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

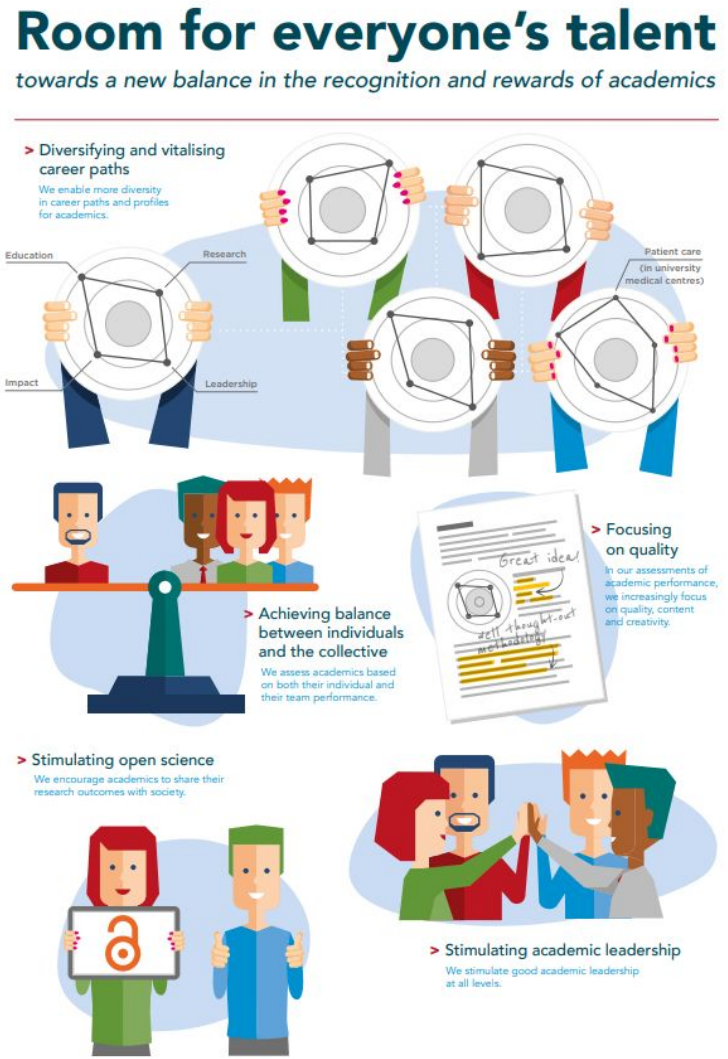
Nina Jeliazkova, Thomas Exner, Egon Willighagen

6 januari 2020, Reykjavik, Iceland

Data as primary research output

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

Agreed? No?



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?

