATION

Once most of the interviews were completed, the eNanoMapper team met in Milan, Italy 19 June 2014 to process and consolidate the notes. In this process, all the 500 notes recorded from the interviews were used to create an *affinity diagram:* a hierarchical view of all notes collected, as given in Appendix 12. To create this affinity diagram, each note was printed on a separate post-it paper. Team members of the eNanoMapper team then sorted the notes into main areas, topics and subtopics by distributing them on the walls of the meeting room. These three levels of categorizations were marked with post-its of different colors: The main areas with green post-its, the topics with purple and the subtopics with blue ones.

The affinity diagram was then reviewed and discussed by the eNanoMapper team. Design ideas were collected and attached (the yellow notes in Appendix 12) and carefully analyzed preferably for the higher-level concepts.

Following a short discussion of the key issues discussed in the interviews, further annotations were added to the affinity diagram. The numbers in brackets correspond to the numbers in the Affinity diagram Section 12.1.

6.2 IMMATURITY: "BAD SCIENCE"

There was a broad consensus in almost all interviews, that many of the past studies in the nano safety field were of low quality, difficult or impossible to reproduce, and hence scientifically not valid:

> *(371) "Risk of jeopardizing public and private funding due to bad science (irreproducible results)"*

The scientists see many causes for this: at the root is the complexity of the scientific field, as we discussed it in section 2.1, Scientific Immaturity: the missing identification of ENMs (366 ff) and the problems of characterization (373 ff):

> *(145) "The challenge is to provide systems of classification for novel nanomaterials which could be multi-component (while existing classifications tend to assume homogeneity a bit more), or have novel shapes that have not been produced before."*

To achieve reproducibility and scientific validity, better standards for ENM characterization need to be agreed on and established, and SOPs, templates etc. are to be developed on top.

> *(407) "Big problem: a lot of repetition mainly in nano-related investigations: different SOP's and conditions; lack of standards."*

This is however a challenging task due to the discussed complexity of physical, chemical and biological properties:

> *(135) "There is (still) a fundamental disconnect in the field between chemical and biological research."*

> *(409) "Small details matter. Even the type of glassware, speed of stirring, stirring bar shape can make a difference. Speed of injection, even tiny temperature changes, cleaning of the glassware too. Some particles are easier to reproduce than others. Also, the scale up from small quantities might not work."*

Also, a tool and methodology to analyze literature better and faster would be of great help, as literature research takes up a significant time for the scientists in the field (284 ff).

> *(296) "screening and evaluation of literature is most difficult."*

(100) "Databases in the literatures aren't harmonized, so it's difficult to fit it in a harmonized template, they aren't related, so we end up with a very poor database."

(301) "If eNanoMapper could provide better tools for literature selection would be positive: reduction from 150 articles to 10-20 most relevant would be positive."

6.1 "SCIENTISTS DON'T SHARE"

Another key issue discussed in most of the interviews is, that nano data is simply not shared hence not available (183 ff):

(140) "Individual ambitions and viewpoints prevent raw and meta data publication; data sharing is a rare event."

(426) "…Asking nano researchers to fully share their protocols is a bit like asking Granny to share her special recipe for her best cake."

The only way today for a scientist to get data from the outside is, through his network of friends (248 ff, data liberation discussion):

(249) "… the only working way to find information is having good friends."

One scientist concluded for the eNanoMapper team (outside an interview): "There are two important things in the life of a scientist: funding and publications. If you want to make scientist share data, you have to bind it to one of them – or best to both!"

(320) "If pushed by funding or access to publishing, researchers would definitely change!"

The project set up and the set goal is critical too for data sharing:

(258) "Ultimate goal of a project is of utmost importance: (is it) data or knowledge? If it is knowledge, data are not shared."

(255) "… in the engineering world big consortia have formed to generate and share data; not for free/public, but accessible to consortium partners"

The awareness of the importance of data sharing is however growing quickly: More journals demand, that all raw data, protocols and processed data must be publicly shared upon publication. Funding agencies such as the EU Horizon 2020 program are starting to require open data sharing as a part of their project requirements.

However, data sharing is of limited use, if it is difficult to access either because it can't be found, and/or because it is saved in propriety formats.

Hence, a key objective of the eNanoMapper team must be, to establish standards, as well as data warehouses, where nano safety data can be stored, found and re-used.

6.2 HARMONIZATION NEED

6.2.1 THE LANGUAGE: ONTOLOGIES

The problem of standardization however starts at a higher level of the language used: nano scientists often use different terms for the same thing.

(175) "If you go into the literature, you will find, that everyone calls things differently (ontology!) Example: we spent the first part of our (graphine) project just to define, what graphine is (physical chemical properties)."

In our interviews, we found a few teams using a glossary, one using a folksonomy (72, 173), but these are not shared between teams, and no one using ontologies in regular work (165 ff).

