

Everything You Always Wanted to Know About Data Citations (*But Were Afraid to Ask)*

Nina Jeliaskova, Thomas Exner, Egon Willighagen
6 januari 2020, Reykjavik, Iceland



Data as primary research output

- Data curation is research
- Data is more precise (better) than articles

Agreed? No?

Room for everyone's talent

towards a new balance in the recognition and rewards of academics



Data as primary research output

- Data curation is research
- Data is more precise (better) than articles

Agreed? No?

Well, we tended to hoard our data, but gave away our articles for free. Which one did we really value more?

Room for everyone's talent

towards a new balance in the recognition and rewards of academics



Workshop

10:00 Introduction (these slides)

10:20 Split up in groups of 3 or 4 people, possible project specific

10:25 Perform one or more of these tasks related to your project:

- Use DataCite to format a data citation to a data set from your project
- List one or more DOIs of datasets released by your project
- Determine with data search engine finds a useful dataset
- State which EU project pages should list datasets

10:45 Each group summarizes their results

Data and findability

- Traditional: central databases
 - Chemical Abstracts
 - Ensembl/UniProt, NCBI
 - PubChem
 - ...
- Moving to a decentralized world
 - Search.data.enanomapper.net
 - Google Dataset Search
 - DataCite
 - ...
- Archives versus databases
- Data versus dataset
- Data versus information



Image: CC-BY 3.0, WikiCommons, File:Duke Humfrey's Library Interior 6, Bodleian Library, Oxford, UK - Diliff.jpg

Searching on chemistry

- ▶ Projects (7084)
- ▶ Study providers (7084)
- ▶ Nanomaterial type (7084)
- ▶ Nanomaterial (6268)
- ▼ Protocols (21065)

ALAMAR BLUE

ANNEXIN V / PI ...

APOPTOSIS ATR-FTIR

BATCH DISPERS...

CELL COUNT

CELL CYCLE

CHN-ANALYSIS

COLONY FORMI...

COMET

CYTOKINE SECR...

[Hits list](#)[Selection](#)[Predefined Queries](#)[Export](#)

< 1 > displaying 1 to 14 of 14



Ag @ IIT (Ag 20 nm) silver nanoparticle

CORE (1): ...

Results:P-CHEM.Crystalline phase, P-CHEM.Surface chemistry, P-CHEM.Particle size distribution (Granulometry)

[more](#)

[Material](#) [Composition](#) [Studies](#)

[Add to Selection](#)



NFC Fine (Nanofibrillar cellulose 2-15 nm) nanofibrillar cellulose

CORE (1): ...

Results:P-CHEM.Analytical Methods, P-CHEM.Crystalline phase, P-CHEM.Surface chemistry, P-CHEM.Particle size distribution (Granulometry), TOX.Immunotoxicity, TOX.Cell Viability, TOX.Genetic toxicity in vitro, TOX.Genetic toxicity in vivo, TOX.Repeated dose toxicity - inhalation, TOX.Repeated dose toxicity - oral

[more](#)

[Material](#) [Composition](#) [Studies](#)

[Add to Selection](#)



JRCNM01001a (NM-101 (TiO2 6 nm)) titanium oxide nanoparticle

CORE (1): ...

Results:P-CHEM.Specific surface area, P-CHEM.Crystalline phase, P-CHEM.Surface chemistry, P-CHEM.Particle size distribution (Granulometry), P-CHEM.Water solubility, TOX.Barrier integrity, TOX.Genetic toxicity in vitro, TOX.Cell Viability, TOX.Oxidative

Google Dataset Search

The screenshot shows the Google Dataset Search interface. At the top, the Google logo is on the left, and a search bar contains the text 'nanowiki'. To the right of the search bar are icons for help, chat, and a grid menu, along with a user profile picture. Below the search bar, there are filter buttons: 'Updated date', 'Download format', 'Usage rights', and 'Free'. The main content area is divided into two columns. The left column, titled '5 data sets found', lists four results, each with a green circular icon containing the letter 'F'. The first result is highlighted with a light blue background. The right column provides detailed information for the selected dataset, including a title, a 'Related Article' link, a blue 'Explore at figshare.com' button, a 'Unique identifier' with a DOI link, 'Data set updated' date, 'Data set provided by' organization, 'Authors' list, 'Licence' information, and 'Available download formats from providers'.

