
Plan Overview

A Data Management Plan created using DMPonline

Title: Converting supplementary info of 10.1021/acsnano.8b07562 into FAIR data

Creator: Egon Willighagen

Principal Investigator: Egon Willighagen

Data Manager: Egon Willighagen

Affiliation: Other

Funder: European Commission

Template: Horizon 2020 Template

ORCID iD: 0000-0001-7542-0286

Project abstract:

The purpose of this project is to convert the supplementary information of this ACS Nano paper: Labouta HI, Asgarian N, Rinker K, Cramb DT. Meta-Analysis of Nanoparticle Cytotoxicity via Data-Mining the Literature. ACS Nano. 2019 Jan 31; doi:10.1021/acsnano.8b07562

ID: 63977

Last modified: 03-10-2020

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Converting supplementary info of 10.1021/acsnano.8b07562 into FAIR data - Initial DMP

1. Data summary

Provide a summary of the data addressing the following issues:

- **State the purpose of the data collection/generation**
- **Explain the relation to the objectives of the project**
- **Specify the types and formats of data generated/collected**
- **Specify if existing data is being re-used (if any)**
- **Specify the origin of the data**
- **State the expected size of the data (if known)**
- **Outline the data utility: to whom will it be useful**

The purpose of this project is to improve the FAIRness of the original data (doi:10.1021/acsnano.8b07562). It is part of the making existing data more reusable. The input is a spreadsheet and the output RDF using the eNanoMapper ontology. In both bases, the data is a few megabytes at most.

2. FAIR data

2.1 Making data findable, including provisions for metadata:

- **Outline the discoverability of data (metadata provision)**
- **Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?**
- **Outline naming conventions used**
- **Outline the approach towards search keyword**
- **Outline the approach for clear versioning**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**

Data will be made available via Zenodo or Figshare. Standards developed by eNanoMapper will be used, like the ontology and the RDF data model being developed.

2.2 Making data openly accessible:

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**

- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

The script is provided in a GitHub repository and in time will also be archived on Zenodo:
<https://github.com/NanoSolveIT/10.1021-acsnano.8b07562>

2.3 Making data interoperable:

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
 - **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**
- eNanoMapper ontology
 - Resource Description Framework

2.4 Increase data re-use (through clarifying licenses):

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**
- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

Follows from the above.

3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

Cost will be covered by the NanoSolveIT H2020 grant.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Not applicable.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

No human data is involved.

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Data will be further shared by NanoSolveIT.

Converting supplementary info of 10.1021/acsnano.8b07562 into FAIR data - Detailed DMP

1. Data summary

State the purpose of the data collection/generation

The purpose of this project is to improve the FAIRness of the original data (doi:10.1021/acsnano.8b07562).

Explain the relation to the objectives of the project

It is part of the making existing data more reusable.

Specify the types and formats of data generated/collected

The input is a spreadsheet and the output RDF using the eNanoMapper ontology.

Specify if existing data is being re-used (if any)

Data comes from this paper:

- Labouta HI, Asgarian N, Rinker K, Cramb DT. Meta-Analysis of Nanoparticle Cytotoxicity via Data-Mining the Literature. ACS Nano. 2019 Jan 31; doi:[10.1021/acsnano.8b07562](https://doi.org/10.1021/acsnano.8b07562) ([Scholia](#))

Specify the origin of the data

Experiments in this paper:

- Labouta HI, Asgarian N, Rinker K, Cramb DT. Meta-Analysis of Nanoparticle Cytotoxicity via Data-Mining the Literature. ACS Nano. 2019 Jan 31; doi:[10.1021/acsnano.8b07562](https://doi.org/10.1021/acsnano.8b07562) ([Scholia](#))

State the expected size of the data (if known)

In both bases, the data is a few megabytes at most.

Outline the data utility: to whom will it be useful

People in the field of risk assessment of nanomaterials.