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

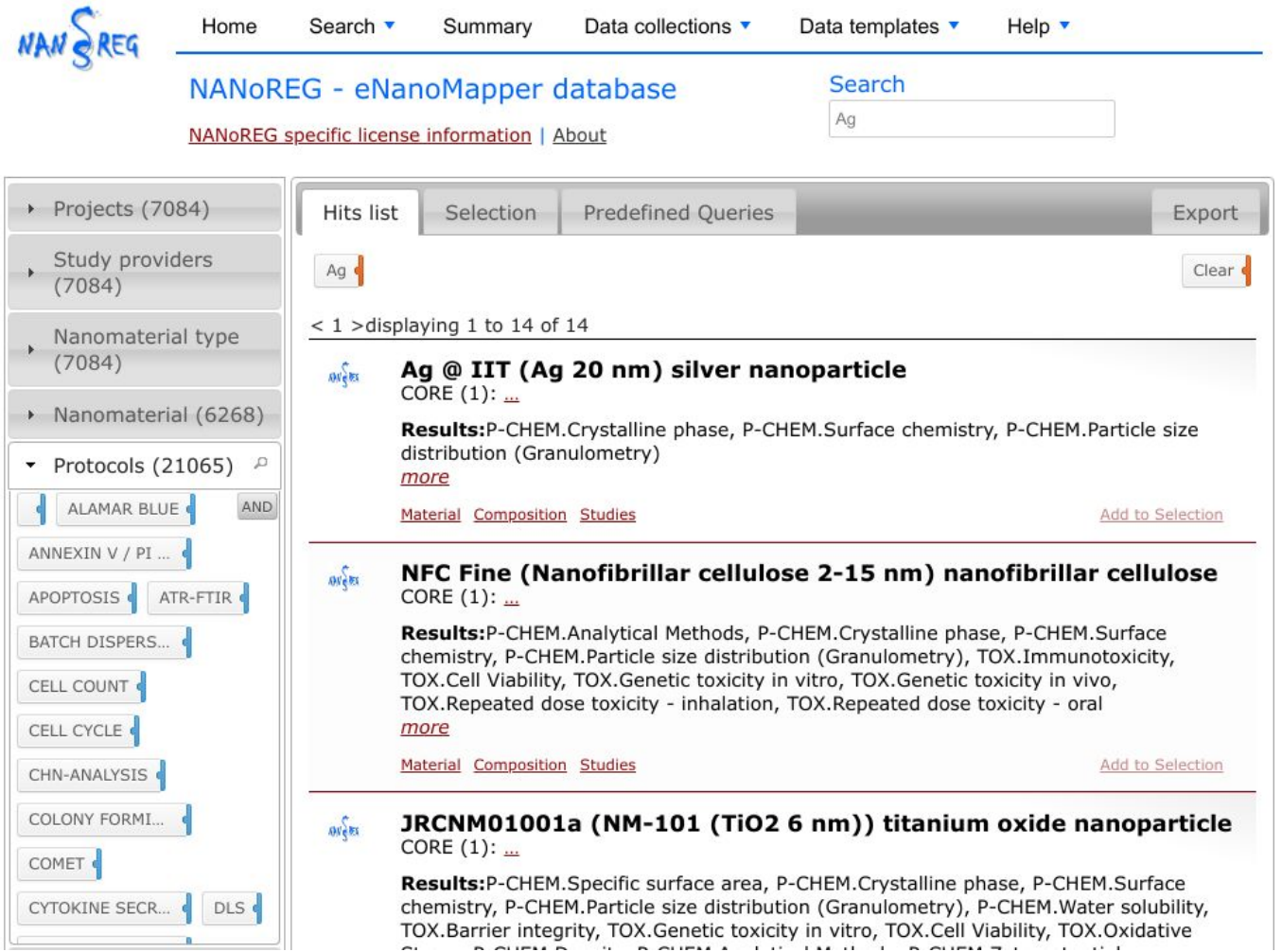
Data and findability

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

