

Information Infrastructure for Data Assimilation

Z03



Remember the Research Data Lifecycle?

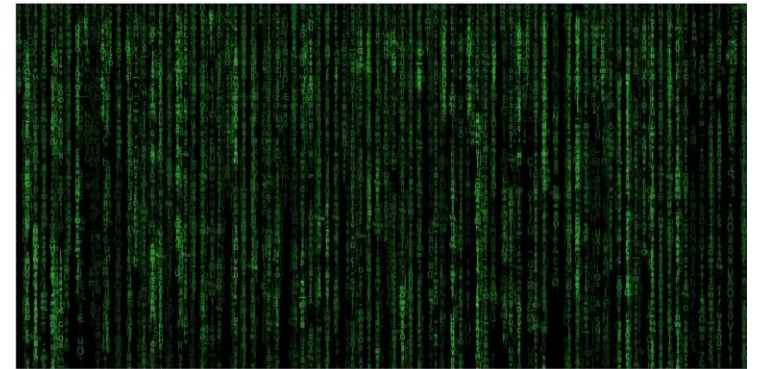
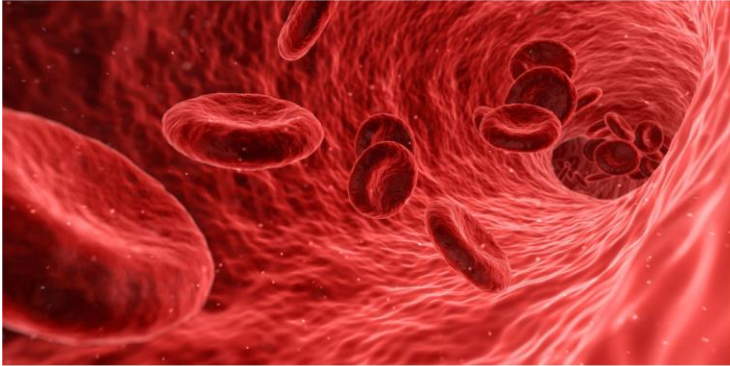
Planning and Monitoring

are proven instruments
for quality assurance.

(But, time is money ...)



What research data are you working with?



Definition of digital research data



No
exact
definition

General:

„...all digital data that are created during the research process or result from it.“

(Kindling & Schirmbacher, 2013)

The crisis of reproducibility



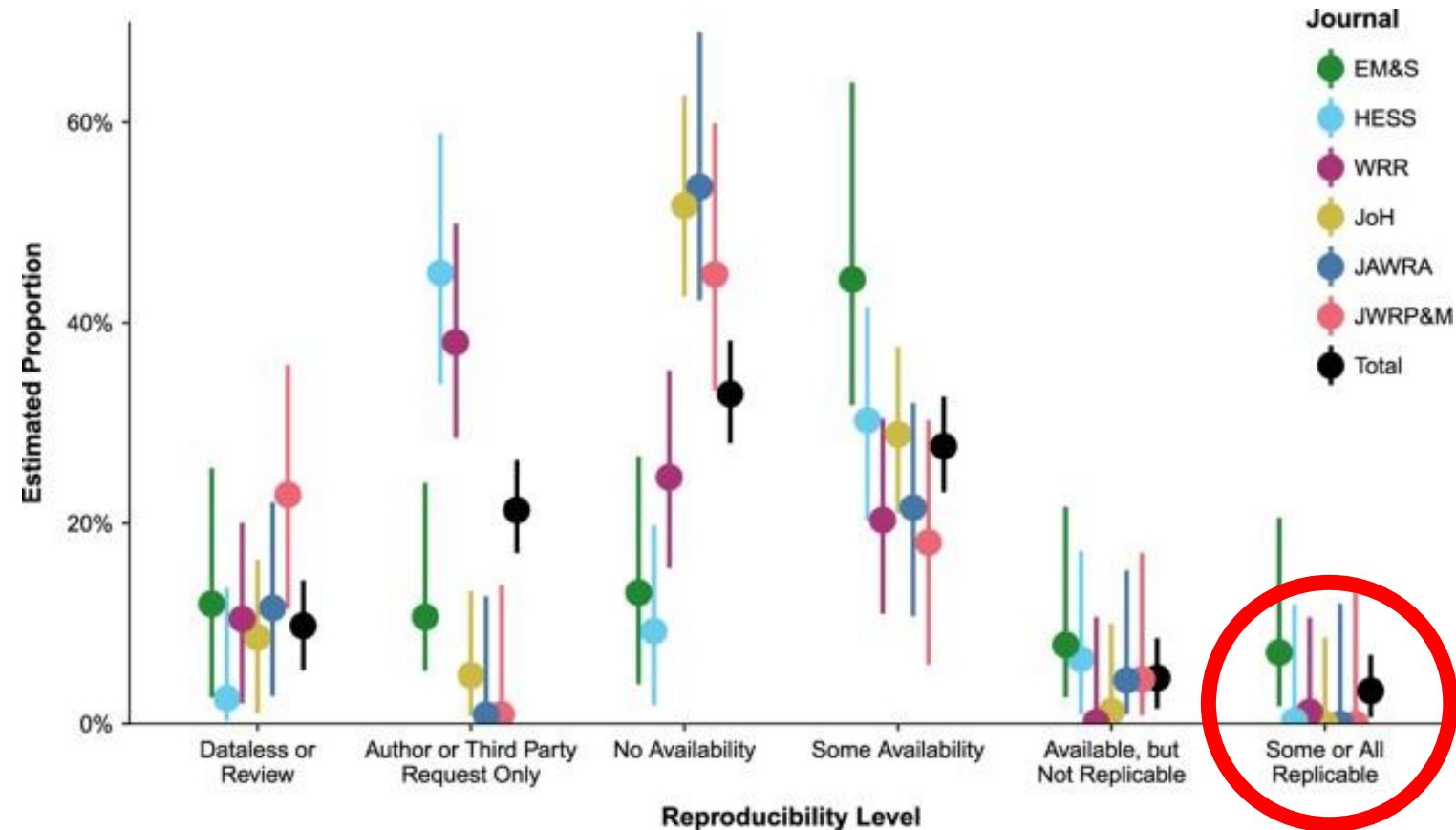
The reproducibility of scientific results