Google

nanowiki

Updated date Download format Usage rights Free

5 data sets found

Data from: NanoWiki (release 1)
figshare.com
Updated Jan 19, 2016

NanoWiki 5
figshare.com
search.datacite.org
Updated Sep 11, 2018

NanoWiki (release 4)
figshare.com
search.datacite.org
Updated Nov 6, 2016

NanoWiki (release 3)
figshare.com
Updated May 14, 2016

Data from: NanoWiki (release 1)

Related Article

Explore at figshare.com

Unique identifier
<https://doi.org/10.6084/m9.figshare.1330208.v1>

Data set updated Jan 19, 2016

Data set provided by
figshare

Authors
Egon Willighagen

Licence
[Attribution 4.0 \(CC BY 4.0\)](#)
Licence information was derived automatically

Available download formats from providers
application/gzip

DataCite

DataCite Search [Works](#) [People](#) [Repositories](#) [Members](#) [Support](#) [Sign in](#)

nanowiki

13 Works

Networked experiments and scientific resource sharing in cooperative knowledge spaces


Sabine Cikic, Sabina Jeschke, Nadine Ludwig, Uwe Sinha & Christian Thomsen
Work published 2007 via Technische Universität Berlin
Dieser Beitrag ist mit Zustimmung des Rechteinhabers aufgrund einer (DFG geförderten) Allianz- bzw. Nationallizenz frei zugänglich.

i No citations were reported. No usage information was reported.

<https://doi.org/10.14279/depositonce-9426> **“** Cite

NanoWiki 5

Egon Willighagen
Dataset published 2018 via Figshare
New release with more JRC nanomaterials annotated with ENM ontology terms, and data from a 2013 NanoQSAR study from Small by Lin et al. on 24 metal oxides.




Registration Year

<input type="checkbox"/> 2015	1
<input type="checkbox"/> 2016	8
<input type="checkbox"/> 2017	1
<input type="checkbox"/> 2018	2
<input type="checkbox"/> 2019	1

Resource Types

<input type="checkbox"/> Dataset	11
<input type="checkbox"/> Text	1

 FEEDBACK

How much data is out there?

DataCite Statistics

Registrations by Member

Registrations by Repository

Resolutions by Month

Download CSV

Statistics

ID	Name	DOI Registrations				DOI Metadata	
		Total	2020	2019	This month	Findable	Registered
AKBILD	Academy of Fine Arts Vienna	199	0	198	0	199	0
ANDS	Australian Research Data Commons	343,740	2,937	54,016	806	341,032	2,708
ARIDHIA	Aridhia Informatics Ltd.	6	0	6	0	6	0
AU	American University Library	243	36	170	11	243	0
AUSTINTX	City of Austin, Texas	513	0	484	0	499	14
BF	Blackfynn Inc.	126	4	122	0	118	8
BIBSYS	Unit – The Norwegian Directorate for ICT and Joint ...	6,997	245	2,995	119	6,988	9
BL	The British Library	535,584	9,855	99,715	984	520,362	15,222
BRAINL	Brain Life	283	5	138	0	283	0
BROAD	Eli and Edythe L. Broad Institute of MIT and Harvard	13,674	0	0	0	13,673	1
BROWN	Brown University Library	6,015	37	2,590	7	6,014	1
CALTECH	California Institute of Technology	23,663	498	6,883	4	23,652	11
CCDC	The Cambridge Crystallographic Data Centre	834,207	6,321	64,560	675	830,329	3,878
CCOM	Center for Coastal & Ocean Mapping - University of ...	4	0	0	0	0	4
CDL	California Digital Library	225,065	490	10,641	47	224,742	323
CERN	CERN - European Organization for Nuclear Research	1,711,597	33,830	834,371	3,101	1,552,386	159,211
CHOP	Children's Hospital of Philadelphia	2	0	1	0	1	1

Where to archive data?



The Zenodo website header features the Zenodo logo on the left, a search bar in the center, and navigation links for 'Upload' and 'Communities' on the right. A user profile icon and email address 'egon.willighagen@gmail.com' are also visible.

Recent uploads

September 16, 2019 (v12) Dataset Open Access

View

Binary black-hole surrogate waveform catalog

Scott E. Field; Chad R. Galley; Jan S. Hesthaven; Jason Kaye; Manuel Tiglio; Jonathan Blackman; Béla Szilágyi; Mark A. Scheel; Daniel A. Hemberger; Patricia Schmidt; Rory Smith; Christian D. Ott; Michael Boyle; Lawrence E. Kidder; Harald P. Pfeiffer; Vijay Varma

This repository contains all publicly available numerical relativity surrogate data for waveforms produced by the Spectral Einstein Code. The base method for building surrogate models can be found in Field et al., PRX 4, 031006 (2014). Several numerical relativity surrogate models are currently...

