

...-data needs meta-...

Paul Vierkant

16 November 2022

NFDI4Culture-Forum Datenpublikation und
-archivierung #5: Persistent Identifiers

<https://doi.org/10.5281/zenodo.7326660>



[@paul4kant@scicomm.xyz](mailto:paul4kant@scicomm.xyz)



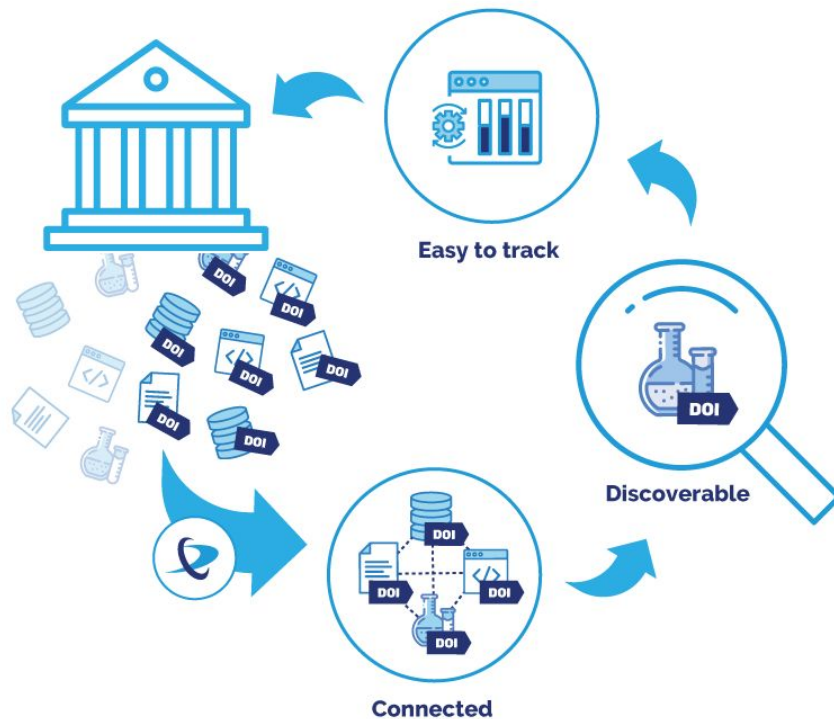
[@datacite](https://twitter.com/datacite)



DataCite

About DataCite

Global non-profit membership organization working with 2100+ repositories in the world to provide Digital Object Identifiers (DOIs) for **data and other research outputs.**



DataCite Community



2100+

Repositories

238

Members

44

Countries

27m+

DOIs

1000+

Organizations

Metadata

Schema

What is the DataCite Metadata Schema?

- A metadata schema specifically for DataCite DOIs

Who develops it?

- The [DataCite Metadata Working Group](#) and DataCite staff

Where can I find it?

- All versions: <https://schema.datacite.org/>
- Latest version (4.4):
<https://schema.datacite.org/meta/kernel-4.4/>

When is it updated?

- About every 1-2 years

What is in the schema?

- The schema consists of 20 metadata **properties** (sometimes called “fields” or “elements”).
- Hierarchical structure: some properties have sub-properties.
- Some are mandatory, some recommended or optional.
- Some can be repeated.
- Some have controlled list values.

20 metadata properties:

6 mandatory

- Identifier, Creator, Title, Publisher, PublicationYear, ResourceType

6 recommended

- Subject, Contributor, Date, RelatedIdentifier, Description, GeoLocation

8 optional

- Language, AlternateIdentifier, Size, Format, Version, Rights, FundingReference, RelatedItem