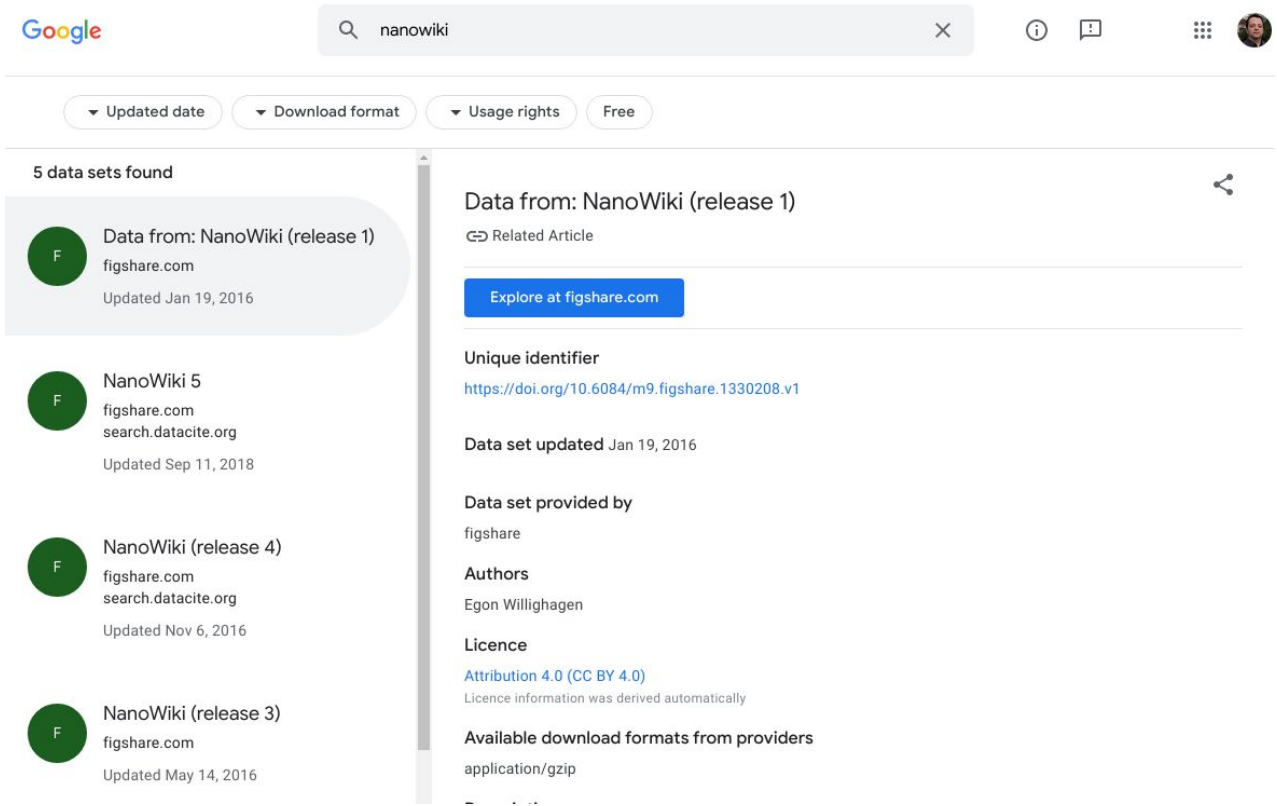
- Archives versus databases
- Data versus dataset
- Data versus information