Study by Stagge et al. 2019

Investigation of 360 publications published 2017 in six Hydrology & Water resource journals

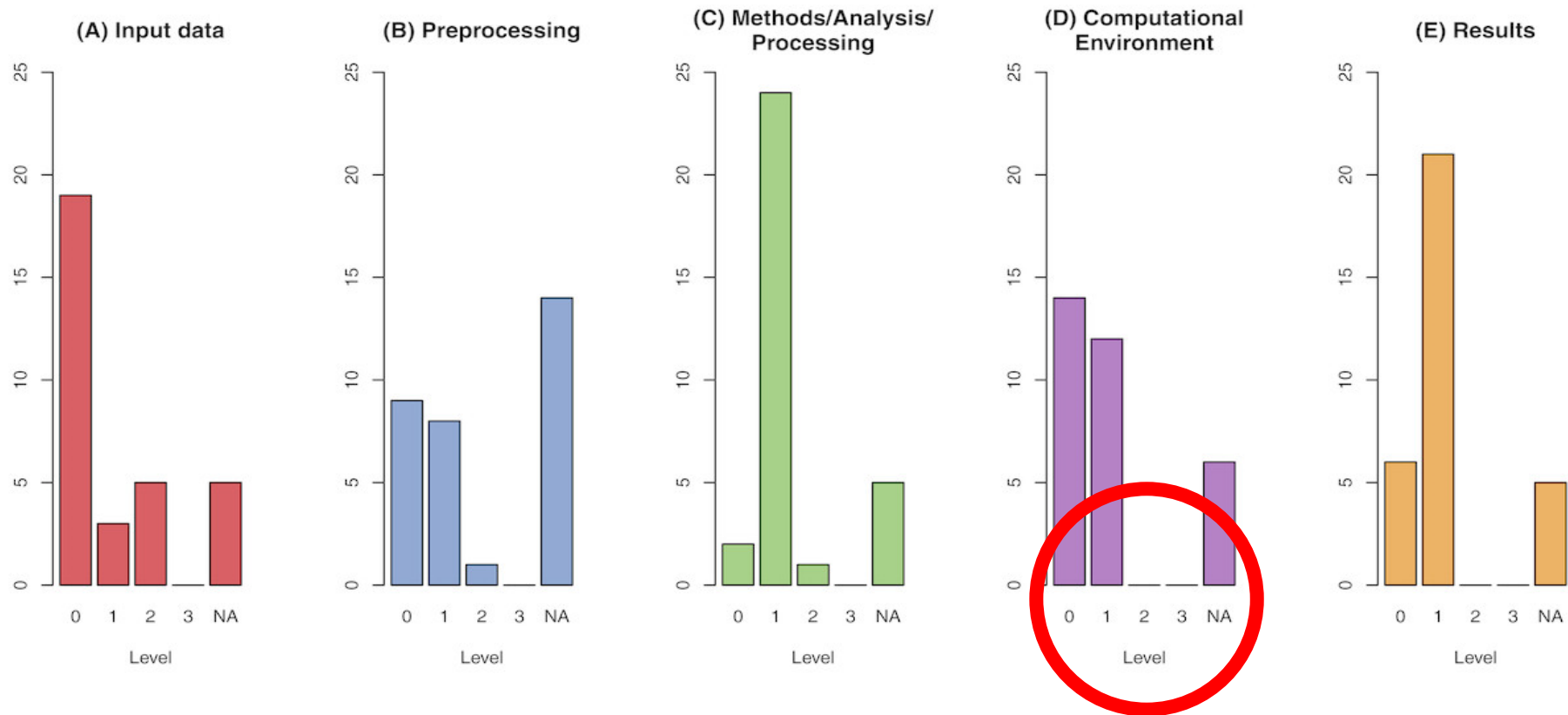
Only **1.7%** of the investigated publications allowed for a reproduction of the results.