Lots of sub-properties

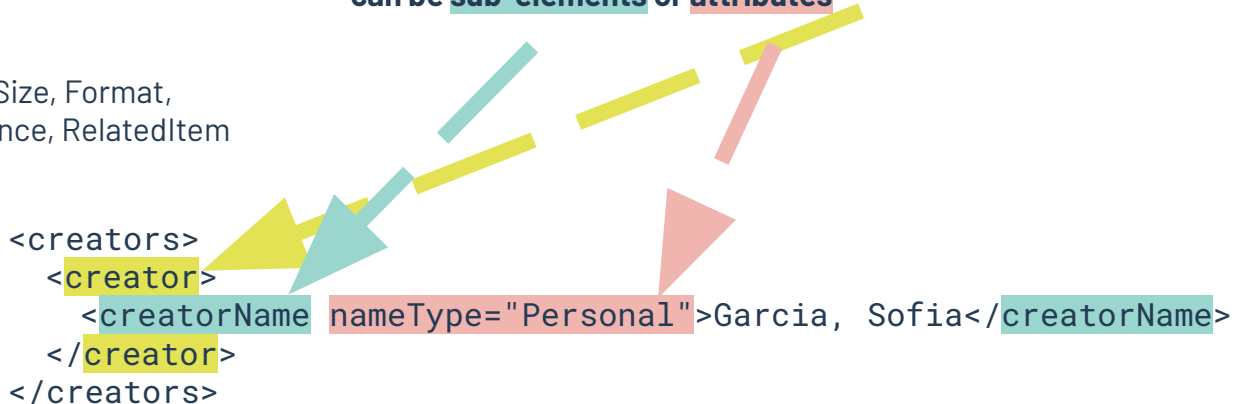
Some mandatory, some optional

Some have their own sub-properties

Some repeatable, some non-repeatable

Metadata is primarily represented in XML

Properties are represented as **elements**; sub-properties can be **sub-elements** or **attributes**



```
<creators>  
  <creator>  
    <creatorName nameType="Personal">Garcia, Sofia</creatorName>  
  </creator>  
</creators>
```

variety of resource types:

Audiovisual	Model
Book	OutputManagementPlan
BookChapter	PeerReview
Collection	PhysicalObject
ComputationalNotebook	Preprint
ConferencePaper	Report
ConferenceProceeding	Service
DataPaper	Software
Dataset	Sound
Dissertation	Standard
Event	Text
Image	Workflow
InteractiveResource	Other
Journal	
JournalArticle	

Recommended and optional properties:

- **Subject**
- **Contributor**
- **Date**
- **RelatedIdentifier**
- **Description**
- **GeoLocation**
- **Language**
- **AlternateIdentifier**
- **Size**
- **Format**
- **Version**
- **Rights**
- **FundingReference**
- **RelatedItem**

RelatedIdentifier

RelatedIdentifier is used for identifiers of related resources.

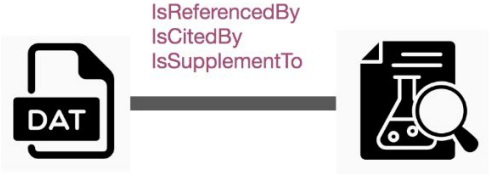
Mandatory sub-properties:

- `relationType`: the type of *relationship*
- `relatedIdentifierType`: the type of identifier

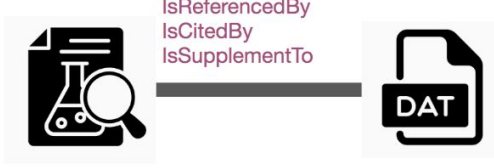
relationTypes	relatedIdentifierTypes
IsCitedBy / Cites	ARK
IsReferencedBy / References	arXiv
IsSupplementTo / IsSupplementedBy	bibcode
IsNewVersionOf / IsPreviousVersionOf	DOI
HasVersion / IsVersionOf	EAN13
IsObsoletedBy / Obsoletes	EISSN
IsIdenticalTo	Handle
IsDescribedBy / Describes	IGSN
HasMetadata / IsMetadataFor	ISBN
IsDocumentedBy / Documents	ISSN
IsReviewedBy / Reviews	ISTC
IsPartOf / HasPart	LISSN
IsPublishedIn	LSID
IsCompiledBy / Compiles	PMID
IsVariantFormOf / IsOriginalFormOf	PURL
IsDerivedFrom / IsSourceOf	UPC
IsRequiredBy / Requires	URL
IsContinuedBy / Continues	URN
	w3id