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

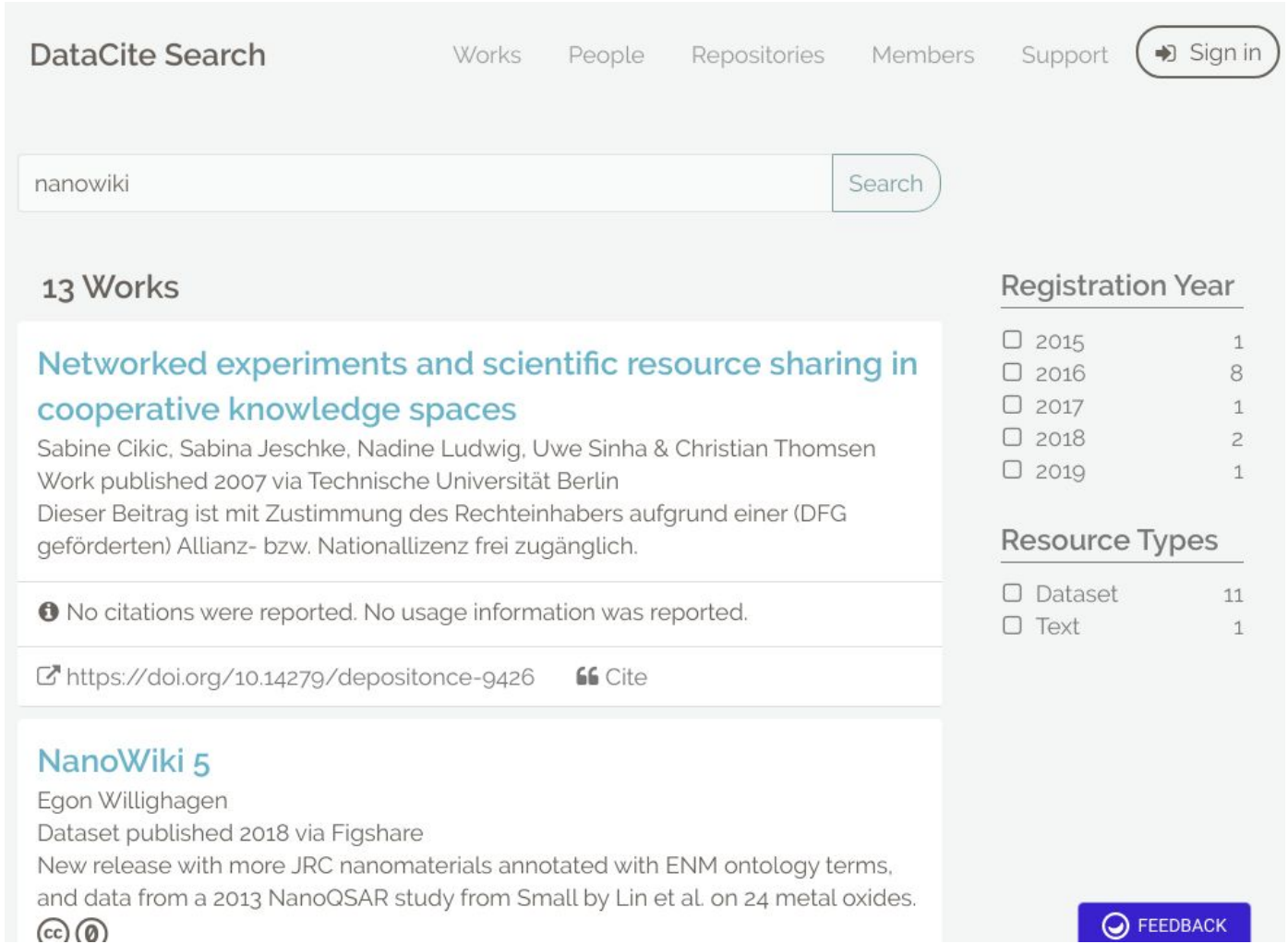
Searching on chemistry



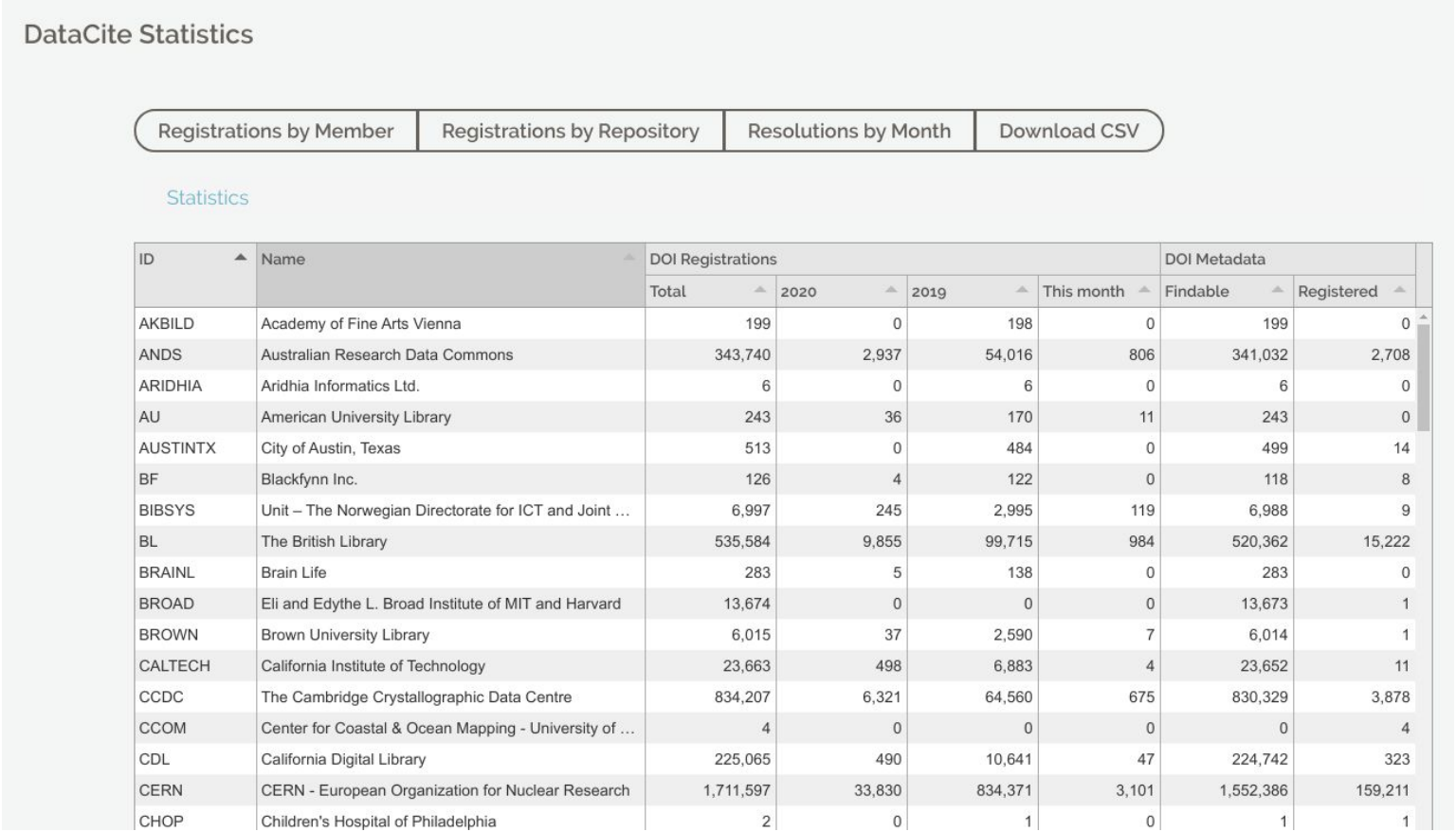
Google Dataset Search



DataCite


DataCite Search
Works
People
Repositories
Members
Support
Sign in
nanowiki
Search
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.
