

# The EU NanoSafety Cluster as Linked Data visualized with Scholia

Egon Willighagen<sup>1</sup>, Najko Jahn<sup>2</sup>, Finn Årup Nielsen<sup>3</sup>

1. Department of Bioinformatics - BiGCaT, NUTRIM, Maastricht University, NL
2. State and University Library Göttingen, University of Göttingen, DE
3. Cognitive Systems, DTU Compute, Technical University of Denmark, DK

DOI: [10.6084/m9.figshare.6727931](https://doi.org/10.6084/m9.figshare.6727931)

Copyright © 2018 The Authors. Creative Commons BY 4.0 Int.

## Abstract

At a recent hackathon organized by the European Research Council, GeneWiki, and others, a group of 25 researchers came together in Berlin to work on ontologically modelling research grants in Wikidata. During this meeting the EU NanoSafety Cluster was used as use case study, resulting in new linked data around the cluster. The Scholia platform was extended with a Project aspect, more than 40 projects have been added to Wikidata, and almost 500 journal articles associated with their project. The result can be viewed at [tools.wmflabs.org/scholia/project/Q27949537](https://tools.wmflabs.org/scholia/project/Q27949537).

## Introduction

One of the goals of the EU NanoSafety Cluster (NSC, [www.nanosafetycluster.eu](http://www.nanosafetycluster.eu)) is to maximise the synergy between the EU-funded projects that participate in the cluster. As such, sufficient interactions and exchange of knowledge between the projects are required. Of course, a common language and data exchange infrastructures are being developed [1,2], but the infrastructure is also about getting researchers together.

Research outputs, like journal articles, datasets, standard operating procedures, models, software, etc, from the various projects should be shared, but also be easy to find. FAIR and Open Data are therefore essential components of the European Commission H2020 and FP9 (Horizon Europe) research funding frameworks, generally referred to as Open Science [3]. A collection of research articles resulting from NSC projects up to 2016 was recently made available as a ScienceOpen collection [4]. Obviously, we need a powerful platform if we also want to integrate information about NSC project grants, funders, participating institutes, and research topics. For that we need a linked data platform.

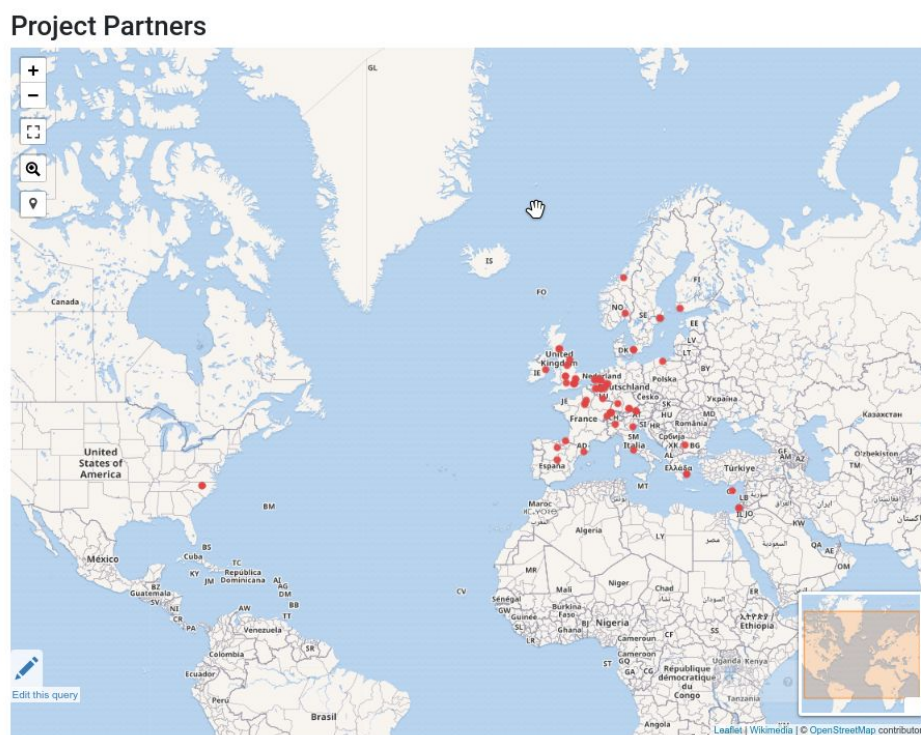
We present here some of the outcomes of a recent hackathon organized by the [European Research Council](http://www.eur-council.eu) (ERC), GeneWiki, and others [5]. The goal of this meeting was to model grant information and see how this can be used in the Wikidata platform for linked data ([www.wikidata.org](http://www.wikidata.org)) (see e.g. [6]). The NSC was proposed as a use case [7], which several participants worked on during the event, including the authors of this write up.