relatedIdentifier metadata objects

Citations



References



Metadata functions

Descriptive metadata supports discovery and citation

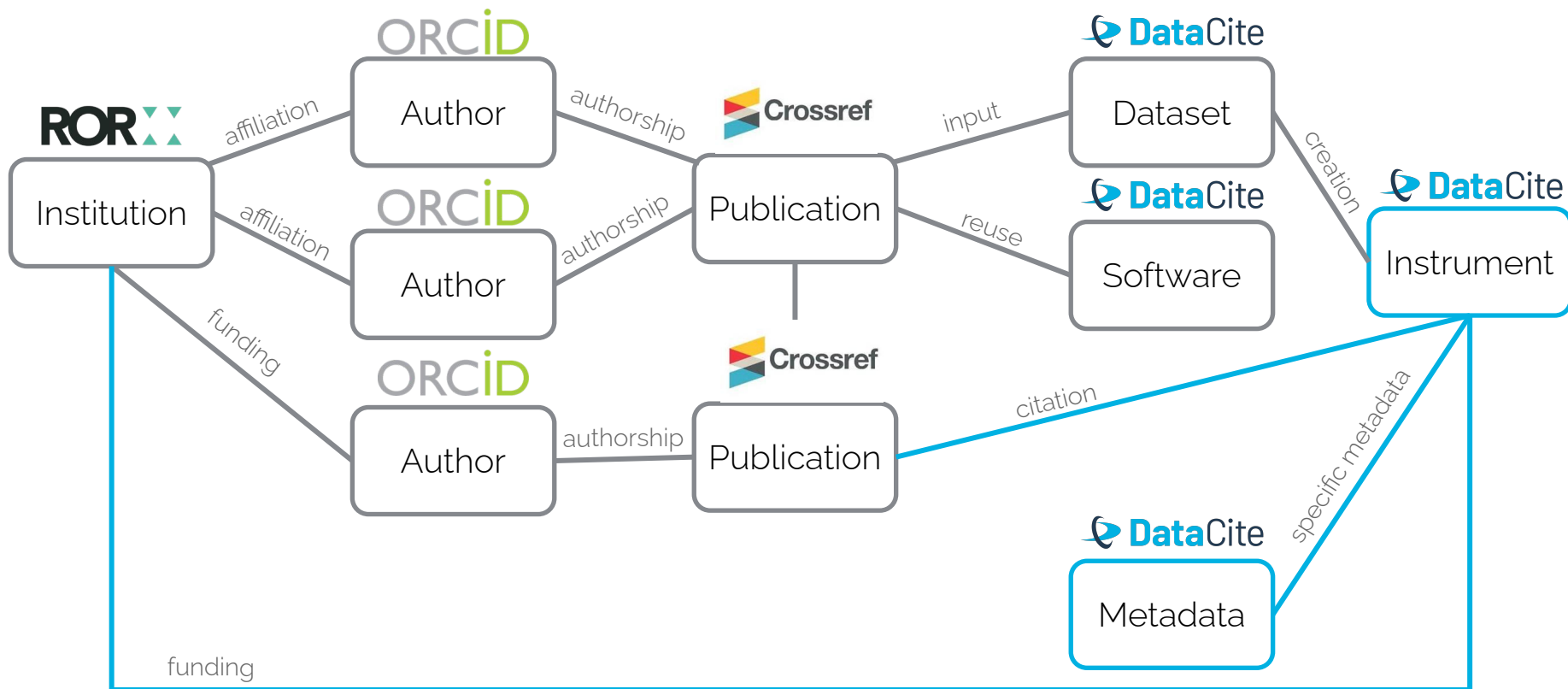
- Titles, creators, subjects, description, dates...

Connection metadata promotes interoperability

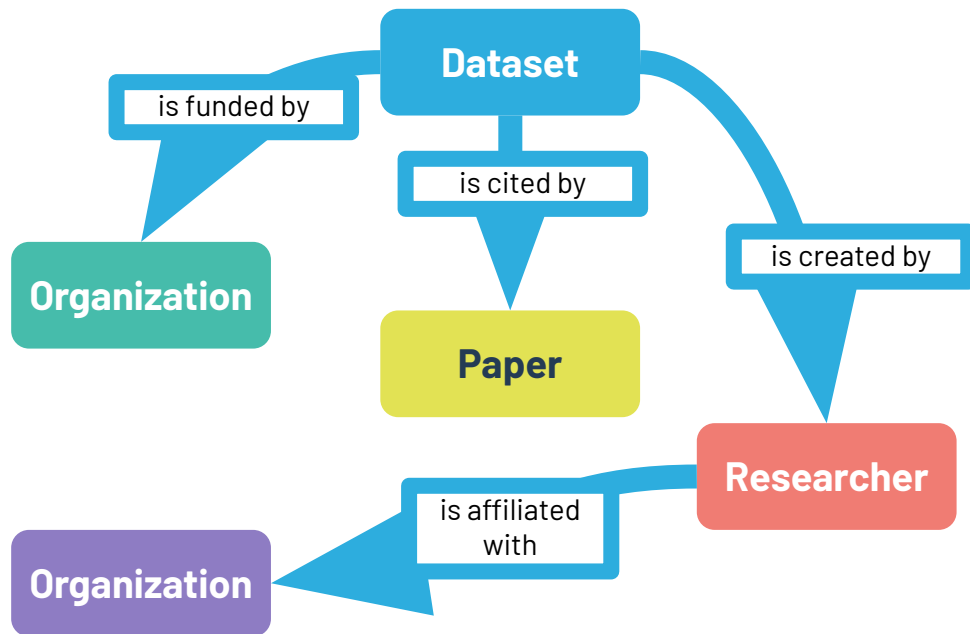
- Related identifiers
- Name identifiers - for creators and contributors
- Affiliation identifiers - for creators and contributors
- Funder identifiers - for funders



