

University of Stuttgart
Germany



Data sharing and publication via DaRUS in CRC 1333

Dr. Elisabeth
Rüthlein
07.12.2020

What is this about?

The CRC requires all data related to CRC 1333 publications of the funding period to be published with DOI by 30.06.2022 at the latest.

This will be crucial for the success of the next funding proposal!

→ CRC 1333 guidelines on how to publish the data

Aim:

All publication data of CRC 1333 published in this manner by 31.07.2021

We do this because...

- a) ...we said we would do so for all key-data of the funding period in our CRC proposal from 2017.
- b) ...in Jan 2020 the DFG has provided new Guidelines on “Good Scientific Practice” that say, all people that want funding from the DFG in the future need to start doing so. (End of transition period **31.07.2021**)
- c) ...managing our data in a structured manner, making it accessible and publishing it to the science community will benefit your research, the CRC 1333 and the whole science community.

DFG: Guidelines for Safeguarding Good Research Practice (from August 1st, 2019)

Guideline 13, p. 18:

*„In the interest of transparency and to enable research to be referred to and reused by others, **whenever possible researchers make the research data and principal materials on which a publication is based available in recognised archives and repositories** in accordance with the FAIR principles (Findable, Accessible, Interoperable, Reusable). Restrictions may apply to public availability in the case of patent applications. If self-developed research software is to be made available to third parties, an appropriate licence is provided.*

*As of 1 August 2019, all [...] research institutions will be required to implement the 19 guidelines in a legally binding manner to be eligible to receive DFG funding. [...] there is a two-year transition period for implementing the guidelines in the Code. **This period** begins on 1 August 2019 and **ends on 31 July 2021.***

Research Data

Citation from the CRC 1333 Funding Proposal 2017:

*For the whole duration of the CRC, the respective **data of promising findings will be entered and organized on a suitable platform** (including key parameters such as development and anchoring of the catalyst, catalyst turn over numbers, characterization data, simulation results...). [...]*

To this end, a suitable database will be installed, which will enable swift access to relevant information.

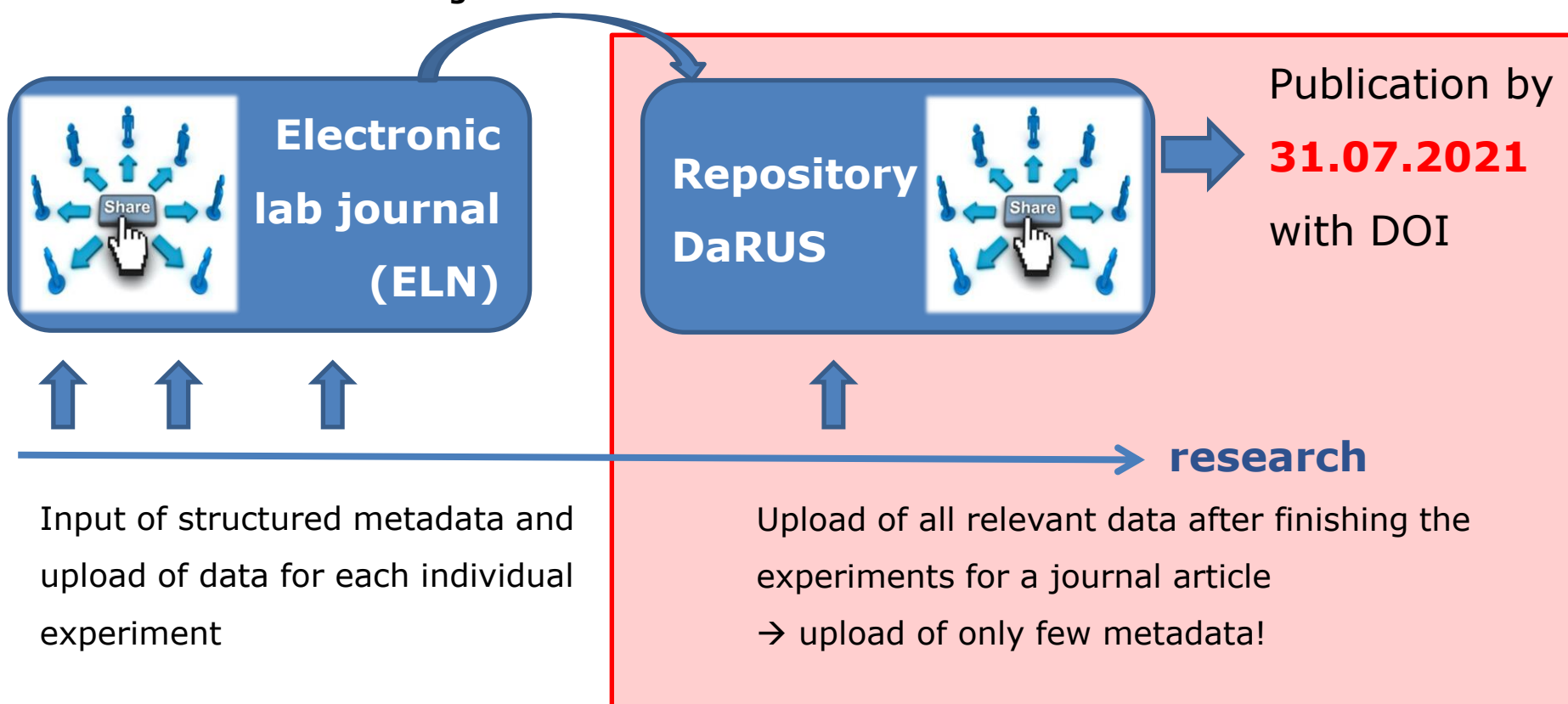
After the first funding period we will publish these data on a suitable platform (complete with DOI labelling).

