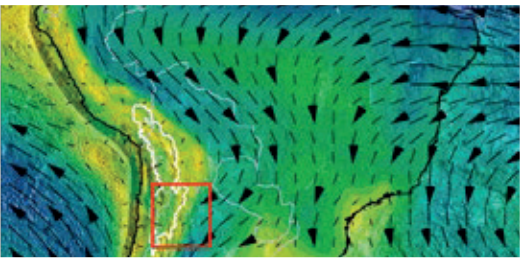
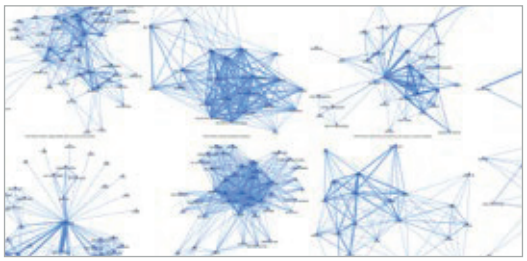
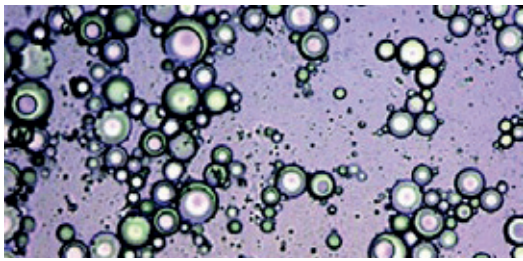
University of Potsdam

Research Data Policy and Recommendations for the Handling of Research Data







Potsdam University Press

University of Potsdam

Research Data Policy

Recommendations for the Handling of Research Data

Version: Oktober 2019

Potsdam University Press 2020
http://verlag.ub.uni-potsdam.de/

Am Neuen Palais 10, 14469 Potsdam
Tel.: +49 (0)331 977 2533 / Fax: 2292
E-Mail: verlag@uni-potsdam.de

>>

Contact:
If you have any questions about the Research Data Policy, the recommendations, or about research data management (RDM), please contact the joint research data team of the University Library and the Computer and Media Center (ZIM). Further information can be found on the website of the University of Potsdam.

E-Mail: forschungsdaten@uni-potsdam.de
https://www.uni-potsdam.de/forschungsdaten.

Cover photos:
Network graphs of German-language dramas from DLINA Corpus (1731–1929). Fischer et al., Figshare (2015).
Micrographs of Janus droplets. Raju et al., RSC Advances 9 (2019), 19271, fig. 5.
Crystal structure of erioflorin. Paz et al., Acta Crystallographica E 73 (2017), 334, fig. 1.
850-hPa austral summer months mean wind speed over South America. Ziarani et al., Atmosphere 10 (2019), 379, fig. 1.
Composition and layout: Kristin Schettler
Print: SAXOPRINT GmbH

Also published online on the publication server of the University of Potsdam:
https://doi.org/10.25932/publishup-44438

Contents


Universität
Potsdam

Research Data Policy of the University of Potsdam

Non-official translation of the policy ratified by Senate on September 25, 2019 and published in *Amtliche Bekanntmachungen "Official Notices" Nr. 18 of September 30, 2019, pp. 1361–1362.* The information regarding the novel DFG Code of Conduct in footnote 3 has been added editorially.

The University of Potsdam, as Brandenburg's largest university, is oriented towards distinctive interdisciplinary research profiles in cooperation with a large number of non-university research institutions in the natural and human sciences, the humanities, and the social sciences. In many of these areas, data are an essential basis of scientific knowledge; since 2019, *data-centric sciences* have been designated as one of four focus areas for research at the University of Potsdam.

Research data can be the foundation for further research as well as the transfer of knowledge and technology, including for purposes beyond those for which the data were collected. Therefore, publishing and archiving research data serve not only for the replicability and verification of earlier results but also for the finding of new results. Recent developments in good scientific practice have given rise to new requirements regarding the handling of research data. Moreover, the digital transformation in academia confronts researchers and research support staff with challenges regarding capabilities and infrastructural prerequisites for research data management. Making data publishing and research data management a self-evident part of research practice will require cultural changes.

This policy governs the handling of data that is generated, reused, or processed as part of the research methods for research projects. This includes the methods and test procedures available as data, such as questionnaires or software. It also serves to implement the "Principles for Handling of Research Data" by the Alliance of Science Organisations in Germany[1] and the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" signed by the University of Potsdam in 2015[2].