(169) "They do not use ontologies, but have a glossary, as well as defined key words for SEO (search engine optimization)."

A key task for the eNanoMapper must be, to harmonize the language used, according to interviewees. Within WP2, eNanoMapper has a team working on a common ontology for the European NanoSafety cluster and for the field of nano safety more generally.

(171) "(in a skeptical tone) For eNanoMapper do you foresee developments and creation of software in these areas of meta-data? You would need the correct ontology!"

(172) "Data from other projects? > Ontologies? - to understand / determine, how data will show up in their DB."

6.2.1 PROCESSES: SOP AND PROTOCOL

A harmonized language however is only the starting point. The core for scientific reproducibility and hence for scientific work in general is, to have a set of common protocols (399 ff) and SOP (Standard Operating Procedures) and protocols. The lack of it was a key issue in most of the interviews (456 ff):

(458) "Reproducibility of experiments is very difficult; lack of controlled SOP's; e.g.: specifications of media and suspension of nanomaterials."

(457) "Challenge is to identify robust and reproducible procedures on all levels: data generation, processing and conclusions."

This lack of reliable protocols and SOPs goes to such an extent, that a few interviewees came to the conclusion, that most of the current data in the nano safety field is not valid, as not reproducible, and hence, before generating data, there need to be SOPs and protocols:

(445) "Data is not (yet) important to us - All our focus is on methods: SOP? Protocols? etc."

We saw above (409), how difficult it is to create reliable protocols and SOPs. The NANoREG project is dedicated to agree on robust SOPs

(460) "NanoReg: „testing the tests": Finding, developing, creating the building blocks for regulators and policy makers."

eNanoMapper will deliver a service to store and retrieve SOPs, protocols, templates and other documents, see 8.2.2 "The services".

6.2.2 INTERFACES: API

What an ontology is for the scientific language, an API (Application Programming Interface) is for computer interfaces: defining the way, how computers interact. Standard APIs are essential to use data and algorithms from different databases (103 ff & 155 ff).

(157) "View eNM: 2) Make sure, the DBs around the world use a common / the same language …"

(158) "wishful from eNM: common standards of knowledge sharing and retrieval; …"

The definition of standard APIs is an essential part of the eNanoMapper project: We base the definitions of our APIs on the open standard OpenTox – see Section 8, "System Design" and specially 8.2.1, "The REST API".

6.2.3 DATA

But even if data is available, but stored in a proprietary way, it is very difficult for scientist to interpret and compare. Data needs agreed rules of the overall structure and design, and associated meta data information, in order to be easy to use in science.

The most common standard today is ISA-TAB nano. Even though approximately 70% of the interviewees knew about ISA-TAB, and approx. 30% plan or have decided to use it in future, the adaption today is still very low (68 ff):

> *(69) "Not using ISA-TAB (not yet), but may use it in NanoDefine;"*

But there are a lot of concerns about the usability of ISA-TAB:

> *(71) "For the modelers ISATAB is great, but for the scientists??? Too time-consuming,"*

eNanoMapper has decided to make ISA-TAB nano the standard format for the project and we work on the standard on multiple levels:

- Creating templates in ISA-TAB
- Storing our data in ISA-TAB, wherever feasible, and serving it via interfaces (see 8.2.2)
- Creating tools and services for ISA-TAB creation and handling

6.3 ENANOMAPPER - A DATABASE OR A STANDARD?

Should the eNanoMapper project develop a data warehouse for the cluster? Or should the project rather focus on clarifying standards, supporting harmonization of the existing databases to render data interchangeable and into a common search index, so it can be easily found?

The reality today seems to be, that there is already a lot of data shared on the NanoHub database, but in all interviews there was not a single interviewee (!), who actually got data out of the NanoHub. (232 ff):

> *(233) "… he does not think that anyone who has had access to Nanohub, has actually extracted data from there."*

A reason for this might be, that NanoHub was developed for the regulators, not the scientists primarily:

> *(64) "Iuclid (db software behind nanohub) has been developed especially for regulators, not very research friendly; it is including administration issues, no semantic interrelationships for scientific questions; …"*

There was no clear conclusion by the interviewees, if eNanoMapper should improve existing, or create a new data warehouse:

> *(112) "Wishful: eNanoMapper DB and software tools should be easy-to-use, intuitive and as close as possible to current formats; Building acceptance in scientific/experimental community; translation services?"*

> *(265) "eNanoMapper should not create another data base: There is enough out there, just make it better to work with: i) make it compatible; ii) create tools (QSAR, Data entry) iii) create a meta search."*

A slight majority of interviewees however concluded, that it will probably be easier to create a new database, instead of improving the current ones. And a clear majority felt, that there needs to be at least a new reference implementation for a modern data warehouse (111 ff).