Resource Types and Relations



Illustrating connection metadata

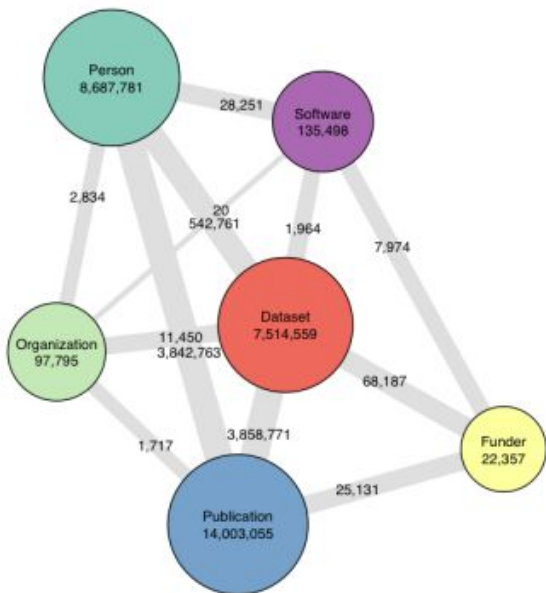


```
<creator>
  <creatorName nameType="Personal">Garcia, Sofia</creatorName>
  <givenName>Sofia</givenName>
  <familyName>Garcia</familyName>
  <nameIdentifier schemeURI="https://orcid.org/"
  nameIdentifierScheme="ORCID">0000-0001-5727-2427</nameIdentifier>
  <affiliation affiliationIdentifier="https://ror.org/03efmqc40"
  affiliationIdentifierScheme="ROR" SchemeURI="https://ror.org">Arizona State
  University</affiliation>
</creator>

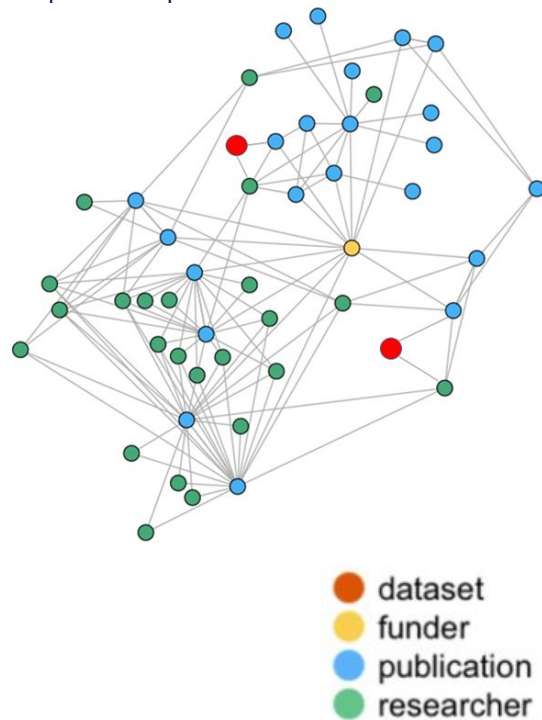
<relatedIdentifier relatedIdentifierType="DOI"
relationType="IsCitedBy">10.5438/ExampleArticle</relatedIdentifier>

<fundingReference>
  <funderName>European Commission</funderName>
  <funderIdentifier funderIdentifierType="Crossref Funder
  ID">https://doi.org/10.13039/501100000780</funderIdentifier>
</fundingReference>
```

PID Graph



Beispiel-Graph:



Impact of connection metadata

- Makes interoperability possible!
- Helps illustrate how research outputs, people, and institutions are connected
- Disambiguation based on identifier
 - Example: Using ROR IDs to retrieve DOIs by organization—instead of relying on affiliation strings
- Systems talk to each other
 - Example: ORCID Auto-Update: update your ORCID profile when a DOI is registered using your ORCID iD