RDM Concept – CRC 1333

In principle, there are two entry points for organizing our research data on a collaboration platform

Publication in Chemotion Repository with DOI

→ **Linking DOI to Dataset in DaRUS**





What is DaRUS?



Research-Data Repository of University of Stuttgart

Based on Open-Source-Software: „Dataverse“

Developed by Harvard University

Usable for sharing, citing, analyzing and archiving of datasets from all scientific disciplines according to FAIR Data principles



- Organisation in **virtual archives, called Dataverses**
- Management and sharing of published and unpublished data
- Permanent Identifiers (DOI) for all published datasets
- Support of different discipline-specific metadata schemes (no Chemistry metadata available, but support of molecular structures as SMILES or INCHI)
- APIs for Integration of external Software, e.g. to upload data from an ELN, from metadata files of analytical machines
- Data can be found via:



Comments to DaRUS

It's generic!

- Upload easy, usefulness restricted!
- Use other specialized repositories if available and accepted in the community, e.g. Cambridge Structural Data Base

Do we still want the data in DaRUS?

Yes, please!

Link to the dataset in DaRUS via metadata field „related data“.



Time Line

Workshop

07.12.2020

How should the data be uploaded in DaRUS?

Input Welcome!

Status Quo 12/2020

Example entry in DaRUS available,
Data files located non-accessible (neither publicly nor within the CRC) with each researcher/institute

Data Curation Process:

Content check by PIs of contributing projects

Deadline: 30.04.2020

Formal curation by CRC and DaRUS team

Data Upload of all publication-relevant data by the authors

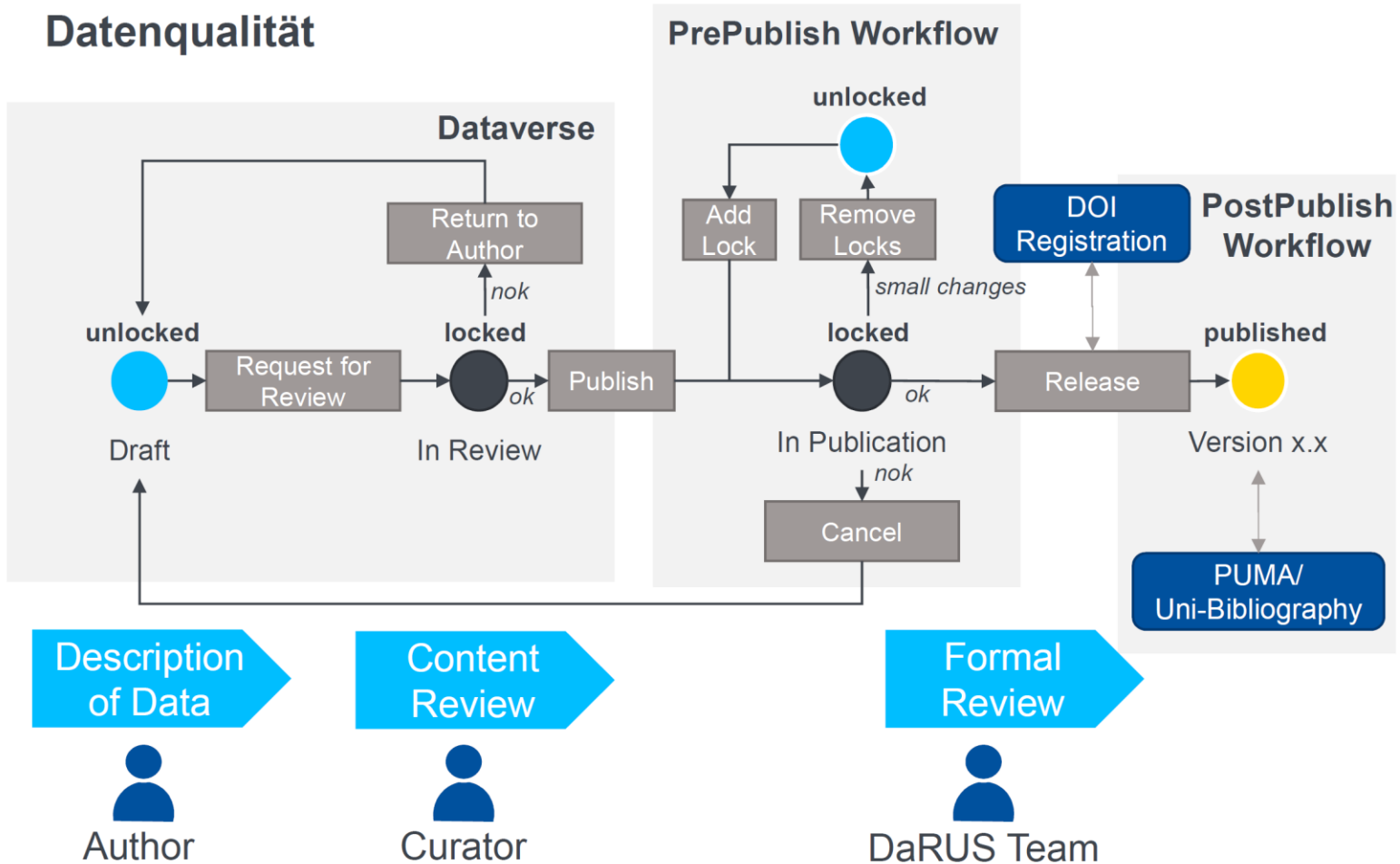
Deadline: 28.02.2021

From 31.07.2021 at the latest

- All analytic data of CRC publications can be publicly accessed via DaRUS
- **Data of new CRC publications should be uploaded asap upon acceptance (or before as data can then be made accessible to reviewers by private URL)**

Please announce beforehand if delays are unavoidable for your data!!