> *(119) "eNanoMapper should organize and elevate the raw data, the protocols and metadata."*

(122) "… I would like to have a tool, a system, capable to dialog with eNanoMapper databases, with other tools produced, available, so that we can strengthen the input data. …"

6.4 SEARCH

Search was clearly a main topic in most of interviews (261 ff, expecially meta search, 273ff):

(276) "… for eNM main focus should be to build a powerful search opportunity."

(277) "Meta search wishful thinking: i) Nanomaterial ii) species; both interactive and hierarchical; iii) end points (mortality, reproduction; oxidative stress)."

(303) "If you can't find it, you repeat experiments that have already been, done. Problematic search of of past experiments."

Even as eNanoMapper creates its own data warehouse, there will always be multiple data stores, and a common meta-search possibility will be an increasing need for the community: not just according to keywords (like google), but according to multiple dimensions, like various characterization properties, species tested, organ, cell lines, genes proteins etc.

7. USE CASES

7.1 LIST AND SPECIFICATION OF USE CASES

With these design ideas (the yellow notes in 12) as well as the major conclusions discussed in the last section, the team identified in Milano 55 use cases for eNanoMapper to specify and analyze in more detail.

A detailed list and description of all these use cases is part of the eNanoMapper Report "Use Cases and Test Suite" (Deliverable 1.2).

Priorities for the use cases included suggestions that addressed common issues across the cluster or those that resulted in the elimination of steps for different tasks and improving the productivity of the task.

Use cases

1. Write protocol
2. Design protocol
3. Rank paper
4. Upload protocol
5. Upload dataset
6. **Design study**
7. Search protocol
8. Search dataset
9. Download protocol
10. Download dataset
11. **Extract data from paper**
12. **Prepare data for modeling**
13. ISA-TAB creation
14. Build QSAR
15. Validate ISA-TAB
16. Validate protocol
17. Validate model
18. Find ISA-TAB template
19. Create template ISA-TAB
20. Create template (xls etc)
21. Map asset to pathway
22. Search for perturbed gene
23. **Search for all knowledge of nano material X**
24. Search for all knowledge of impurities
25. Calculate precaution for X
26. Copy open facts to generate it
27. Find/ link related data sources
28. Find data with similar patterns
29. Save my data in format X
30. Publish my data with my paper
31. Find people with similar X
32. Who else with protocol X
33. What URI to use for X
34. Give me all names for ontology term A
35. Create NSL dictionary
36. Harmonize Terminology
37. Win Nobel prize / Automatic Knowledge discovery tool
38. Find ontological contradiction
39. Comp aided ENM design
40. **Define safe-by design**
41. **Integrate data for risk assessment**
42. **Verify data against regular requirements**
43. Convert from one template to another, e.g from Modclust to OECD
44. Find producer of eNM C
45. Find QC data of eNM C
46. User alert for new info on D (gene, nano tag)
47. Find nano particles used in product
48. Format data to be used by C
49. **Map nano material found in DB 'A' to an entry in DB 'B'**
50. Find all DB's with data of Paper X
51. Write up tutorials on use cases
52. Register and get access (public)
53. Map existing used schemas to ontology IDs (e.g. OECD harmonized templates)
54. **Annotation of data with ontology IDs (e.g. experimental data in ISA-Tab)**
55. Experimental design

Figure 4: Use Cases

7.2 EXAMPLE USE CASE 23: SEARCH NANO-MATERIAL X

We designed a template for describing the use cases along the "casual" template of Alistair Cockburn (Cockburn, 2001).

As a representative example, following the details of use case 23: "Search for all knowledge of nano material X" -

Primary Actor: Researcher: (a) Industry (b) Academic or (c) Regulator.

Scope: Find all general, as well as specific information available on one particular nano-material – according to its name and/or detailed characteristics.

Brief: Classes of nanomaterials have unique names - specific materials may however not have a name, but are characterized through their many characterization properties. A search for nanomaterials can not be based only on keywords, but has to include specifics like size distribution, dissolvability etc. Further, suggestions and a guided search process may be very useful - results should be prioritized.

Trigger: May be …

A. An industrial researcher considers using a nanoparticle for a product. He investigates the impact on humans and the environment (safety study)

B. An academic researcher is integrating a nano material characterization into a cell membrane model.

C. A regulator assesses the safety of a nano material, which company X wants to use in a product.

Basic Flow:

A) Industrial Researcher

1. The researcher goes to google and enters the name, as well as a few characteristics of the ENM.
2. Besides other resources, the eNanoMapper landing page shows in the results. He follows this link.
3. He has two choices: (a) browse services and resources (b) search: basic, as well as advanced. Advanced search gives him a form for entering characterization details.
4. He finds all data and information provided or linked by eNanoMapper either by browsing or in the search results.