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
2015 1
2016 8
2017 1
2018 2
2019 1
Resource Types
Dataset 11
Text 1
FEEDBACK

How much data is out there?



DataCite Statistics

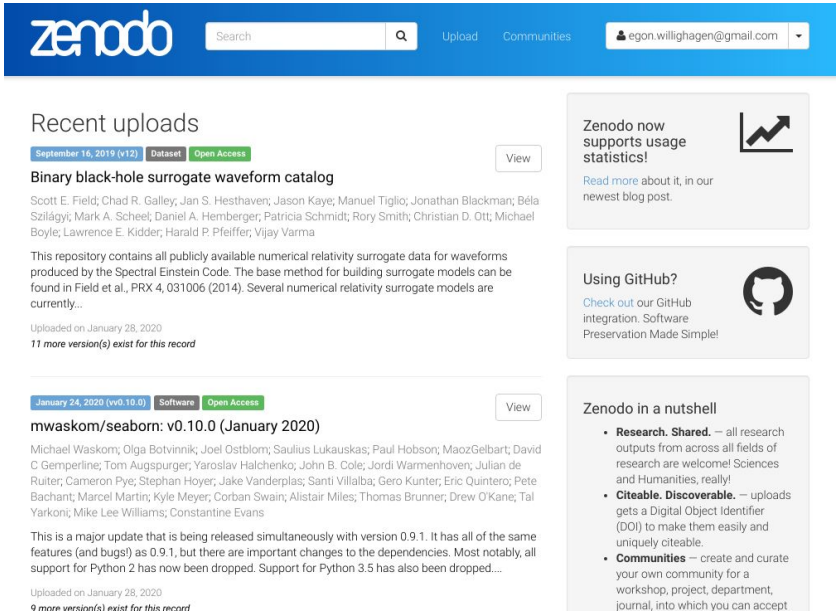
Registrations by Member | Registrations by Repository | Resolutions by Month | Download CSV