This policy applies to all researchers and research support staff at the University of Potsdam. The University of Potsdam will publish recommendations on how researchers and research support staff can comply with the requirements of this policy.

I. Principles

(1) The University of Potsdam is committed to free access to research data. Research data should be made available for reuse as openly as possible, unless interests worthy of protection predominate or legal requirements prevent their publication. Research data that constitute the basis of published results should be made accessible promptly and linked to the publication appropriately. Research data with high reuse potential should be published in a quality-assured manner regardless of their use for a text publication. In order to improve reproducibility and to acknowledge underlying academic achievements, reused data and research software should be cited.

1 Alliance of Science Organisations in Germany. Principles for handling of research data (2010). http://doi.org/10.2312/ALLIANZOA.019.

2 Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003). https://openaccess.mpg.de/Berliner-Erklaerung.

(2) **The University of Potsdam ensures the archiving of research data.** Research data that constitute the basis of a text publication but are not themselves published are to be archived for at least 10 years using a suitable infrastructure. Storage of research data exclusively by the researchers themselves or the immediate deletion of the data are permissible only if interests worthy of protection predominate or legal requirements stipulate this, or in case no suitable infrastructure for archiving these data is available.

(3) **The University of Potsdam recognizes the preparation of research data for reuse and the development of reusable research software as academic achievements.** The same applies to contributions to the advancement of discipline-specific good practice in the handling of research data and research software. The research output of the University of Potsdam includes quality-assured data and software publications that are recognized by the respective scientific community.

(4) **All researchers are responsible for complying with good scientific practice regarding the handling of research data.** The relevant discipline-specific guidelines by DFG review boards and academic associations on the handling of research data are to be taken into account. The following people bear special responsibility:

1. Principle investigators, in particular with regard to a project's documentation, openness, and reusability of research data as well as data protection.
2. Research group leaders with regard to cross-project aspects of data handling (e.g., using common standards for the quality and storage of comparable data from several studies).

(5) Research students and early career researchers are entitled to appropriate information, qualification, and support from teaching staff and supervisors. In disciplines working with data, basics of the practical handling of research data should be taught as comprehensive course content, starting at the undergraduate level. Opportunities for postgraduate training should be created for young scientists, lecturers, and supervisors.

II. Legal Framework

(1) Legally binding regulations on the handling of research data remain unaffected and take precedence over this policy.

(2) Examples of legal and university regulations which result in rights and obligations with regard to research data are:

1. *Regeln zur Sicherung guter wissenschaftlicher Praxis an der Universität Potsdam* "Rules for Good Scientific Practice at the University of Potsdam",[3]
2. *Verfahrensordnung der Ethikkommission der Universität Potsdam* "Rules of Procedure of the Ethics Committee of the University of Potsdam",[4]
3. fundamental rights, in particular privacy rights, including data protection,[5]

3 University of Potsdam. Self-Regulation in Science – Rules for Good Scientific Practice at the University of Potsdam/ Universität Potsdam. Selbstkontrolle in der Wissenschaft – Regeln zur Sicherung guter wissenschaftlicher Praxis an der Universität Potsdam. Amtliche Bekanntmachungen der Universität Potsdam 2 (2002), S. 17–21. https://digital.ub.uni-potsdam.de/periodical/pageview/261536.
Please take note of the novel DFG Code of Conduct Guidelines for safeguarding Good Research Practice of August 2019 (available in English at https://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/index.html) and the pending alignment of regulations at the University of Potsdam. The „Recommendations for the Handling of Research Data at the University of Potsdam" of October 2019 already meet the requirements of the Code of Conduct in the areas of research data and research software.

4 Rules of Procedure of the Ethics Commission of the University of Potsdam/Verfahrensordnung der Ethik-Kommission der Universität Potsdam (2016). https://www.uni-potsdam.de/am-up/2016/ambek-2016-15-1443-1446.pdf.

5 In particular, the provisions of the EU General Data Protection Regulation (GDPR) and the Brandenburg Data Protection Act/Brandenburgisches Datenschutzgesetzes (BbgDSG) must be observed.

4. intellectual property rights (copyright and related rights as well as patent law and related commercial rights, including regulations regarding employee inventions).

(3) Further binding regulations may result from agreements, e.g., in grant, cooperation, and license contracts. In particular, the requirements of funding bodies for the management, publishing, and archiving of research data must be met. Any restrictions on the openness and reusability of research data through contracts and other agreements that contravene the principles of this policy are to be avoided.