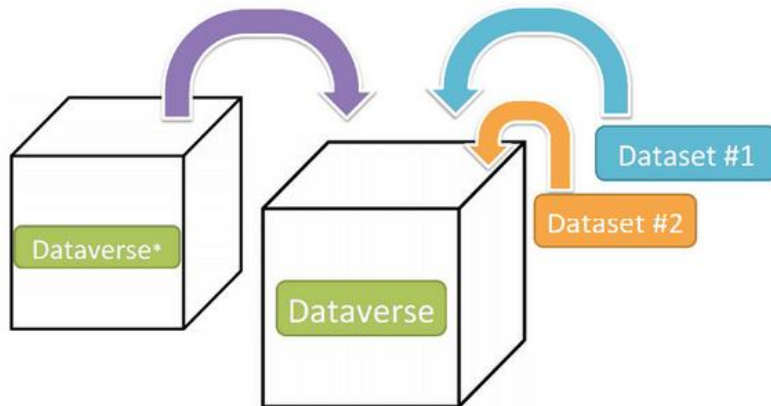
Data Curation Process



Checklists and more information: <https://www.izus.uni-stuttgart.de/en/fokus/darus/publication/>

Dataverses, Datasets, Data files...

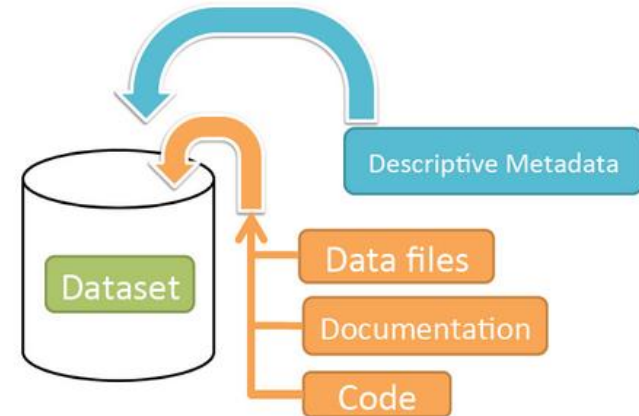
Schematic Diagram of a **Dataverse** in Dataverse 4.0



Container for your **Datasets** and/or **Dataverses***

* Dataverses can now contain other Dataverses (this replaces Collections & Subnetworks)

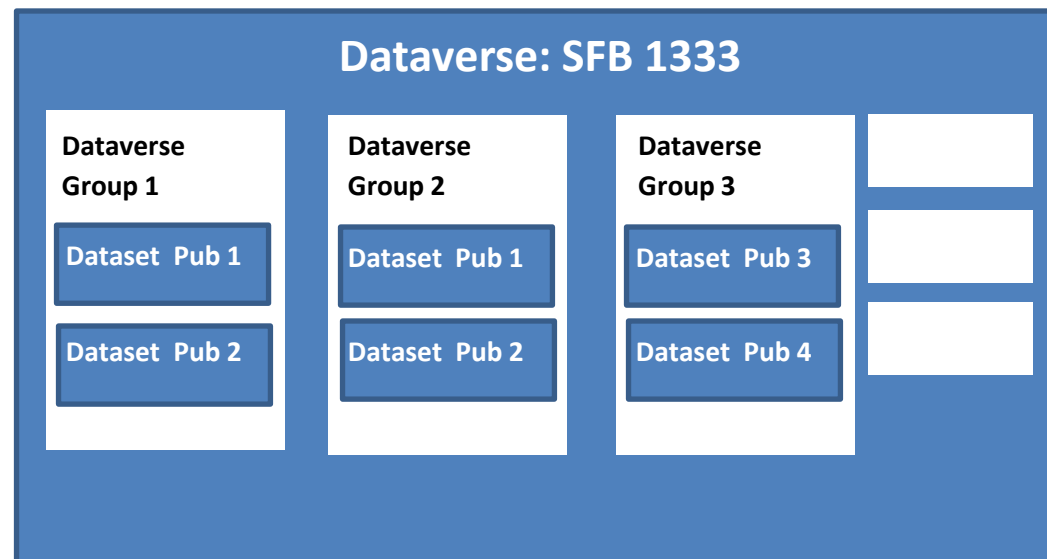
Schematic Diagram of a **Dataset** in Dataverse 4.0



Container for your data, documentation, and code.