B) Academic Researcher

(Discuss with users)

C) Regulator

Similar to Case A), but with emphasis on information to satisfy regulatory review requirements such as “fit for purpose” or validation.

8. SYSTEM DESIGN

This section outlines the high level of the overall system design of eNanoMapper: The basic design patterns and architectural choices of the systems implemented in the eNanoMapper project. Details for each component can be found in the deliverables of the other work packages.

8.1.1 INTRODUCTION

The eNanoMapper system consists of a set of web services, providing access to protocols and data, search services, libraries as well as GUIs (Graphical User Interfaces), offering user-friendly access to the above functionality. The web services, currently developed by partners could run on the same machine, or on geographically dispersed servers, and communicate via the Internet. The design is expected to facilitate adding new services of any kind, for example supporting different data types or search functionality.

eNanoMapper currently adopts the OpenTox framework design (http://opentox.org), based on the following technological choices: (i) The REpresentational State Transfer (REST) software architecture style allowing platform and programming language independence and facilitating the implementation of new data and processing components. (ii) A formally defined common information model, and communication through well-defined interfaces ensuring interoperability of the web components. (iii) Authentication and authorisation, allowing defining access policies of REST resources, based on OpenAM (see the eNanoMapper Report “User registration, authentication and authorisation”, Deliverable 5.2).

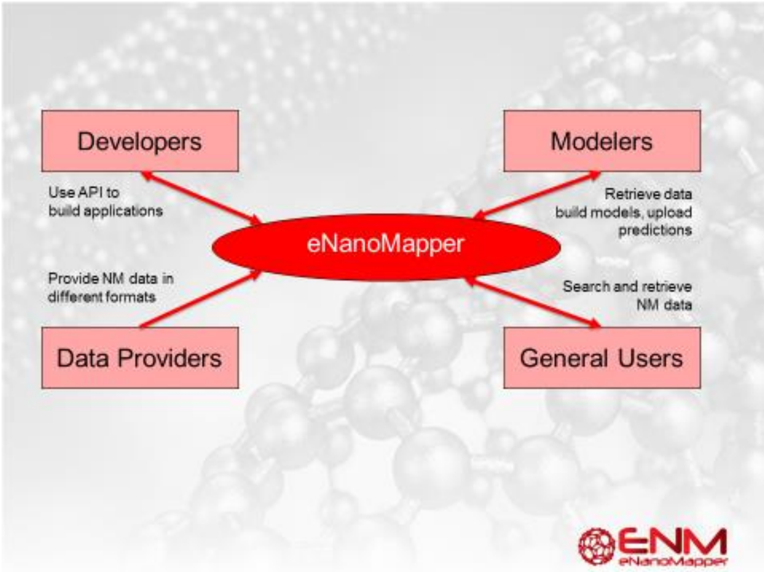

Figure 5: System Design

8.2 SYSTEM ARCHITECTURE CONTEXT DIAGRAM

The system architechture context diagram represents the overview of the external interaction with the system, the components, how they interact with other components and the different services.