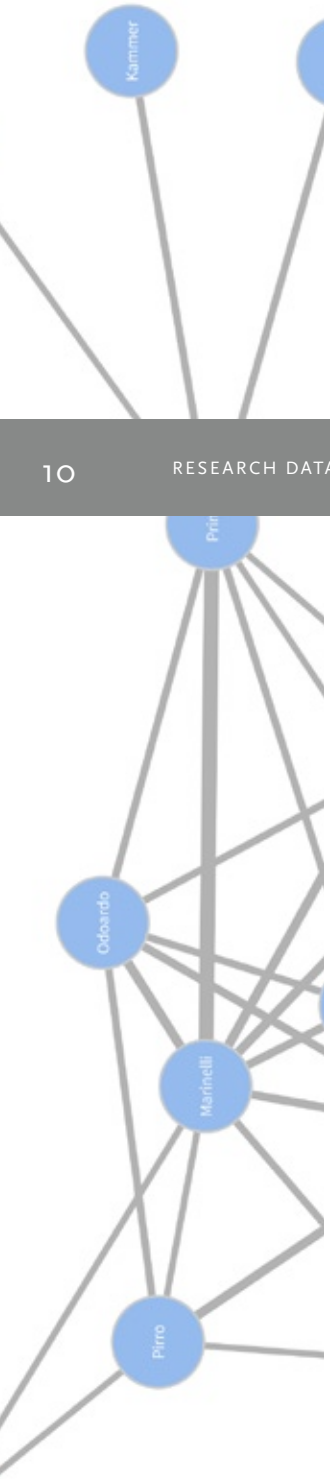
III. Institutional Responsibility

(1) As part of its research data strategy, the University of Potsdam commits itself to create the technical and organizational preconditions that enable compliance with this policy.

(2) The University of Potsdam initiates a cross-faculty discussion and development process on open science and the digital transformation in research, as well as on data handling competence training in teaching for degree courses and in postgraduate training.

(3) The University of Potsdam supports declarations and initiatives at the state, federal, European, and international levels that promote free access to research data, the recognition of data and software publications as research output, good practice in research data management, and good conditions for data-intensive research.

(4) The activities of the University of Potsdam in the field of research data take gender and diversity aspects into account.



IV. Validity

(1) This policy shall enter into force after the day of its publication in *Amtliche Bekanntmachungen* "Official Notices" of the University of Potsdam.

(2) It will be reviewed together with the associated research data strategy of the University of Potsdam when necessary, but no later than in 2023.

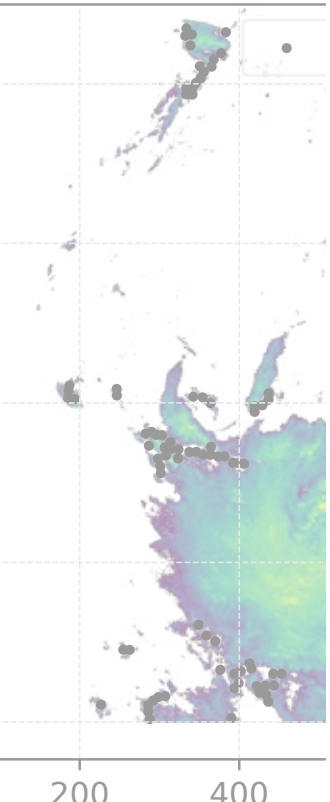
Recommendations for the Handling of Research Data at the University of Potsdam

Non-official translation of the recommendations adopted by Senate's Commission for Research and Junior Academics (FNK) on June 26, 2019, duly noted by Senate on September 25, 2019, with changes adopted by FNK on October 9, 2019 to match the requirements of the novel DFG Code of Conduct *Guidelines for Safeguarding Good Research Practice*.

These recommendations specify and complement the Research Data Policy of University of Potsdam. They are aimed at all researchers and research support staff.

I. Working with Research Data

The requirements of good scientific practice regarding honest, methodologically correct, thorough, reliable, and replicable work through good documentation also apply to the handling of research data. Principle overarching elements of secure and well-documented work with research data are listed below. For the planning and structured documentation of these elements it is recommended to use data management plans (DMP) as instruments for project and quality management and to update them regularly. DMP can be created at the level of a project, a data-intensive research instrument, or a working group, as needed. In the case of projects which pose an exceptional burden on centrally-operated infrastructure, central facilities may require the submission of a DMP.