Uploaded on January 28, 2020

11 more version(s) exist for this record

January 24, 2020 (v0.10.0) Software Open Access

View

mwaskom/seaborn: v0.10.0 (January 2020)

Michael Waskom; Olga Botvinnik; Joel Ostblom; Saulius Lukauskas; Paul Hobson; MaozGelbart; David C. Gemperline; Tom Augspurger; Yaroslav Halchenko; John B. Cole; Jordi Warmerhoven; Julian de Rutter; Cameron Fyfe; Stephan Hoyer; Jake Vanderplas; Santi Villalba; Gero Kunter; Eric Quintero; Pete Bachant; Marcel Martin; Kyle Meyer; Corban Swain; Alistair Miles; Thomas Brunner; Drew O'Kane; Tal Yarkoni; Mike Lee Williams; Constantine Evans

This is a major update that is being released simultaneously with version 0.9.1. It has all of the same features (and bugs!) as 0.9.1, but there are important changes to the dependencies. Most notably, all support for Python 2 has now been dropped. Support for Python 3.5 has also been dropped...

Uploaded on January 28, 2020

9 more version(s) exist for this record

Zenodo now supports usage statistics!



Read more about it, in our newest blog post.

Using GitHub?



Check out our GitHub integration. Software Preservation Made Simple!

Zenodo in a nutshell

- **Research. Shared.** — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
- **Citeable. Discoverable.** — uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- **Communities** — create and curate your own community for a workshop, project, department, journal, into which you can accept



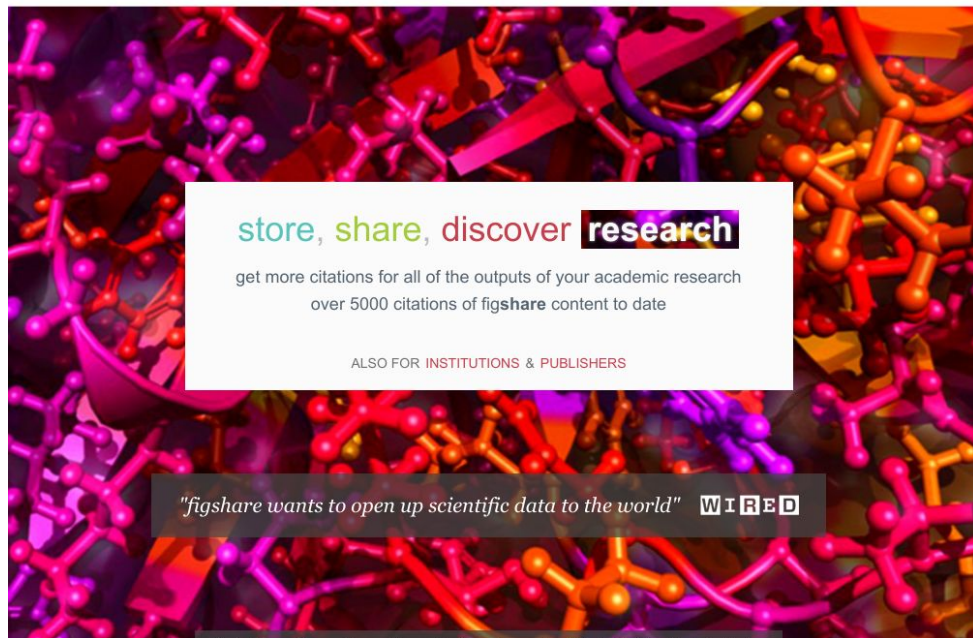
Browse

Search on figshare...



Log in

Sign up



The banner features a background of colorful molecular models. A central white box contains the text 'store, share, discover research' in a mix of colors. Below this, it states 'get more citations for all of the outputs of your academic research over 5000 citations of figshare content to date'. At the bottom, it says 'ALSO FOR INSTITUTIONS & PUBLISHERS'. A dark grey box at the bottom right contains the quote 'figshare wants to open up scientific data to the world' followed by the WIRED logo.

How to cite data?

DOI Citation Formatter

Paste your DOI:

For example 10.1145/2783446.2783605

Select Formatting Style:

Begin typing (e.g. Chicago or IEEE.) or use the drop down menu.

Select Language and Country:

Begin typing (e.g. en-GB for English, Great Britain) or use the drop down menu.

Format

Willighagen, E. (2018). NanoWiki 5 [Data set]. <https://doi.org/10.6084/M9.FIGSHARE.7075214>

Copy to clipboard





Do you want to integrate this service? Check the [Documentation](#)

DOI Registration Agencies



How to make data findable?