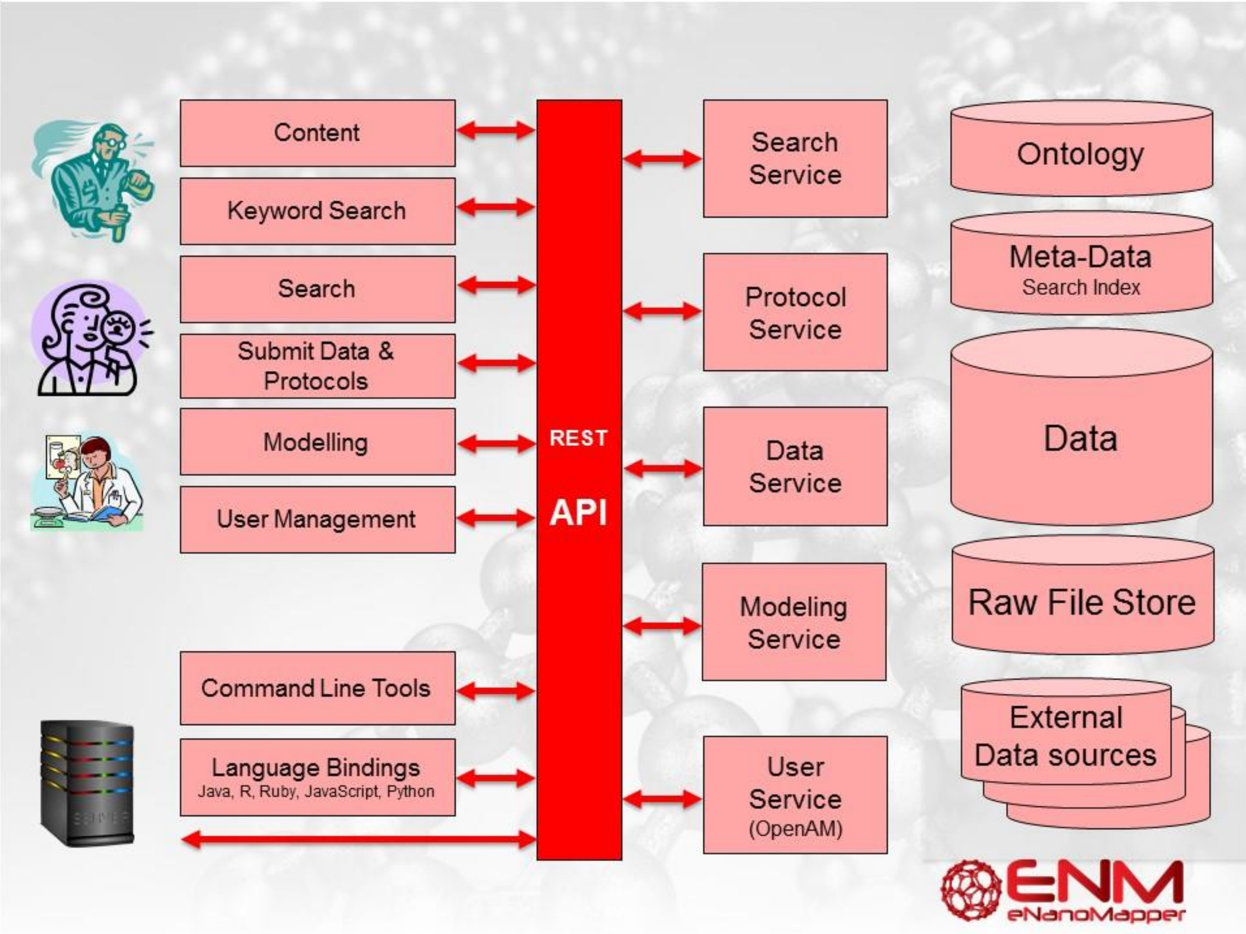

Figure 6: System Architecture Context Diagram

The core of the eNanoMapper system consists of the 4 services for search, protocol, data and modelling. These services load and store into various internal as well as external data sources, supporting ontologies, meta data like search indexes, the various data standards of the nano safety field, as well as raw files.

The connecting backbone of the system is the REST API, which facilitates all interactions within the components, as well as with outside systems. The API connects to various UI systems, as well as to external applications – providing the various uses functionalities matching with their profiles.

8.2.1 THE REST API

The REST API is the central backbone of the eNanoMapper systems. All internal components of eNanoMapper are accessible via the API and all interactions between the components, as well as outside systems use the API.

The API also connects the various services with the various user applications for data scientists, chemists, biologists, researchers, software developpers etc. These user applications are based on the main use cases discussed in the previous chapter and in D 1.2.
The API is based on the OpenTox API standards (http://opentox.org/dev/apis), extended with support for substances and documents. The API supports JSON for facilitating UI (User Interface) development and an RDF-serialized output of data, using the ontology.
The enanoMapper resources will be identified by dereferenceable URIs to ensure that all primary entities in the eNanoMapper storage can be identified and looked up. This ensures that any eNanoMapper instance can be a component in the Linked Data Cloud.
Implementation details, as the specific technologies used for storage and processing, are discussed in the relevant deliverables of the other work packages.

8.2.2 THE SERVICES

The core functionality of the eNanoMapper system is provided by its five services:

- The **Protocol Service** stores and serves standard protocols, templates, SOPs, reports and various other documents. A detailed discussion of functionality and API can be found in Deliverable D3.1.
- The **Search Services** establish and maintain varous indexes along multiple criteria for finding relevant data and protocols: within the eNanoMapper data warehouse, as well as in supported outside databases.
- The eNanoMapper **Data Service** is the central part of the eNanoMapper data warehouse. Functionality and REST API are defined in D3.1. The input and output to and from the storage will support various formats, based on community needs. The input will be a set of importers, the output will support a set of formats, facilitating the consumption of the web services in different scenarios:
    - RDF (for integration with Linked Data Cloud),
    - JSON (for UI development),
    - ISA-TAB-Nano (for supporting standardized format for describing experiments).
- **Modelling services** run modeling algorithms in a standardized and reproducable environment.
- **User Services** adding **Authorization and Authentication** run on top of all services, as described in D 5.2.

8.2.3 DATA STORE

Various stores are supported by the eNanoMapper system for the various types of meta data, processed and raw data and raw files:

- The **Ontology** module captures the domain languages. As such, the module is an integration of smaller more specific ontologies, including those listed below. However, the outcome is a single OWL ontology that people can import into their ontology, covering all the subdomains. For all included ontologies, a slimmed down version is used, including all required bits, but solving typical ontology integration issues. See eNanoMapper Report “Ontology Initial Release” (Delivarable Report 2.3)
- Further **meta data** is stored, mainly facilitating the search services, but also other services, like the user service (see eNanoMapper Report “Technical Specification and initial implementation of the protocol and data management web services”, Deliverable 3.1).
- The main **data** is stored in relational structure, allowing fast and easy SQL access to it.