1. **Purpose of the data.** It is recommended to determine at an early stage which data will be published or archived and which data will continue to be stored by the researchers themselves. It is useful to define holding periods. Data that are no longer needed, nor worth publishing or archiving, should be deleted regularly. In particular, personal data are subject to the principle of storage limitation: storage is only permitted as long as strictly necessary for the respective purpose. Therefore, research data that contains personally identifyable information must be anonymized as soon as possible according to the research purpose.[1]

2. **Intellectual property.** Research data usually is not covered by copyright or related intellectual property rights, but in certain cases multiple rights may accrue to different persons. Ownership and usage rights of research data are therefore often unclear, which can limit its reuse. For each project with multiple participants it is therefore recommended to document at an early stage which rights may have accrued to whom, the intention to publish the data, and the reciprocal granting of the relevant usage rights.

3. **Secure storage.** Data loss is to be prevented through the use of suitable storage services or of storage media and appropriate backups. The use of storage services run by the scientific community is recommended; the use of local storage media and of commercial storage services as a private customer is discouraged. In many cases version-control is helpful. The necessary level of data security is to be ensured by appropriate technical and organizational measures, e.g. effective access restrictions and pseudonymization of personal data that cannot be fully anonymized.

1 To this end, all directly identifying characteristics must be removed, or the pseudonymization key that allows linking the research data to personally identifying information must be destroyed. Further measures may be necessary. Data is only considered anonymous if "any particular information about personal or factual circumstances can no longer be assigned to an identified or identifiable person, or can only be assigned with a disproportionate expenditure of time, cost and labour" (§ 3 BbgDSG).

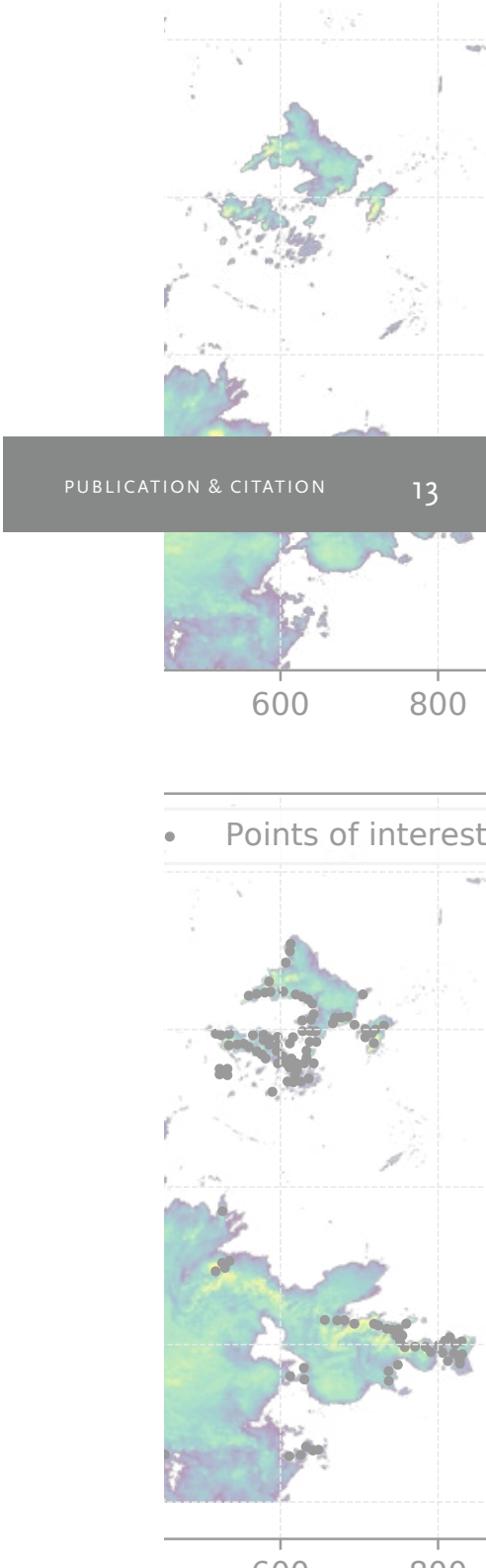
4. **Documentation and use of standards.** To make data reusable it is necessary to document the context in which the data was created and the tools used for data processing. For the sake of interoperability and long-term readability, the use of appropriate file formats is recommended with preference given to formats using open standards. To further reproducibility it is recommended to define and document conventions for file names and folder hierarchies early on; to use suitable, and when possible subject-specific metadata standards; and to collect relevant metadata already during the research process.