- A Dataverse contains other Dataverses or Datasets (Container defining the **rights management** and **minimum required metadata**)
- A Dataset includes metadata and one or several files, documentation and code

How is the SFB 1333 dataverse organized?



- CRC 1333 dataverse Admin creates dataverse for each group
- Assigns rights to authors to add and change datasets
- Assigns rights to PIs to curate and publish the data

Considerations

- A DOI (persistent Identifier) can be assigned for each dataset and for each file
- Dataverses do NOT get a DOI

Possibility 1:

Create one single dataset with one set of metadata and include all your data files from the publication

- You will get one DOI/publication
- Additional DOIs for single files can be assigned

Possibility 2:

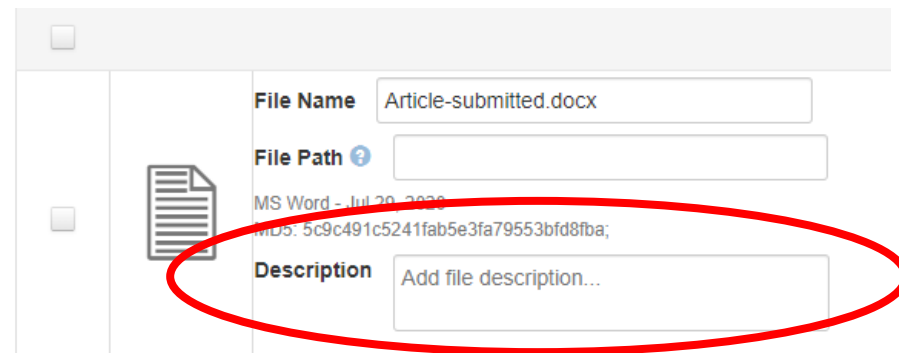
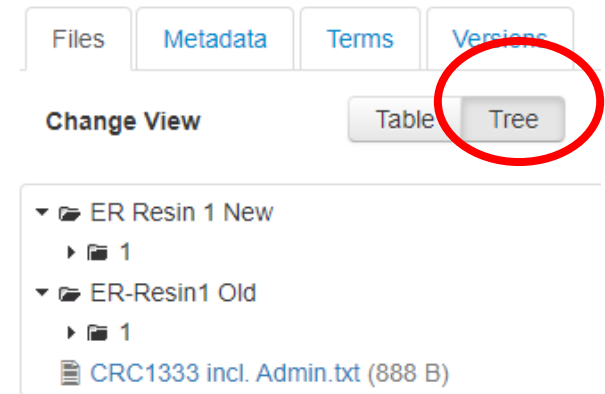
Add several datasets (e.g. one per analytical method) with different metadata which will make finding and reusing the data easier.

- Several DOIs (One DOI/ dataset)

How to organize data files in a dataset?



- If you upload a vast amount of data files in one dataset: Please upload as *.zip with sensible folder structure
 - Files will be unzipped upon upload and can then be represented as a „Tree“
- Name folders by the structure of your publication and also name files meaningfully
- Description of files optionally



Which data?

All data related to CRC 1333 publications	
<ul style="list-style-type: none"> • Output-files of analytical measurements • Input data and results of calculations • Used code and software 	Yes, whenever possible
Data behind figures (edited data, excel sheets)	Yes, please
Which formats?	<ol style="list-style-type: none"> 1) Open data formats whenever possible (e.g. JCAMP-DX) 2) Text-based formats (e.g. *.csv) 3) Then whatever is customary in your community
PDF and Suppl. Inf. of the research article (also subm. +acc. Version) Figures from the paper	Yes! Please add as separate DATASET esp. if non-accessible from Uni Stuttgart Pot. mark as „NOT FOR PUBLICATION“