## Scholia, Wikidata, and project

Scholia was introduced recently as a graphical interface around Wikidata, combining information from multiple entries in the database, with a focus on scholarly literature [8]. Scholia has since

been used to provide overviews of literature, e.g. literature about the Zika virus [9]. Scholia organizes views on data from Wikidata into so-called aspects. Existing aspects include *work*, *venue*, *author*, *topic*, and *organization* among a number of others.

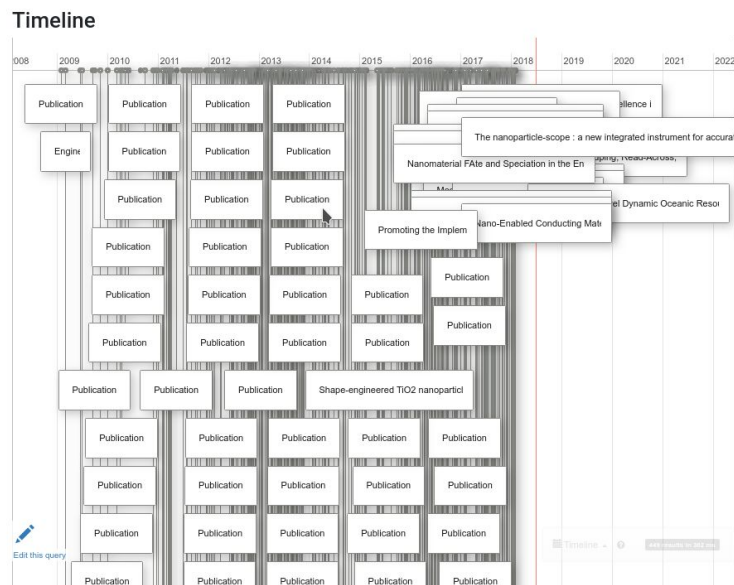
In relation to the Berlin workshop, we developed a new *Project* aspect, which visualizes various bits of information about projects, such as a map of participating institutions (see Figure 1): [tools.wmflabs.org/scholia/project/](https://tools.wmflabs.org/scholia/project/).



**Figure 1:** Map of institutes participating in NSC projects. It is currently incomplete because Wikidata currently does not list all participants for all projects, and for many projects only the coordinating institute.