II. Publication and Citation of Research Data

The rules of good scientific practice for publishing also apply to data and software publications. In particular, it is prohibited to withhold data that does not support the authors' hypotheses, to fragment data and software publications with the aim of increasing the number of publications, or to duplicate publications without disclosing the previous publication. The following points provide further guidance regarding the publication and citation of research data.

1. **Place of publication.** For the publication of research data, well-established discipline- or data-type-specific databases, repositories, and data centers should be preferred. Infrastructures should be used in which data are kept and can be referenced independently; research data should not exclusively be provided with the text publication as supplementary material.

2. **Preparation and availability of data.** Research data should be made accessible in a processing stage (raw or structured data) that enables meaningful reuse by third parties. It is recommended to strive for consistent adherence to the FAIR data principles (findable, accessible, interoperable, and reusable) when preparing data for publication and selecting



the place of publication.[2] The four principles articulate the following basic requirements:

a. *Findability*: The data is sufficiently described with relevant metadata and referenced by a unique persistent identifier (e.g., a DOI).
b. *Accessibility:* The data is readable by both humans and machines and is stored in a trusted repository.
c. *Interoperability*: Data and metadata use a formal, accessible, shared, and broadly applicable vocabulary for knowledge representation.
d. *Reusability*: The data is unambiguously licensed, contains correct provenance information, and is well documented.[3]

Self-programmed research software should be made publicly available as source code. Source code of published software should be persistent, citable, and documented.[4]

3. **Authorship.** A person who makes a genuine, accountable contribution to the content of academic data or software publication is an author. Such a contribution is established especially through substantial research-based involvement in the creation, collection, acquisition, or provision of the data, software, or sources.[5]

2 Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, u.a. "The FAIR Guiding Principles for scientific data management and stewardship". Scientific Data 3 (2016). https://doi.org/10.1038/sdata.2016.18.
The current version of the FAIR principles can be accessed at https://www.go-fair.org/fair-principles.

3 League of European Research Libraries. "Implementing FAIR Data Principles". Factsheet (n.d.[2017]). https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf.

4 Version-control software and repositories commonly used in software development do not usually meet these requirements, as they neither guarantee continuity nor offer persistent identifiers (such as DOIs). The releases to be cited therefore should also be submitted to a suitable research data repository. GitHub offers an easy-to-use interface to Zenodo: https://guides.github.com/activities/citable-code.

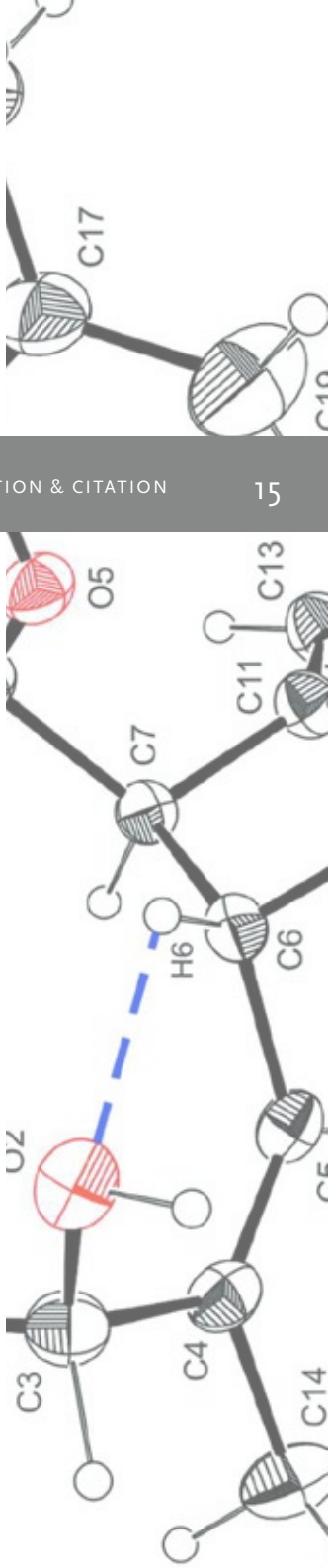
5 Criteria for whether a contribution is genuine, accountable, substantial, and research-based may vary by subject area. Within the framework of the discipline-specific publication culture, it is possible that the authors of a text publication differ from the authors of the publications of the data and software on which the text publication is based.