- A **Raw files store** supports the storage and exchange of large raw data from omics, characterization etc.

9. PAPER PROTOTYPING

After the use cases were specified and a generic system design has been worked out, we started with "paper prototypes": High level outlines and specification of implementation ideas. They were first mocked-up as hand drawings, then worked out on slides. "Paper prototypes" are the first specifications for individual systems. They are from an end-users perspective and still hide a lot of implementation details in order to keep the focus on the functionality and the overall design and usability. This effort is ongoing with documentation maintained in the eNanoMapper document repository.

9.1 SCOPE OF PAPER PROTOTYPES

A typical "paper prototype" covers three perspectives:

1) The **functionality** covered – Usually this is done by first specifying, which use cases are covered, then specifying a few high level persona- and flow diagrams outlining the main functions of the use cases.
2) The functional **screens flow and customer front-ends.** Emphasis in this step is on working out functional elements and keep it as simple as possible. Hence, all main functionality is shown is screen mock-ups, but no graphical design or UX details are included. These screens are later shown to customers in feedback sessions.
3) The **architecture** of this system and raw implementation idea.

A paper prototype typically starts with one use case. Because the scope of the prototype is not a single use case but rather one technical system, functionality is then extended to cover more use cases in parts or in full, which logically fit into the same technical system.

9.2 REFINING AND TESTING THE PROTOTYPES

The paper prototypes are currently further tested with NanoSafety Cluster scientists in multiple rounds of feedback. In these user interviews, the paper prototypes are "used" to actually perform a series of tasks: the screen flows and customer front end GUI components on paper or slides are presented in response to the actions taken by the user. Changes are made on-the-fly to ensure the GUI performs the tasks the users expect. These changes are then consolidated with the eNanoMapper team to form the blueprint for the eNanoMapper applications.

9.3 EXAMPLE: PAPER PROTOTYPE FOR USE CASE 23

As an example for a paper prototype, following a discussion about an excerpt of use case 23: Search for all knowledge of nano material X

9.3.1 USE CASES COVERED BY THIS PROTOTYPE

- 8 - Search dataset
- **23 - Search for all knowledge of nanomaterial X**
- 44 - Find producer of eNM C
- 45 - Find QC data of eNM C

9.3.2 USE CASES, PARTLY COVERED BY THIS PROTOTYPE:

- 4 - Upload protocol
- 5 - Upload dataset
- 7 - Search protocol
- 12 - Prepare data for modeling
- 13 - ISA-TAB creation
- 22 - Search for perturbed gene
- 24 - Search for all knowledge of impurities
- 28 - Find data with similar patterns
- 31 - Find people with similar X
- 34 - Give me all names for ontology term A
- 36 - Harmonize Terminology
- 42 - Verify data against regular requirements
- 46 - User alert for new info on D (gene, nano tag)
- 47 - Find nano particles used in product

9.3.3 PERSONA / FLOW-DIAGRAM

Via google, another resource or directly, a user is reaching the landing page of eNanoMapper.net, which offers him three options:

- He can **search via keyword** or advanced search interface
- He can **browse the structure** to explore the resources of eNanoMapper.net
- He can **add his own resources** into the eNanoMapper data warehouse. A simple form-driven dialogue is provided for this.

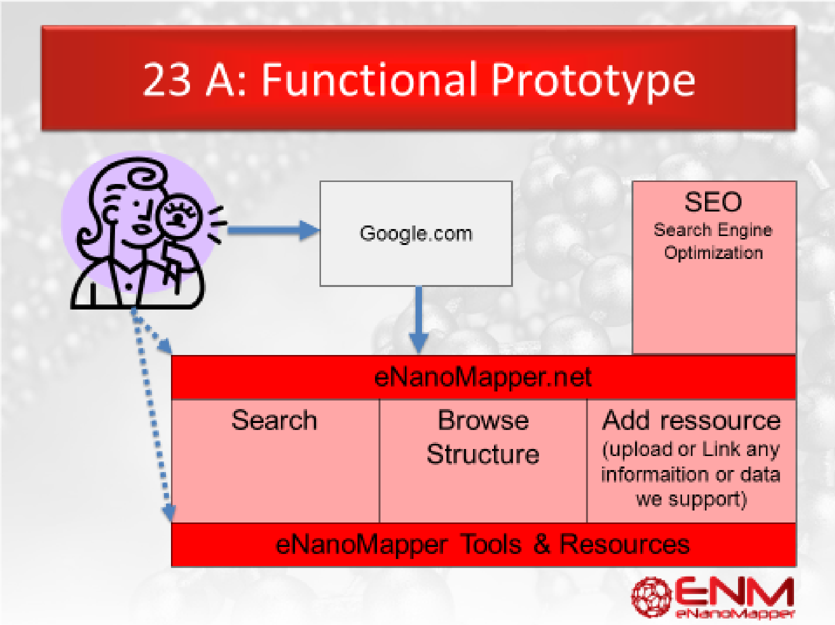

Figure 7: Persona / Flow-Diagram

For searches, the key question of this use case is: along which categories do users perform their searches? In the interviews, we got the following main areas, which all have to be part of the search indexing for keyword search, as well as part of the advanced search form:

- ENM: Characterizations
- Endpoints
- Biology: species, organs, cells
- Micro-biology: genes, proteins, pathways, biomarkers etc
- Modelling: Models, QSARs, algorithms, model components

System 23 is the key connector between the various user groups and the different tools and resources of eNanoMapper.

9.3.4 SCREEN FLOW

A series of screen flows are currently worked out by the team to represent the functionalities of the previous section – following the first screen as an illustration: A user comes to eNanoMapper.net and has first 2 possibility to access the tools:

- The top right search box connects a user to data, resources and tools via the search index.
- Clicking on the "Library" link gives the user an overview of resources, as well as access to advanced search interfaces, as well as possibilities to add his own data and resources.

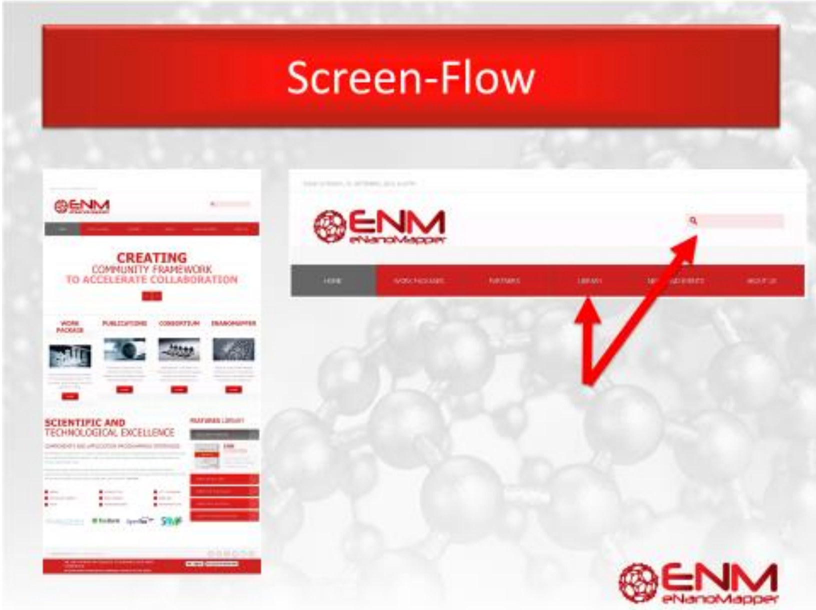

Figure 8: Screen flow

9.4 TECHNICAL ARCHITECTURE

eNanoMapper.net must have four main components to act as the connector between the users and the tools and resources:

- A **Content Management** system, outlining and describing the areas of eNanoMapper
- A **SEO module** (Search Engine Optimization) to make sure, eNanoMapper.net and its resources appears in all major search engines.
- An own **Search Index**, mapping requests to all internal, as well as

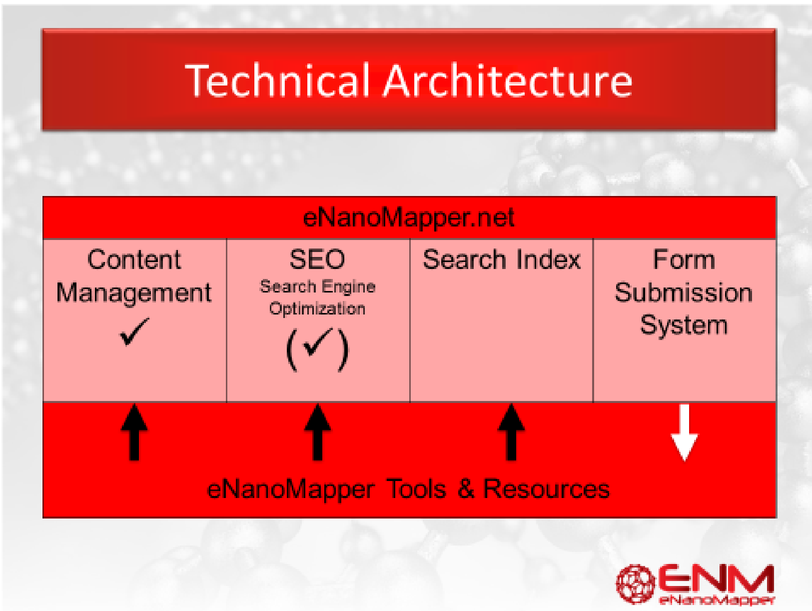

Figure 9: Technical Architecture

supported external resources.