2.1 Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

Data will be made available via various source, include Figshare or Zenodo and potentially also via GitHub. Further distribution will be via the NanoSolveIT Knowledge Infrastructure.

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

The data will get a DOI.

Outline naming conventions used

Labels originate from the paper. Everything gets an IRI (being RDF).

Outline the approach towards search keyword

No keywords will be assigned.

Outline the approach for clear versioning

There will be initially one version. Later, revisions may be developed if changes to the data model or changes in the used ontologies happen.

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Resource Description Framework

2.2 Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

All data will be made available.

Specify how the data will be made available

1. Zenodo or Figshare
2. NanoSolveIT Knowledge Infrastructure
3. (hoped for) other NanoSafety Cluster projects
4. (maybe) GitHub, using Markdown, Bioschemas, etc

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

No special software will be needed. Any tool that can handle data in an RDF format can be used to extract data.

Specify where the data and associated metadata, documentation and code are deposited

<https://github.com/NanoSolveIT/10.1021-acsnano.8b07562>

Specify how access will be provided in case there are any restrictions

Not applicable: no restrictions.

2.3 Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

1. Resource Description Framework
2. eNanoMapper ontology

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

1. eNanoMapper ontology

2.4 Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

Data will be available as CCZero. The original copyright/license claim is as follows:

"Terms & Conditions

Electronic Supporting Information files are available without a subscription to ACS Web Editions. The American Chemical Society holds a copyright ownership interest in any copyrightable Supporting Information. Files available from the ACS website may be downloaded for personal use only. Users are not otherwise permitted to reproduce, republish, redistribute, or sell any Supporting Information from the ACS website, either in whole or in part, in either machine-readable form or any other form without permission from the American Chemical Society. For permission to reproduce, republish and redistribute this material, requesters must process their own requests via the RightsLink permission system. Information about how to use the RightsLink permission system can be found at <http://pubs.acs.org/page/copyright/permissions.html>."

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Soon.

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

No restrictions in reuse.

Describe data quality assurance processes

The data structure is tested with a set of SPARQL queries and manual verification. It has also been reviewed by NanoSolveIT partners.

Specify the length of time for which the data will remain re-usable

Too philosophical.

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Unclear. It is a research project.

Clearly identify responsibilities for data management in your project

The UM partner will develop the script to make the data more FAIR. Integration into the NanoSolveIT Knowledge Infrastructure requires feedback from other partners.

Describe costs and potential value of long term preservation

The data is published data and is expected to have long term value.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Not applicable.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Not applicable.

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

The first author of the publication has been informed about the efforts.

Converting supplementary info of 10.1021/acsnano.8b07562 into FAIR data - Final review DMP

1. Data summary

State the purpose of the data collection/generation

Question not answered.

Explain the relation to the objectives of the project

Question not answered.

Specify the types and formats of data generated/collected

Question not answered.

Specify if existing data is being re-used (if any)

Question not answered.

Specify the origin of the data

Question not answered.

State the expected size of the data (if known)

Question not answered.

Outline the data utility: to whom will it be useful

Question not answered.

2.1 Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

Question not answered.

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

Question not answered.

Outline naming conventions used

Question not answered.

Outline the approach towards search keyword

Question not answered.

Outline the approach for clear versioning

Question not answered.

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Question not answered.

2.2 Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

Question not answered.

Specify how the data will be made available

Question not answered.

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

Question not answered.

Specify where the data and associated metadata, documentation and code are deposited

Question not answered.

Specify how access will be provided in case there are any restrictions

Question not answered.

2.3 Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

Question not answered.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Question not answered.

2.4 Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

Question not answered.

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Question not answered.

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

Question not answered.

Describe data quality assurance processes

Question not answered.

Specify the length of time for which the data will remain re-usable

Question not answered.

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Question not answered.

Clearly identify responsibilities for data management in your project

Question not answered.

Describe costs and potential value of long term preservation

Question not answered.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Question not answered.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Question not answered.

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Question not answered.