### Project information in Wikidata

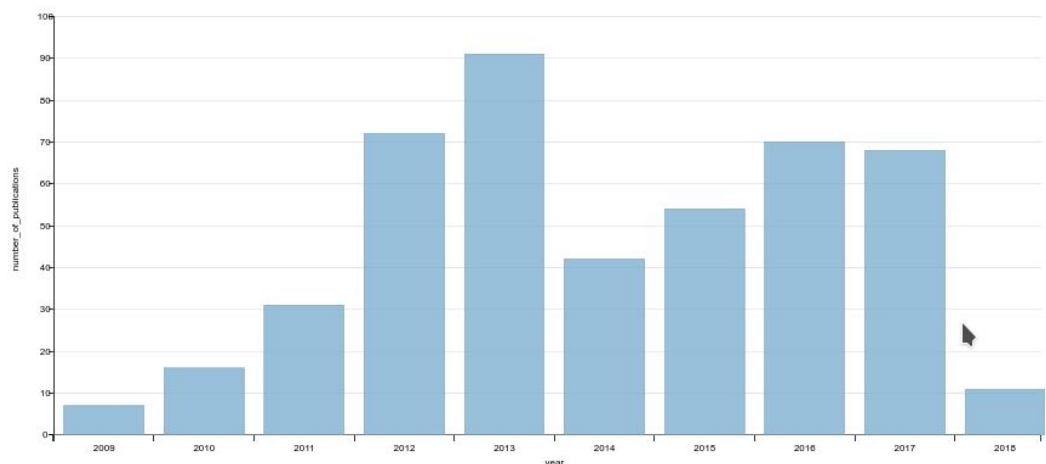
During the hackathon (see also [Wikidata:WikiProject Research projects](https://wikidata.wiki/wiki/WikiProject_Research_projects)), a number of Wikidata items were created for projects participating in the NSC, using grant information from the EC CORDIS database. A list of FP7 and H2020 projects was retrieved from the NSC website and merged with data from CORDIS, a central online databases about projects funded by the European Commission. OpenRefine's Wikidata reconciliation tool ([www.wikidata.org/wiki/Wikidata:Tools/OpenRefine](http://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine)) was used to upload data about the NSC projects to Wikidata in case they were not available, yet. Furthermore, OpenAIRE REST API ([api.openaire.eu](http://api.openaire.eu)) was used to obtain links between items for projects and items for journal articles indexed in PubMed. The relationships between grants and publications were represented according to the QuickStatement syntax (see [www.wikidata.org/wiki/Help:QuickStatement](http://www.wikidata.org/wiki/Help:QuickStatement)), a popular tool for uploading data to Wikidata. This combination allows Scholia to present a timeline of NSC projects and research outputs, as depicted in Figure 2.



**Figure 2:** Timeline of NSC projects and research output of those projects.

This new Scholia Project aspect also lists the literature in other ways, including a table listing more than 450 articles. This list is smaller than the ScienceOpen collection which was compiled and curated by an expert by searching OpenAIRE, Google Scholar, and other sources using the individual project short names and Grant Agreement numbers [4], but also covers literature from 2017 and 2018, which the ScienceOpen collection did not at the time of writing (see Figure 3).