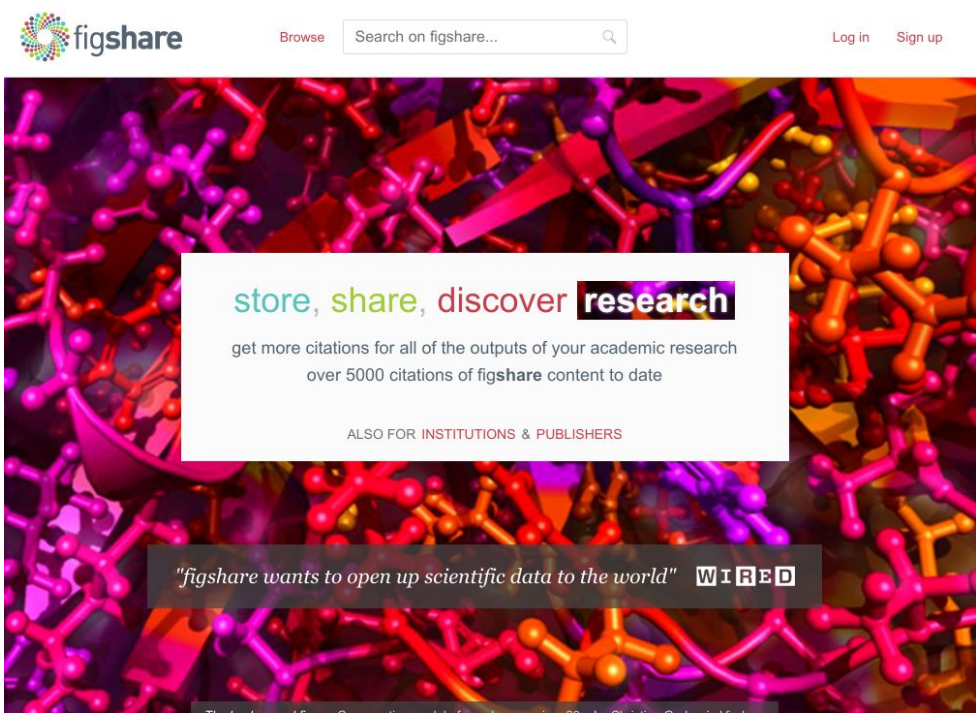
Statistics

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

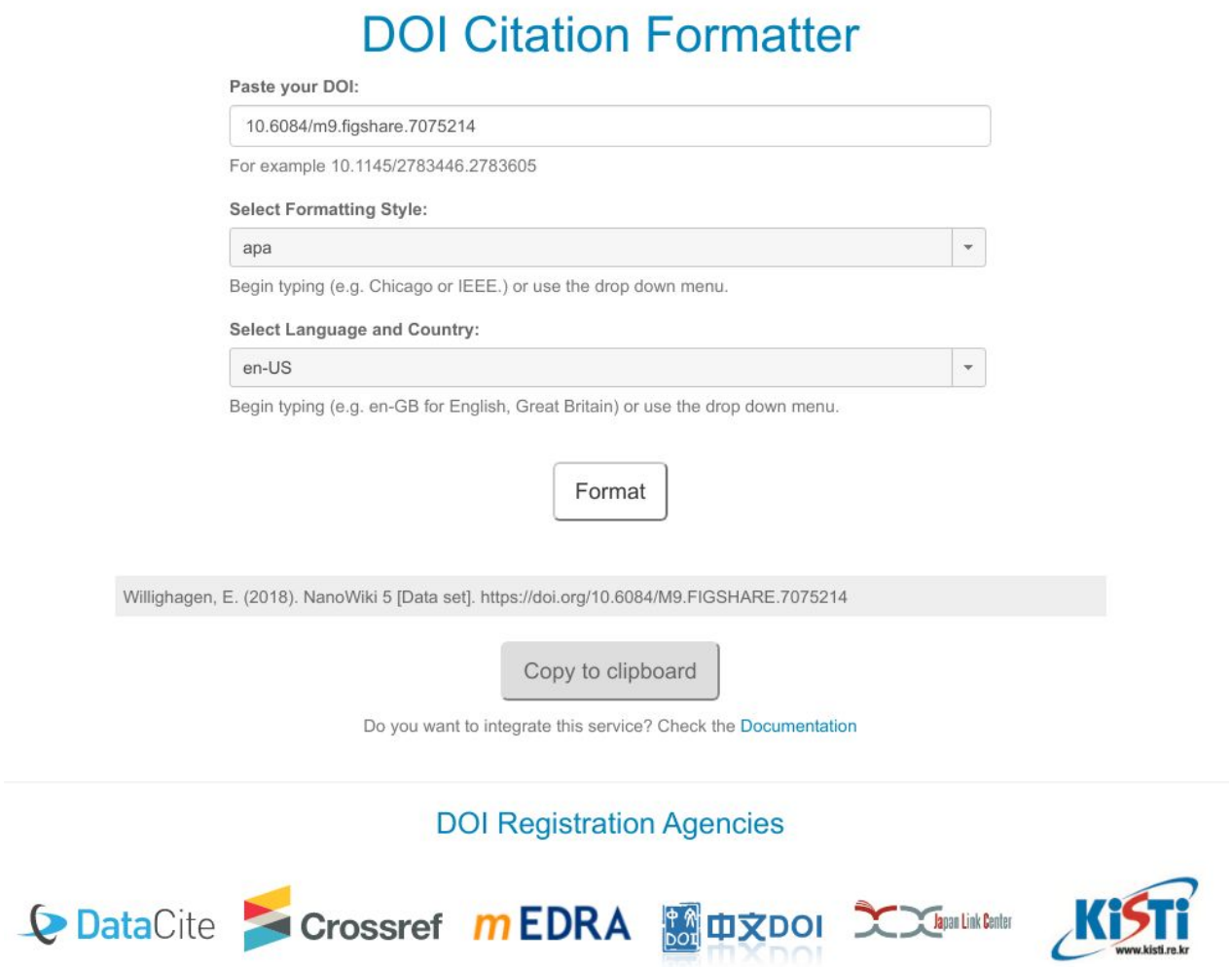
https://stats.datacite.org/

Where to archive data?