Required Metadata

Citation Metadata ^	
Dataset Persistent ID ?	doi:10.18419/darus-832
Title ?	Olefin Metathesis in Confined Geometries: A Biomimetic Approach toward Selective Macrocyclization
Author ?	Felix Ziegler (Universität Stuttgart) - ORCID:
Contact ?	Use email button above to contact. Michael R. Buchmeiser (Universität Stuttgart)
Description ?	All primary data files and processed data of the journal article. (2020-07-29)
Subject ?	Chemistry
Keyword ?	Catalysis, Confinement, Metathesis
Related Publication ?	Felix Ziegler, Johannes Teske, Iris Elser, Michael Dyballa, Wolfgang Frey, Hamzeh Kraus, Niels Hansen, Julia Rybka, Ulrich Tallarek, and Michael R. Buchmeiser Journal of the American Chemical Society 2019 141 (48), 19014-19022 DOI: 10.1021/jacs.9b08776 doi: 10.1021/jacs.9b08776 https://pubs.acs.org/doi/pdf/10.1021/jacs.9b08776
Grant Information ?	DFG: 358283783 – SFB 1333
Project ?	CRC1333
Depositor ?	Felix Ziegler
Deposit Date ?	2020-07-29
Kind of Data ?	experimental data
Related Datasets ?	https://www.ccdc.cam.ac.uk/structures/search?id=doi:10.5517/ccdc.csd.cc23699y&sid=DataCite

→ „Required fields“ chosen for the CRC 1333-dataverse

Using the „metadata-template“ function of a dataverse you can keep entries for every new dataset.

Optional Metadata

- Which metadata do people search for if they are interested in your data?
- Which metadata do people need to understand your data?

Engineering Metadata ^	
Data Generation ?	Simulation; Analysis; Experiment; Prediction
Measured Variables ?	Selectivity for macrocyclization
Controlled Variables ?	Specification of controlled (observed / independent) variables. Pore Size
System or Phase Components ?	(InChICode: InChI=1S/C18H28N2O3Si.C10H12O.2ClH.Ru/c1-15-12-16(2)18(17(3)13-15)20-10-9-19(14-20)8-7-11-24(21-4,22-5)23-6;1-8(2)11-10-7-5-4-6-9(10)3;;;/h9-10,12-13H,7-8,11H2,1-6H3;3-8H,1-2H3;2*1H;/q;;;+2/p-2) (SmilesCode: >>CC1=CC(C)=CC(C)=C1N2C=CN(CCC[Si](OC)(OC)OC)[C]2.CC(C)OC(C=CC=C3)=C3/C=[Ru](Cl)/Cl)

If you actually want people to find, understand and potentially reuse your data without relation to your publication:

- Please make a new dataset with more extensive metadata
- Or better: Put it in a discipline-specific repository and link the data to your DaRUS entry

What should I do now?

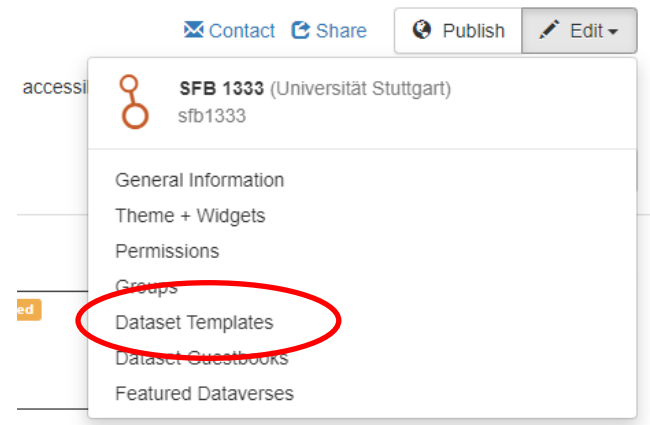
All: Please sign in to DaRUS once via <https://darus.uni-stuttgart.de/>

→ Then I can give you rights to access the data within the dataverse „SFB 1333“

→ Please check if there is a dataverse for your group – if not request one with the dateverse admin (me ;-))

First authors of CRC 1333 publications:

- Add a metadata template with the metadata of your publication to the SFB 1333 dataverse and save it
- Write an email to your fellow authors: Ask them to upload their data as new datasets (in their group's dataverse) with (slightly) modified metadata.