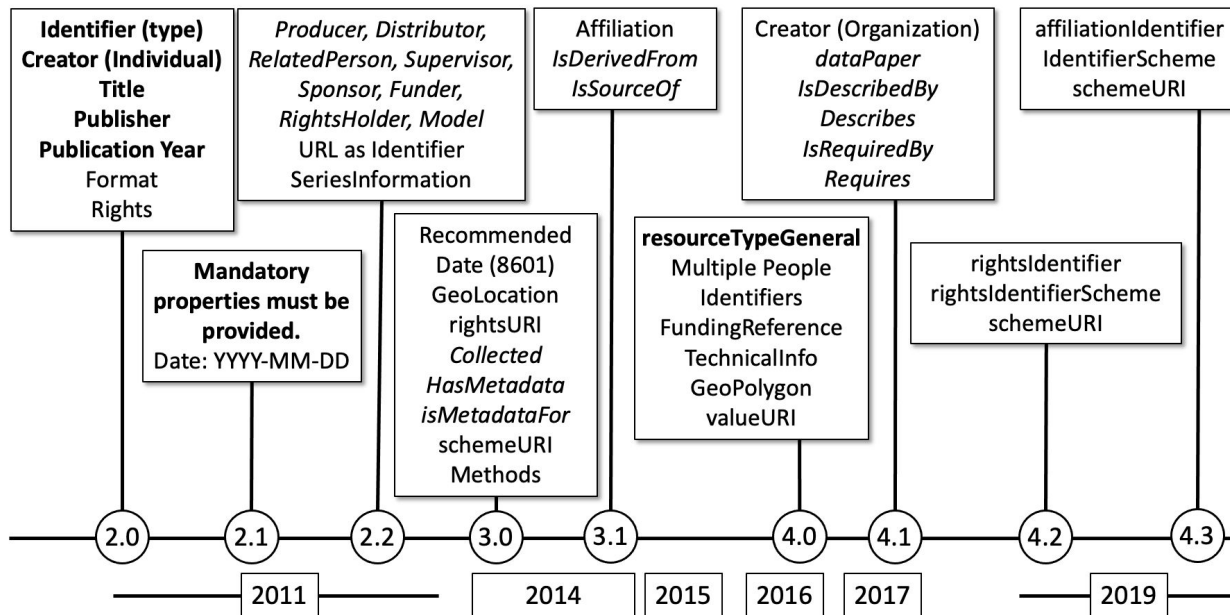
Relation to FAIR principles

More complete metadata...

- makes research outputs easier to **find**
- clarifies conditions of **access**
- promotes **interoperability** between systems
- communicates the context to promote **re-use**

Also to be FAIR

DataCite FAIR: Dialect Evolution



DataCite
Commons

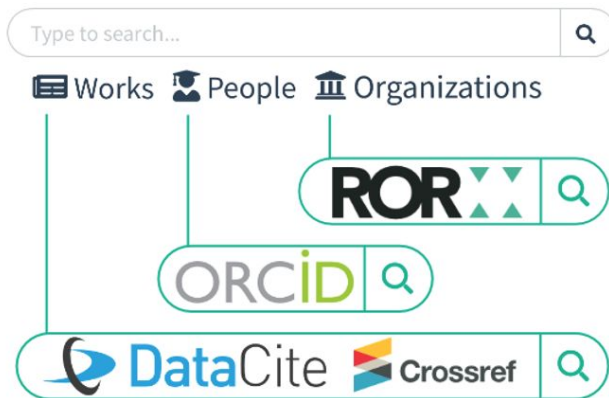
DataCite Commons

DataCite Commons shows:

- The relationships between DOIs in the form of citations, versions, and collections.
- The relationships between research output with DOIs, researchers (ORCID), research organizations (ROR), and funders (Crossref Funder ID), e.g., all research output/funder or all research output/institution.

Additional information:

<https://doi.org/10.5438/f4df-4817>



<https://commons.datacite.org/>

Aggregation: researcher (ORCID)

<https://orcid.org/0000-0001-5492-3212>

Markus Stocker

Markus Stocker is Head of the Knowledge Infrastructures Research Group at the TIB Leibniz Information Centre for Science and Technology. He holds a PhD in Environmental Informatics from the University of Eastern Finland; a MSc in Environmental Science from the University of Eastern Finland; and a Diploma (MSc) in Informatics from the University of Zurich, Switzerland. His research interests lie at the intersection between research infrastructures and research communities, and how such infrastructures acquire, maintain, and share scientific knowledge about human and natural worlds. Prior to TIB, Markus held a postdoctoral research associate position at PANGAEA, the Data Publisher for Earth & Environmental Science, at the MARUM Center for Marine Environmental Sciences, University of