4. **Free licensing and open access.** Research data and scientific software should be made available under established, standardized and, as far as possible, free and open licenses. Access conditions and, if applicable, embargo periods should adhere to the principle of "as open as possible, as closed as necessary". The following points provide specific guidance for licensing:

   a. *Licensing and attribution:* The obligation to attribute academic achievements is a principle of good scientific practice. Licenses and waivers for data and software that do not contractually require the attribution of the authors do not supersede this obligation. The preferred way to promote data and software citation is not through restrictive licensing but by including a recommended citation.

   b. *Public research data:* Creative Commons tools have become the most common way to license research data that is made publicly available. Data that is free of copyright and related rights should be labelled as such upon release using the "Creative Commons Zero" (CC0) waiver. For data that is protected by copyright or related rights, using the CC0 waiver is also recommended because it is the best way to ensure reusability. The license "Creative Commons Attribution" (CC BY) can also be considered, but due to the strict formal requirements regarding attribution and the provision of license information it already limits subsequent use. Creative Commons licenses with terms that go beyond the "Attribution" element are not suitable for licensing research data ("Non-Commercial", NC; "No Derivative Works", ND), or are suitable only in certain cases and only to a limited extent ("Sharing under Equal Conditions", SA).[6]

6 More information about Creative Commons is available at https://creativecommons.org.



c. *Research data for "Scientific Use" only:* Research data to which access has to be restricted should be licensed in consultation with the appropriate data center using the licenses recommended by them.

d. *Software:* Creative Commons licenses (such as CC BY) are not suitable for software. A CC0 waiver can be granted for completely self-written code (to which the programmers own all rights). This is especially useful for short scripts. Free software licenses should be used for more complex products. Examples of common free software licenses are the MIT license and the GNU GPLv3.[7]

5. **Registration of research output.** Quality-assured data and software publications that are recognized by the respective scientific community and can be referenced independently should be reported to the University Library by the authors for inclusion in the University Bibliography.

6. **Citation of data and software.** If no subject-specific standards have been established yet for the citation of data and software and if no guidelines by the journal or the publisher are available, it is recommended to refer to the Data Citation Principles.[8]

7 Support in the selection of a suitable software license is provided on the website https://choosealicense.com.
8 Data Citation Synthesis Group. "Joint Declaration of Data Citation Principles - FINAL". FORCE11 (2013). https://doi.org/10.25490/a97f-egyk.

III. Contracts and Cooperations

1. When negotiating grant agreements, in particular with private sponsors, cooperation agreements, and license agreements, researchers and research support staff should ensure that these agreements observe the principles of the Research Data Policy of the University of Potsdam as much as possible, in particular with regard to the openness and reusability of research data. When transferring rights for reuse, publication, or commercial utilization, care should be taken to ensure that the data remain freely available for scientific purposes, and in particular that no exclusive rights are granted to commercial actors.

2. Cross-institutional research cooperations should align their practices with the Research Data Policy of the University of Potsdam, unless other parties mandate equivalent or stricter requirements. Within the framework of their governance, institutionalised cooperations like clusters and networks should establish clear and binding regulations on joint data management as well as on the openness and usability of their research data at an early stage.

IV. Institutional Responsibility

1. All faculties are advised to consider whether qualification theses should contain statements on data availability in the future and, if so, to set up regulations with an appropriate degree of bindingness.

2. The study commissions are advised to reassess curricula to ensure appropriate consideration of the practical handling of research data as comprehensive course content in undergraduate and consecutive Masters programs.

3. Departments and research groups are advised to appoint research data contacts to enable institutional archiving of the research data at the University of Potsdam.

Illustrations

4/5 **Roese, Thomas**.

6/7 **Network graphs of German-language dramas from DLINA Corpus (1731–1929)**. Fischer et al., Figshare (2015). http://doi.org/drkf
CC BY 4.0

8/9 **Micrographs of Janus droplets**. Raju et al., RSC Advances 9 (2019), 19271, fig. 5. http://doi.org/drj9
CC BY 3.0 Unported

10/11 **Network graph of Lessing's Emilia Galotti**. Trilcke & Fischer, ZfdG S3 (2018), art. 3, fig. 4. http://doi.org/drkg
CC BY 4.0

12/13 **Radar-based precipitation nowcasting model output.** Ayzel et al., Geoscientific Model Development 12 (2019), 1387, fig. 5. http://doi.org/drj7
CC BY 4.0