What should I do now?

All: Please sign in to DaRUS once via <https://darus.uni-stuttgart.de/>

→ Then I can give you rights to access the data within the dataverse „SFB 1333“

→ Please check if there is a dataverse for your group – if not request one with the dateverse admin (me ;-))

First authors of CRC 1333 publications:

- Add a metadata template with the metadata of your publication to the SFB 1333 dataverse and save it
- Write an email to your fellow authors: Ask them to upload their data as new datasets (in their group's dataverse) with (slightly) modified metadata.

All authors contributing data to CRC publications:

- Add a dataset to your group's dataverse using the metadata template, add your data as zip-archive with sensible folder structure
- click „Request for review“ when the data is complete, sufficiently annotated by metadata and sensibly structured, to your belief.

Please complete this task by 28.02.2021

What should I do now?

CRC 1333 PIs:

- Make sure your data is being uploaded to DaRUS in time by your coworkers
- Curate the datasets from your working group by April 30, 2021 and contact your coworkers about any adjustments or missing data
- Click „Publish“ as soon as the data is complete, sufficiently annotated by metadata and sensibly structured


The DaRUS team will formally curate the dataset and will potentially get back to you before it actually gets published.

PIs will receive requests for review for all datasets within their group's dataverse before publication.

Important: Every PI is still responsible for a safe storage of their data for 10 years according to the rules of good scientific praxis!

Example Entry

Please find our example entry [here](#) (Private URL)


Olefin Metathesis in Confined Geometries: A Biomimetic Approach toward Selective Macrocyclization

Draft
Unpublished

Felix Ziegler, 2020, "Olefin Metathesis in Confined Geometries: A Biomimetic Approach toward Selective Macrocyclization", <https://doi.org/10.18419/darus-832>, DaRUS, DRAFT VERSION

[Cite Dataset](#) [Learn about Data Citation Standards.](#)

Dataset Metrics
 0 Downloads

Subject Chemistry

Keyword Catalysis, Confinement, Metathesis

Related Publication Felix Ziegler, Johannes Teske, Iris Elser, Michael Dyballa, Wolfgang Frey, Hamzeh Kraus, Niels Hansen, Julia Rybka, Ulrich Tallarek, and Michael R. Buchmeiser Journal of the American Chemical Society 2019 141 (48), 19014-19022. doi: [10.1021/jacs.9b08776](https://doi.org/10.1021/jacs.9b08776)

Data Generation Simulation; Analysis; Experiment; Prediction

Files Metadata Terms Versions

Change View Table Tree

- Data Buchmeiser group
 - Analytics
 - BET
 - 3@SBA-15_50A.qps (32.8 KB)
 - 3@SBA-15_50A_DMDMS.qps (32.8 KB)
 - 3@SBA-15_62A.qps (32.8 KB)
 - 3@SBA-15_62A_DMDMS.qps (32.8 KB)

Example Entry

DaRUS > SFB 1333

[Contact](#) [Share](#)

[Publish](#)

[Edit](#)

All primary data files of published research from CRC 1333, e.g. from journal articles. All data will be made accessible to the general public by the end of the funding period on 30.June 2022.

[Find](#)

[Advanced Search](#)

[+ Add Data](#)

☐ [Dataverses \(6\)](#)

☒ [Datasets \(7\)](#)

☐ [Files \(248\)](#)

Publication Year

2019 (1)

Publication Status

Draft (6)

Unpublished (6)

Published (1)

Author Name

Felix Ziegler (3)

Kästner, Johannes (2)

Kraus, Hamzeh (1)

Michael Dyballa (1)

Rybka, Julia (1)

Subject

Chemistry (7)

Keyword Term

Catalysis, Confinement, Metathesis (4)

Project Name

Project C4 (2)

CRC1333 - project B2 (1)

1 to 7 of 7 Results

[Sort](#)

Olefin Metathesis in Confined Geometries: A Biomimetic Approach toward Selective Macrocyclization Draft Unpublished

Dec 2, 2020 - SFB 1333 - Hansen group, ITT



Felix Ziegler, 2020, "Olefin Metathesis in Confined Geometries: A Biomimetic Approach toward Selective Macrocyclization", <https://doi.org/10.18419/darus-1168>, DaRUS, DRAFT VERSION

Primary data files and processed data of the journal article from the Hansen group.