Aggregated Citations, Views and Downloads

3 Citations

36 Views

Accessibility Achievements

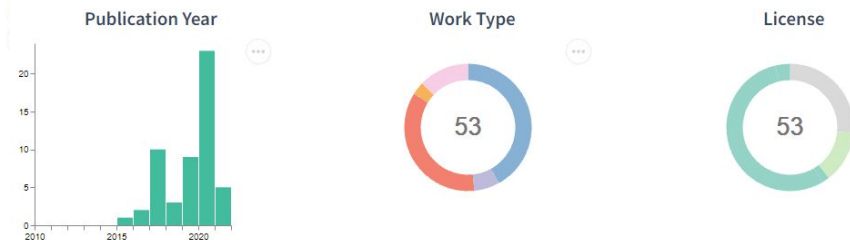


58% of the researcher's associated DOIs have metadata with rights as CC-BY, CC0 or public domain license.



Congratulations, you hit the trifecta. You have an open access paper, open dataset, and open source software.

53 Works



Other Identifiers

Scopus Author ID: [44461998000](#)

Loop profile: [1242536](#)

Registration Agency

<input type="checkbox"/> DataCite	43
<input type="checkbox"/> Crossref	10

Co-Authors ?

<input type="checkbox"/> Auer, Sören	16
<input type="checkbox"/> Magagna, Barbara	10
<input type="checkbox"/> Oelen, Allard	9
<input type="checkbox"/> Jaradeh, Mohamad Yaser	8
<input type="checkbox"/> Zhao, Zhiming	7
<input type="checkbox"/> Thijssse, Peter	6
<input type="checkbox"/> Prinz, Manuel	6
<input type="checkbox"/> Fiebig, Markus	5
<input type="checkbox"/> Jeffery, Keith	5

Aggregation: institution (ROR)

Filter Works

 ✕ 🔍

Publication Year

<input type="checkbox"/> 2021	1,710
<input type="checkbox"/> 2020	3,031
<input type="checkbox"/> 2019	3,896
<input type="checkbox"/> 2018	2,515
<input type="checkbox"/> 2017	1,901
<input type="checkbox"/> 2016	1,521
<input type="checkbox"/> 2015	878
<input type="checkbox"/> 2014	403
<input type="checkbox"/> 2013	306
<input type="checkbox"/> 2012	243
<input type="checkbox"/> 2011	212
<input type="checkbox"/> 2010	143

<https://ror.org/04xfq0f34>

RWTH Aachen University

RWTH Aachen, RWTH

Founded 1870

Links

Homepage

Wikipedia

Other Identifiers

GRID [grid.1957.a](#)

Crossref Funder ID [10.13039/501100007210](#)

ISNI [000000010728696X](#)

Wikidata [Q273263](#)

Geolocation

50° 46′ 40″ N, 6° 04′ 41″ W

Germany Education DataCite Member

<https://ror.org/04xfq0f34>

Share

Email

Twitter

Facebook

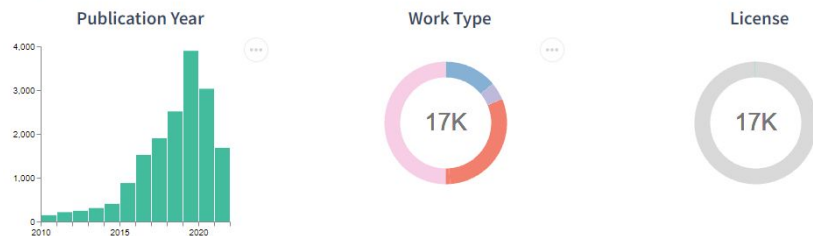
Aggregated Citations, Views and Downloads

256 Citations

1,705 Views

352 Downloads

17,138 Works



Field of Science

<input type="checkbox"/> Languages and literature	4
<input type="checkbox"/> Chemical engineering	3
<input type="checkbox"/> Computer and information sciences	1
<input type="checkbox"/> Earth and related environmental sciences	1
<input type="checkbox"/> History and archaeology	1

Registration Agency

<input type="checkbox"/> DataCite	16,874
<input type="checkbox"/> Crossref	301
<input type="checkbox"/> DataCite	1

Authors ?