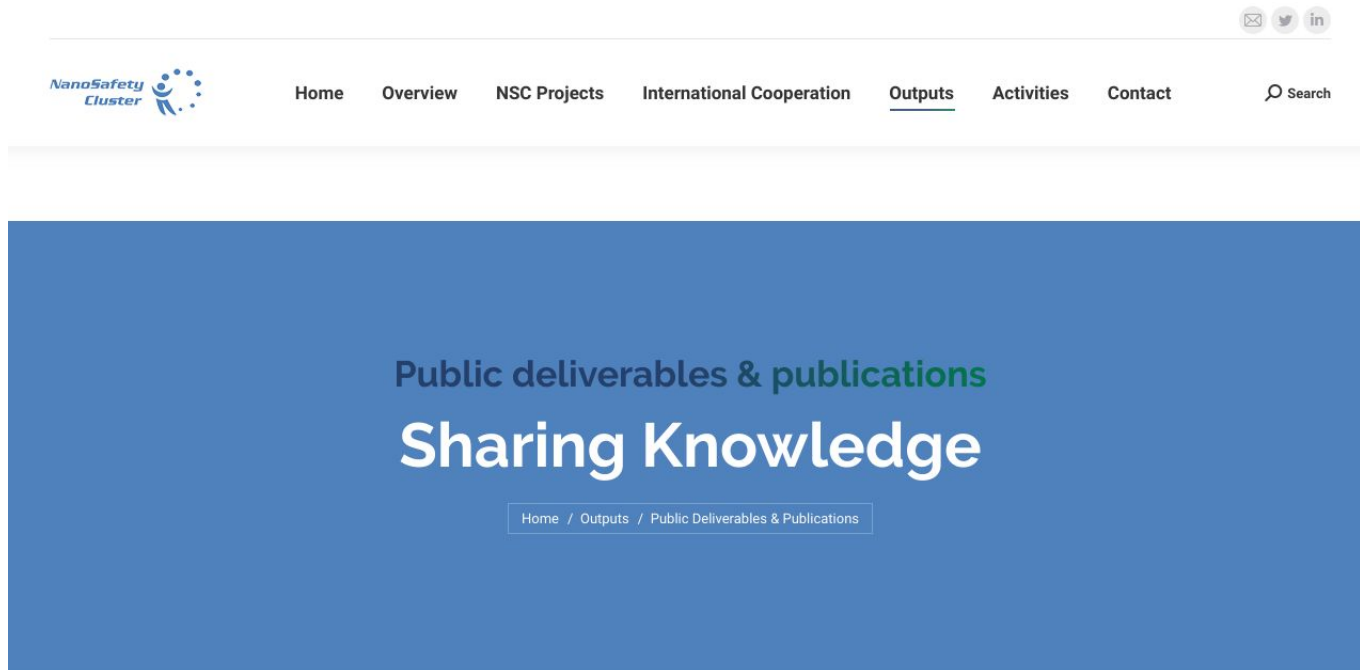
Examples 

Property	Expected Type	Description	CD	Controlled Vocabulary	Example
Marginality: Minimum.					
<u>description</u>	<u>Text</u>	Schema: A description of the item. Bioschemas: A short summary describing a dataset.	ONE		
<u>identifier</u>	<u>PropertyValue</u> <u>Text</u> <u>URL</u>	Schema: The identifier property represents any kind of identifier for any kind of Thing, such as ISBNs, GTIN codes, UUIDs etc. Schema.org provides dedicated properties for representing many of these, either as textual strings or as URL (URI) links. See <u>background notes</u> for more details.	MANY		
<u>keywords</u>	<u>Text</u>	Schema: Keywords or tags used to describe this content. Multiple entries in a keywords list are typically delimited by commas. Bioschemas: These keywords provide a summary of the dataset.	MANY		
<u>name</u>	<u>Text</u>	Schema: The name of the item.	ONE		



Bioschemas

How to brag
about your
data?



The screenshot shows the NanoSafety Cluster website. The top navigation bar includes the logo, a search icon, and links for Home, Overview, NSC Projects, International Cooperation, Outputs, Activities, and Contact. Social media icons for email, Twitter, and LinkedIn are in the top right. The main content area has a blue background with the text 'Public deliverables & publications' and 'Sharing Knowledge'. A breadcrumb trail below reads 'Home / Outputs / Public Deliverables & Publications'.

Overview

This page provides available knowledge from the nanosafety community, as produced by projects and published



Workshop

10:00 Introduction (these slides)

10:20 Split up in groups of 3 or 4 people, possible project specific

Coffee break

10:45 Perform one or more of these tasks related to your project:

- Use DataCite to format a data citation to a data set from your project
- List one or more DOIs of datasets released by your project
- Determine with data search engine finds a useful dataset
- State which EU project pages should list datasets

11:00 Each group summarizes their results (until 11:15)

Acknowledgments

This presentation was given as part of the three projects funded by the European Commission, under grant n° 731032 (NanoCommons), 814425 (RiskGONE), and 814572 (NanoSolveIT).