Olefin Metathesis in Confined Geometries: A Biomimetic Approach toward Selective Macrocyclization Draft Unpublished

Dec 2, 2020 - SFB 1333 - Buchmeiser group, IPOC-MSF



Felix Ziegler, 2020, "Olefin Metathesis in Confined Geometries: A Biomimetic Approach toward Selective Macrocyclization", <https://doi.org/10.18419/darus-1169>, DaRUS, DRAFT VERSION

Article PDF, submitted and accepted version Figures and ChemDraw files with structures as published. - For publication as we have Open Access rights.

Olefin Metathesis in Confined Geometries: A Biomimetic Approach toward Selective Macrocyclization Draft Unpublished

Dec 2, 2020 - SFB 1333 - Dyballa group, ITC



Michael Dyballa, 2020, "Olefin Metathesis in Confined Geometries: A Biomimetic Approach toward Selective Macrocyclization", <https://doi.org/10.18419/darus-1166>, DaRUS, DRAFT VERSION

Primary data files and processed data of the journal article from the Dyballa group.

Olefin Metathesis in Confined Geometries: A Biomimetic Approach toward Selective Macrocyclization Draft Unpublished

Jul 29, 2020 - SFB 1333 - Buchmeiser group, IPOC-MSF



Felix Ziegler, 2020, "Olefin Metathesis in Confined Geometries: A Biomimetic Approach toward Selective Macrocyclization", <https://doi.org/10.18419/darus-832>, DaRUS, DRAFT VERSION

All primary data files and processed data of the journal article from the Buchmeiser group and links to related data sets from other groups. - For

Getting Help

In that Order, please 😊

- Dataverse User Guide: <http://guides.dataverse.org/en/latest/user/index.html>
- Dataverse API Guide: <http://guides.dataverse.org/en/latest/api/index.html>
- DaRUS FAQ: <https://www.izus.uni-stuttgart.de/fokus/darus/>
- Is it a bug? <https://github.com/TIK-NFL>
- Ask your local admin:
It's the one with the role „DaRUS Admin“ in the Identity Management
- Ask the DaRUS-User-Mailinglist: darus@listserv.uni-stuttgart.de
- Ask the FoKUS-Team: fokus@izus.uni-stuttgart.de

Questions?
Comments?
Suggestions?

Useful tips & tricks

- Upload of large files
- Upload via Web interface works well for files up to 2 GB
- for larger files please use „Upload via API“ or ask for „Direct Upload“
<https://www.izus.uni-stuttgart.de/en/fokus/darus/bigdata/>
- Upload of data >100 GB - please announce beforehand to FoKUS team

- If you experience issues with zip archives (might happen if many subdirectories are named equally within the same dataset):

E.g.: The following files are duplicates of (an) already uploaded file(s): format.ased, format.temp, uxnmr.par(skipping)

1) „Double zip“ (e.g. zip each NMR, then zip all the data for the dataset including the NMRs, then upload the whole dataset)

2) Create tar archives instead of zip archives

- Install WinRAR for transforming many Files to individual Zip-archives at the same time.

Update to
DaRUS 5.2
should solve
theses issues
with Zip
archives

Data already elsewhere published?



- What if we have already uploaded the data to our own institute's dataverse within DaRUS?
 - That's ok!
 - Please provide me with the link, then I can link it to the „SFB 1333“ dataverse as well.
- What if I have published the data already in a different (discipline specific) data repository.
 - Also not a problem.
 - Please provide the link to the first author of the publication, so he/she can add the link to the metadata field „Related data“ in the DaRUS Dataset