<input type="checkbox"/> Stahl, Achim	961
<input type="checkbox"/> Nowack, Andreas	929
<input type="checkbox"/> Kreß, Thomas	819
<input type="checkbox"/> Pook, Tobias	806
<input type="checkbox"/> Lipinski, Martin	726
<input type="checkbox"/> Heidemann, Carsten Andreas	623
<input type="checkbox"/> Nehrkorn, Alexander Rafael Maria	609
<input type="checkbox"/> Teroerde, Marius	415
<input type="checkbox"/> Klein, Katja	401
<input type="checkbox"/> Fischer, Benjamin	251

Aggregation nach Publikation (DOI)

<https://doi.org/10.5061/dryad.5rm2f>

Data from: Environmental risk assessment for the small tortoiseshell *Aglais urticae* and a stacked Bt-maize with combined resistances against Lepidoptera and Chrysomelidae in central European agrarian landscapes

Mechthild Schuppener, Julia Mühlhause, Anne-Katrin Müller & Stefan Rauschen

Version 1 of Dataset published 2012 in *DRYAD*

The cultivation of Lepidoptera-resistant Bt-maize may affect non-target butterflies. We assessed the risk posed by event MON89034 x MON88017 (expressing Cry1A.105 and Cry2Ab2 against corn borers) to non-target Lepidoptera. Using the small tortoiseshell *Aglais urticae*, a butterfly species common in central Europe, as a test organism we (1) assessed the toxicity of Bt-maize pollen on butterfly larvae; (2) measured pollen deposition on leaves of the host plant *Urtica dioica*; (3) mapped the occurrence and distribution of host plants and larvae in two arable landscapes in Germany during maize anthesis; and (4) described the temporal occurrence of a one-year population of *A. urticae*. (1) Larvae fed 200 Bt-maize pollen grains/cm² had a reduced feeding activity. Significant differences in developmental time existed at pollen densities of 300 Bt-maize pollen grains/cm² and in survival at 400 grains/cm². (2) The highest pollen amount found was 212 grains/cm² at the field margin. Mean densities were much lower. (3) In one region over 50% of *A. urticae* nests were located within five metres of a maize field, while in the other all nests were found in more than 25 m distance to a maize field. (4) The percentage of larvae developing during maize anthesis was 19% in the study area. The amount of pollen from maize MON89034 x MON88017 found on host plants is unlikely to adversely affect a significant proportion of larvae of *A. urticae*. This paper concludes that the risk of event MON89034 x MON88017 to populations of this species is negligible.

DOI registered June 22, 2012 via DataCite.



1 Citation 131 Views 14 Downloads

Dataset English

<https://doi.org/10.5061/dryad.5rm2f>

Creators

Mechthild Schuppener
RWTH Aachen University

Julia Mühlhause
RWTH Aachen University

Anne-Katrin Müller
RWTH Aachen University

Stefan Rauschen
RWTH Aachen University

ROR

Download

Full Metadata

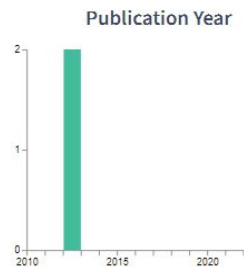
DataCite XML
DataCite JSON
Schema.org JSON-LD

Citation Metadata

Citeproc JSON
BibTeX
RIS

2 References

2 Citations



Work Type



License



Environmental risk assessment for the small tortoiseshell and a stacked Bt-maize with combined resistances against Lepidoptera and Chrysomelidae in central European agrarian landscapes

MECHTHILD SCHUPPENER, JULIA MÜHLHAUSE, ANNE-KATRIN MÜLLER & STEFAN RAUSCHEN

Journal Article published 2012 in *Molecular Ecology*

DOI registered via Crossref.

1 Citation

Journal Article

<https://doi.org/10.1111/j.1365-294x.2012.05716.x>

Aggregation: funder (ROR)

Publication Year

<input type="checkbox"/> 2021	623
<input type="checkbox"/> 2020	14,626
<input type="checkbox"/> 2019	15,303
<input type="checkbox"/> 2018	12,281
<input type="checkbox"/> 2017	6,560
<input type="checkbox"/> 2016	5,184
<input type="checkbox"/> 2015	3,411
<input type="checkbox"/> 2014	1,791
<input type="checkbox"/> 2013	778
<input type="checkbox"/> 2012	750
<input type="checkbox"/> 2011	529
<input type="checkbox"/> 2010	781

Work Type

<input type="checkbox"/> Text	3,728
<input type="checkbox"/> Dataset	44,330
<input type="checkbox"/> Collection	2,855
<input type="checkbox"/> Image	1
<input type="checkbox"/> Software	12
<input type="checkbox"/> Sound	9
<input type="checkbox"/> Audiovisual	8
<input type="checkbox"/> Physical Object	8
<input type="checkbox"/> Model	5
<input type="checkbox"/> Event	2