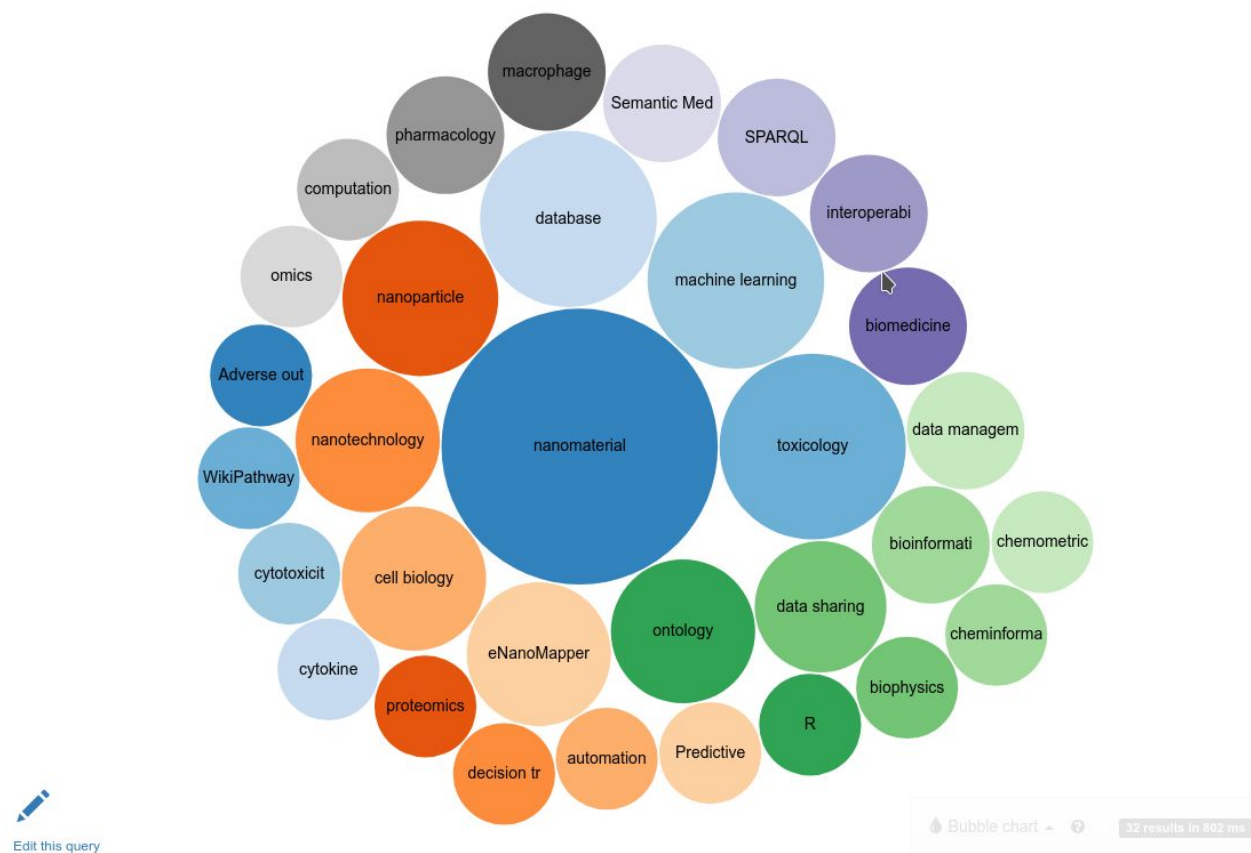
### Publications per year



**Figure 3:** The Scholia Project aspect for the NSC features a bar diagram showing the number of research outputs (mostly journal articles) over time, including recent literature.

## Topic scores

Topics based on a weighting between fields of work, topics of research output and topics of citing works.



**Figure 4:** Topics associated with articles published by the eNanoMapper project.

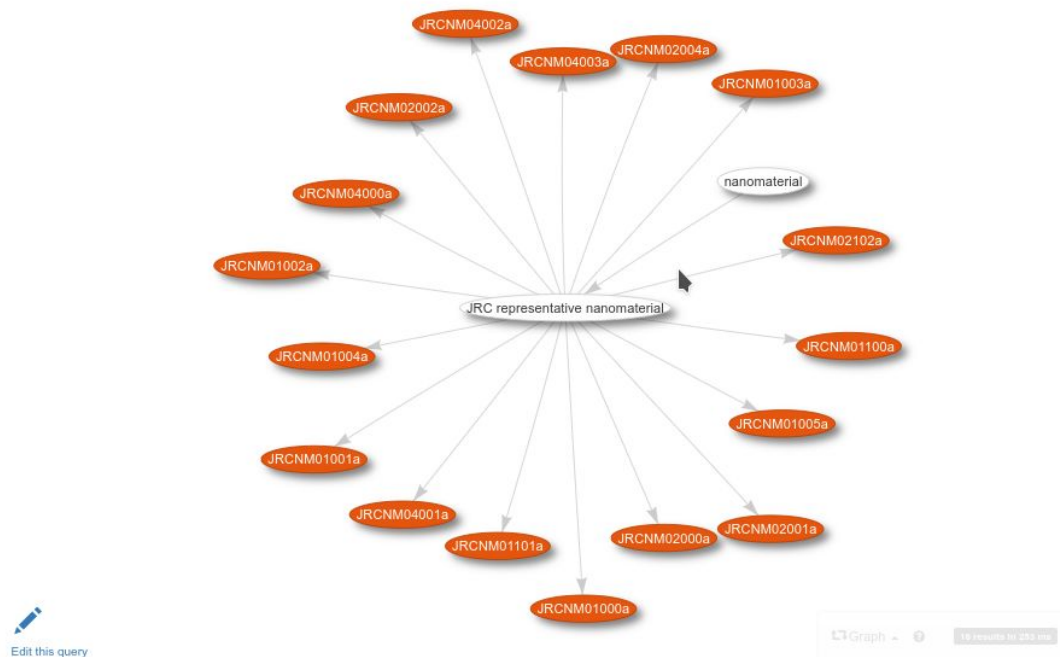
Of course, the Project aspect of Scholia can also be used for individual projects. For example, Figure 4 shows the topics related to articles funded by the eNanoMapper project. Importantly, this topics overview only works if articles are annotated with *main subject* information in Wikidata.