14/15/20/21 **Crystal structure of erioflorin**. Paz et al., Acta Crystallographica E 73 (2017), 334, fig. 1. http://doi.org/drkc
CC BY 2.0 UK

16/17 **850-hPa austral summer months mean wind speed over South America**. Ziarani et al., Atmosphere 10 (2019), 379, fig. 1. http://doi.org/dpjs
CC BY 4.0

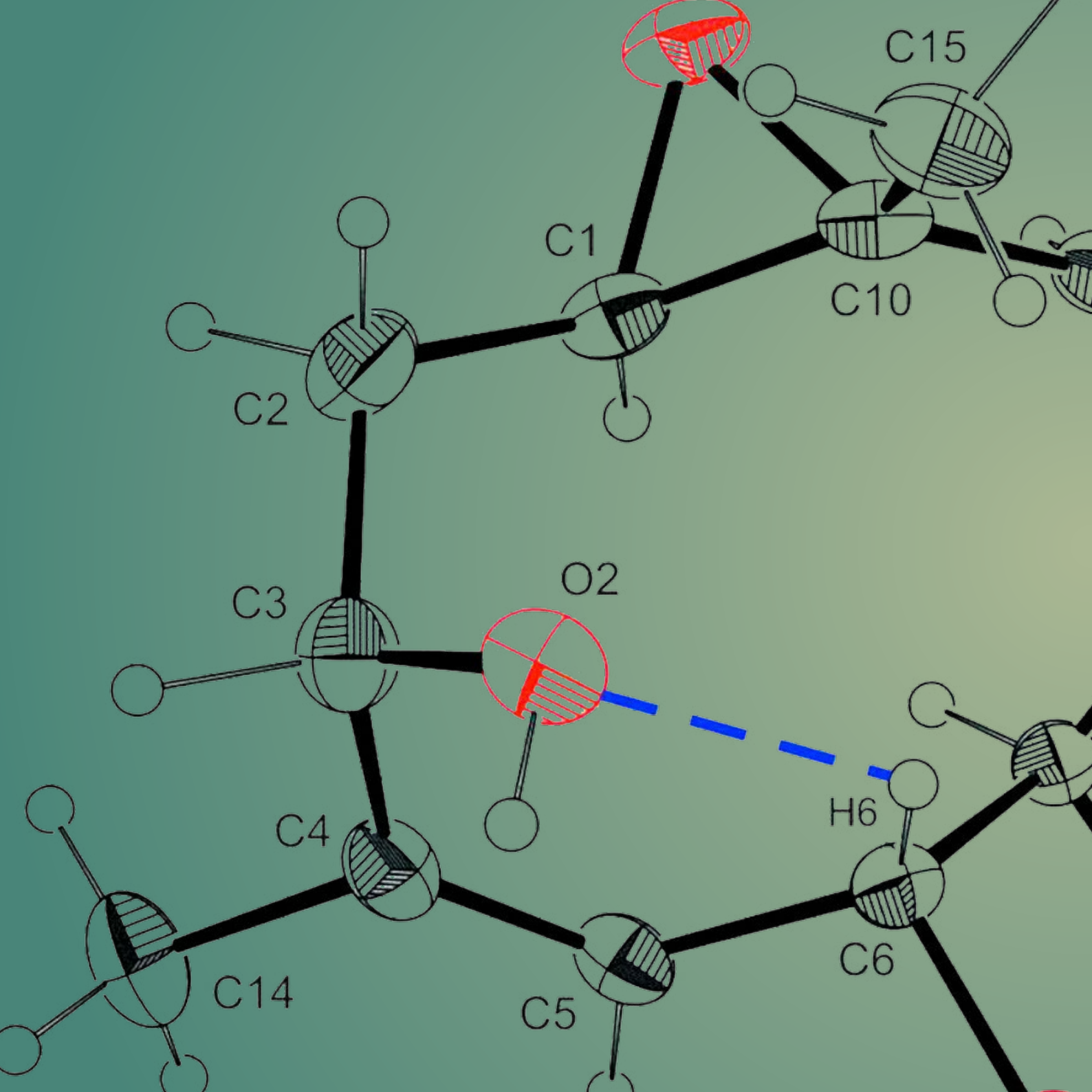
18/19 **Co-occurrence networks of prokaryotic and eukaryotic taxa on microplastics**. Kettner et al., Frontiers in Microbiology 10 (2019), 538, fig. 4. http://doi.org/drj8
CC BY 4.0