License

<input type="checkbox"/> CC-BY-3.0	44,532
<input type="checkbox"/> CC-BY-4.0	6,356
<input type="checkbox"/> CC-BY-NC-ND-4.0	852
<input type="checkbox"/> CC-BY-NC-4.0	442
<input type="checkbox"/> CC-BY-NC-SA-4.0	122
<input type="checkbox"/> CC-BY-NC-ND-3.0	41
<input type="checkbox"/> CC-BY-NC-SA-3.0	30
<input type="checkbox"/> CC0-1.0	29
<input type="checkbox"/> CC-BY-SA-4.0	25
<input type="checkbox"/> CC-BY-NC-3.0	10

<https://ror.org/018mejw64>

Deutsche Forschungsgemeinschaft

DFG

Founded 1951

Links

[Homepage](#)
[Wikipedia](#)
[Twitter](#)

Geolocation

50° 41' 57.9948" N, 7° 08' 51.972" W

[Germany](#) [Nonprofit](#)

<https://ror.org/018mejw64>

Share

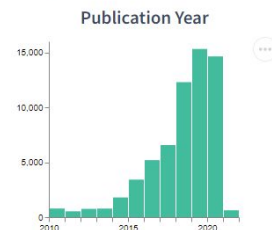
Email
 Twitter
 Facebook

Aggregated Citations, Views and Downloads

528,503 Citations

73 Views

102,451 Works



Field of Science

<input type="checkbox"/> Biological sciences	4
<input type="checkbox"/> Computer and information sciences	4
<input type="checkbox"/> Economics and business	3
<input type="checkbox"/> Agriculture, forestry and fisheries	1
<input type="checkbox"/> Chemical sciences	1
<input type="checkbox"/> Earth and related environmental sciences	1
<input type="checkbox"/> History and archaeology	1
<input type="checkbox"/> Law	1
<input type="checkbox"/> Natural sciences	1
<input type="checkbox"/> Political science	1

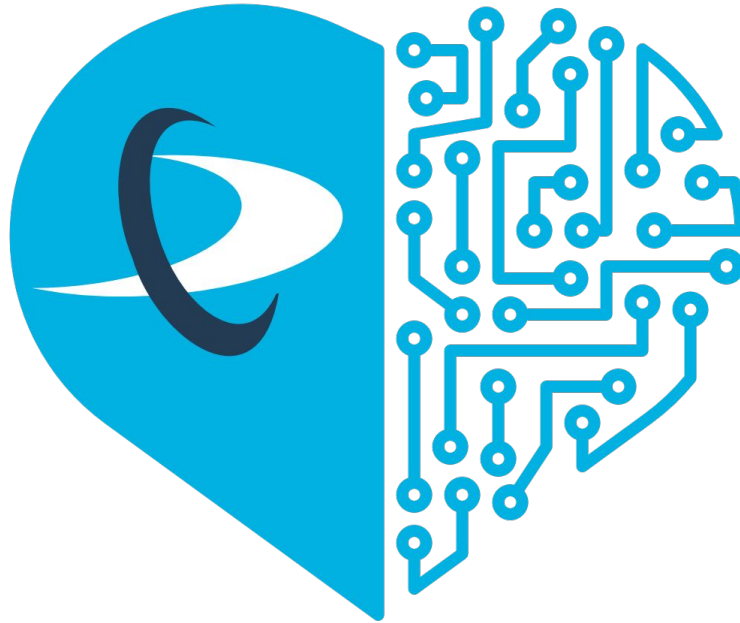
Registration Agency

<input type="checkbox"/> Crossref	53,682
<input type="checkbox"/> DataCite	48,766
<input type="checkbox"/> Crossref	3

Authors

<input type="checkbox"/> Malone, Mitchell J	2,001
<input type="checkbox"/> Acton, Gary D	1,683
<input type="checkbox"/> Lyle, Mitchell W	1,408
<input type="checkbox"/> Bohrmann, Gerhard	1,153
<input type="checkbox"/> Wefer, Gerold	893
<input type="checkbox"/> Emels, Kay-Christian	798
<input type="checkbox"/> Tiedemann, Ralf	788
<input type="checkbox"/> Hodell, David A	594
<input type="checkbox"/> Gersonde, Rainer	567
<input type="checkbox"/> Zachos, James C	535

Love your data...



...persistently



CONNECTING RESEARCH,
IDENTIFYING KNOWLEDGE



info@datacite.org



pidforum.org



datacite.org
blog.datacite.org



support@datacite.org
support@datacite.org



[@datacite](https://twitter.com/datacite)



[DataCite](https://www.youtube.com/DataCite)



[@datacite](https://www.linkedin.com/company/datacite)