- A **Form Submission System** to provide a simple dialogue to add resources, like in vitro or in vivo data, reports, SOPs, QSARs etc.

10.CONCLUSION

The eNanoMapper project started one year ago with the objective to create computational standards and infrastructure for nanomaterials design and safety assessment for the European NanoSafety cluster, as well as for the nano science community in general.

We started in work package 1 with a requirement analysis, preforming a series of interviews throughout the relevant nano safety communities. We then interpreted and consolidated the interviews in multiple sessions and identified the main topics for us to work on.

The main challenge of the eNanoMapper project lies in creating and establishing standards in a young scientific field: The immaturity of the nano safety field has two main roots:

- The complexity of toxicology is inherited from chemistry and biology.
- The complexity of characterization of ENMs, leading to difficulties of ENM identification and hence reproducibility. This causes uncertainty of the validity of experiments and data.

This makes it today very difficult for scientists, regulators and the industry to share and compare data, models, protocol, SOPs and knowledge in general.

Hence, for the remaining two years, it must be the main goal of the eNanoMapper project to increase the “sharing” in the community though a harmonization on various levels:

1. To establish a harmonized **language** with the help of a standardized ontology.
   - ✓ We already created in WP 2 an initial release (see eNanoMapper D2.3: “Ontology Initial Release”)
2. A harmonization of **processes** by establishing an infrastructure for harmonized SOPs, protocols, templates etc.
   - ✓ Work package 3 will establish the possibility to store and find standard documents. We are also adopting the current industry standards, like ISA-TAB, the OECD harmonized templates etc and are building tools on top of these standards (see eNanoMapper D3.1: “Technical Specification and initial implementation of the protocol and data management web services”)
3. Harmonized **data sets** by setting up a standardized data warehouse on one side, and by defining standard APIs for the distributed resources of the community.
   - ✓ We already did in work package 3 an initial release of a data warehouse, incorporating flexible import facilities for the different templates of the nano safety community, as well as for the OECD harmonized templates (see D3.1 above). Also, the high level API guidelines are defined (see 8.2.1 The REST API) and in work package 3 and 4, a lot of work for its implementation is already done.
4. Harmonized and reproducible computational toxicology **models**, by establishing modelling and API standards.
   - ✓ We already did in work package 3 an initial release of a data warehouse, incorporating flexible import facilities for the different templates of the nano safety community, as well as for the OECD harmonized templates (see D3.1 above). Also, the high level API guidelines are defined (see 8.2.1 The REST API) and in work package 3 and 4, a lot of work for its implementation is already done. (see eNanoMapper D4.1: “Analyzis and Modelling Specifications”)

With the requirement analysis and its processing to use cases, the system design, the prototypes etc, we have (tried to) make sure we do the right things for the community. For a successful eNanoMapper project it will be however key, to also deliver the standards, infrastructure and tools in the right way to the community. This is mainly the work of WP 5 and WP 6:

- In WP 5 “User Application Development, Integration and Testing” we will ensure, the top priority use cases identified in WP1 will be covered by user friendly, easy to use and reliable applications.
- In WP 6 “Dissemination and Training” we will make sure, the tools are easy to use, associated with tutorials and documentation, trainings and (virtual) seminars are available, so that the standards and tools are well adopted by the community.

For the rest of the project, it will be crucial to move more and more focus from WP1 to WP 5 and 6.

11.BIBLIOGRAPHY

Cockburn, A. (2001). *Agile Software Development.*

Hugh Beyer, K. H. (1998). *Contextual Design - Defining Customer-Centered Systems.* Morgan Kaufmann Publishers.

Karen Holtzblatt, J. B. (2005). *Rapid Contextual Design: A How-To Guide to Key Techniques for User-Centered Design.* Elsevier.

www.eNanoMapper.net

12. ANNEXES

12.1 AFFINITY DIAGRAM

The Affinity Diagram is the product of the eNanoMapper Milan meeting, June 19 2014. This matrix represents the preliminary Requirements that were developed and the structural design that began to evolve during the Design Meeting.

The process of generating this affinity diagram is discussed in detail in 6.1 Notes consolidation.

Color coding:

White	Note: The original notes, captured from the interviews
Green	Level 1 Category: The main areas.
Purple	Level 2 Category: The topics
Blue	Level 3 Category: The sub-topics
Yellow	Design Idea

FOR THE DETAILS OF THE AFFINITY DIAGRAM, PLEASE CONTACT THE ENANOMAPPER TEAM.