## Discussion and Conclusion

While Wikidata now better covers the activities and results of the NSC, the effort is not complete yet. For example, not all EC grants are available in Wikidata yet, and not all project partners are connected to projects. In this respect, other results of the Berlin hackathon concerning grants and projects data, and data modelling will be presented in other venues and soon available. However, the Scholia system will become increasingly informative as more data get linked, which anyone with [a Wikidata account](#) can do, and everyone in the NSC cluster is invited to contribute to.

Particularly, Wikidata and Scholia allow annotation of research articles with their “main subject”, as was done for Zika [9]. Similarly, it would be nice to have NSC cluster articles annotated with the nanomaterials they study. This is particularly interesting for those articles that describe JRC representative industrial nanomaterials [10], which are already in Wikidata: [tools.wmflabs.org/scholia/chemical-class/Q47461491](https://tools.wmflabs.org/scholia/chemical-class/Q47461491) (see Figure 5) [11].

## Class Hierarchy



**Figure 5:** JRC representative nanomaterials found in Wikidata.

But while the data can be more complete, it nicely shows how a public knowledge base can be used to support European nanosafety research. The power is of having information about our NSC as linked data is only limited with the data we made available. Using custom search queries we can extract all sorts of information from the system. Just to give some idea, follow [this link](#) to get a list of all publications funded by at least two NSC projects.

## Acknowledgments

This work received funding from the European Union's H2020 research infrastructure project NanoCommons under grant agreement no. 731032.

## References

1. Hastings, J. et al. [eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment](#). Journal of Biomedical Semantics 6, (2015).
2. Jeliazkova, N. et al. [The eNanoMapper database for nanomaterial safety information](#). Beilstein Journal of Nanotechnology 6, 1609–1634 (2015).
3. <https://ec.europa.eu/research/openscience/>
4. Willighagen, E., Lynch, I. & Dawson, S. A ScienceOpen collection of NanoSafety Cluster publications. (2017). NanoSafety Cluster Newsletter, doi:[10.6084/m9.figshare.4725424.v2](https://doi.org/10.6084/m9.figshare.4725424.v2)
5. [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Wikidata\\_for\\_research/Meetups/2018-06-17-19-Berlin](https://www.wikidata.org/wiki/Wikidata:WikiProject_Wikidata_for_research/Meetups/2018-06-17-19-Berlin)
6. Mietchen, D. et al. [Enabling Open Science: Wikidata for Research \(Wiki4R\)](#). Research Ideas and Outcomes 1, e7573 (2015).

7. Willighagen, E. The WikiPathways and EU NanoSafety Cluster Use Cases. (2018). doi:[10.6084/m9.figshare.6619925.v1](https://doi.org/10.6084/m9.figshare.6619925.v1)
8. Nielsen, F. Å., Mietchen, D. & Willighagen, E. [Scholia, Scientometrics and Wikidata](#). ESWC 2017 Satellite Events, LNCS 10577, pp. 237–259 (2017). doi:[10.1007/978-3-319-70407-4\\_36](https://doi.org/10.1007/978-3-319-70407-4_36)
9. Ekins, S. et al. [Open drug discovery for the Zika virus](#). F1000Research 5, (2016).
10. Totaro, S. et al. [The JRC Nanomaterials Repository: A unique facility providing representative test materials for nanoEHS research](#). Regulatory Toxicology and Pharmacology 81, 334–340 (2016).
11. Willighagen, E. & Chang, J. eNanoMapper Ontology IRIs for the JRC representative industrial nanomaterials. (2018). <http://specs.enanomapper.org/2018/WD-jrc-20180120/>