How to cite data?



How to make data findable?

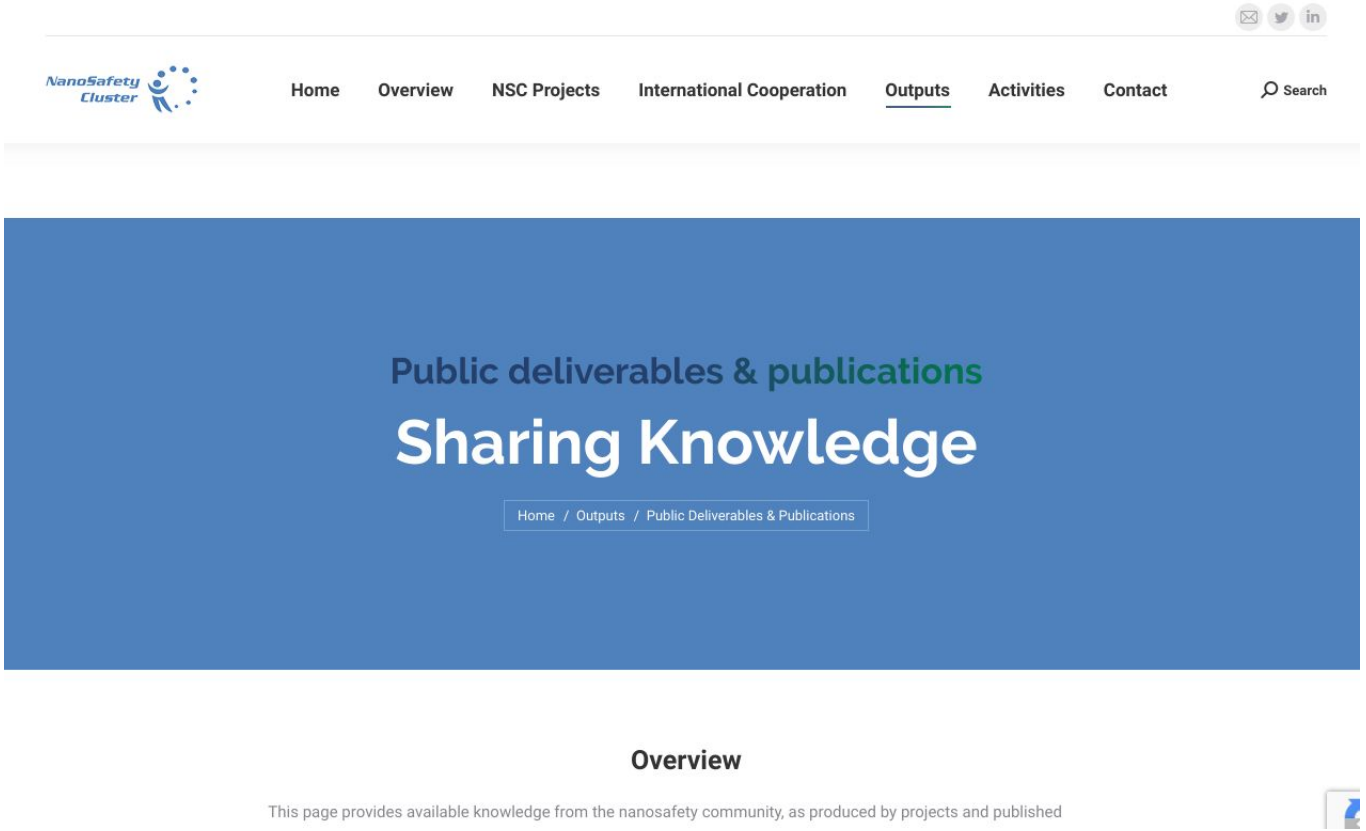
Examples

Property	Expected Type	Description	CD	Controlled Vocabulary	Example
Marginality: Minimum.					
description	Text	**Schema:** A description of the item. **Bioschemas:** A short summary describing a dataset.	ONE		
identifier	PropertyValue Text URL	**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 background notes for more details.	MANY		
keywords	Text	**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		
name	Text	**Schema:** The name of the item.	ONE		

elixir

Bioschemas

How to brag about your data?



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 nº 731032 (NanoCommons), 814425 (RiskGONE), and 814572 (NanoSolveIT).