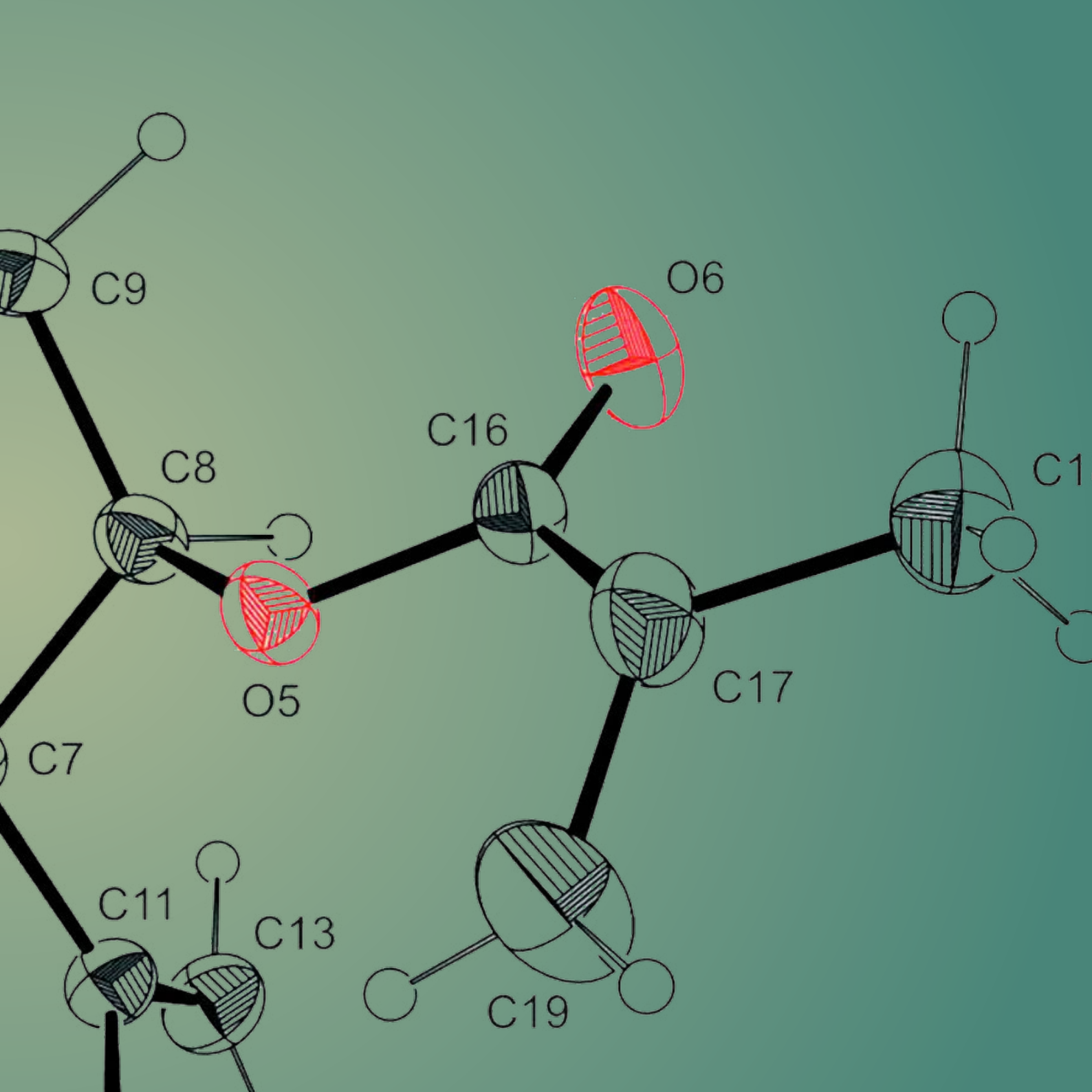
To view the terms of the respective license, please follow the corresponding hyperlink:

CC BY 4.0
https://creativecommons.org/licenses/by/4.0/

CC BY 3.0
https://creativecommons.org/licenses/by/3.0/

CC BY 2.0 UK
http://creativecommons.org/licenses/by/2.0/uk/legalcode


C15
C1
C10
C2
C3
O2
H6
C4
C14
C5
C6


C9
O6
C16
C8
C1
O5
C17
C7
C11
C13
C19