Limitations: Incomplete data or insufficient documentation



Just an isolated case?

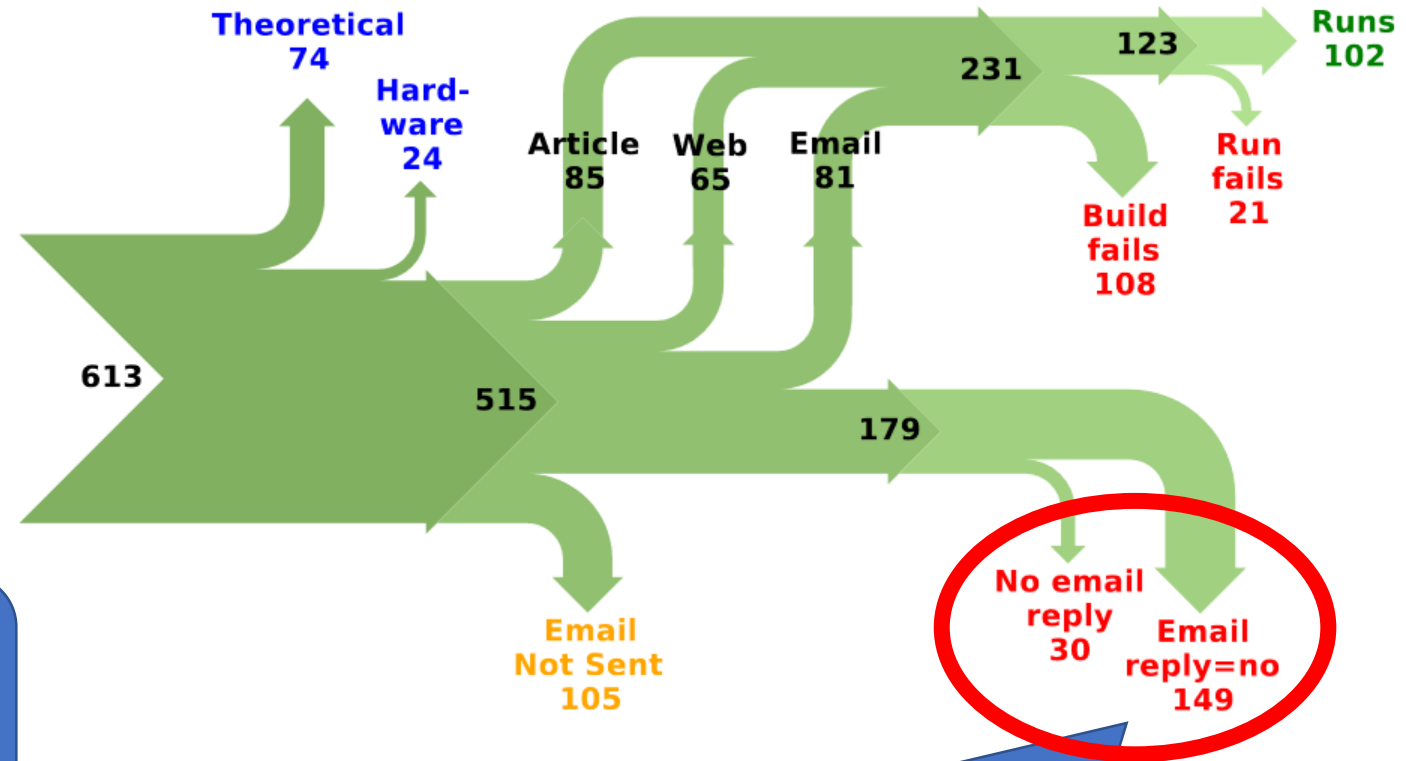
Study by Nüst et. al. 2018 in environmental science



And in other disciplines?

Study by Collberg et al. in 2016

Reasons for non-repeatability in computer systems research



“I’m not very sure whether it is the final version of the code used in our paper, but it should be at least 99% close.”

“Unfortunately the current system is not mature enough at the moment, so it’s not yet publicly available.”

How many percent of the experiments published by CRC 1294 members are reproducible?

*the first 53 publications (~Summer 2020)



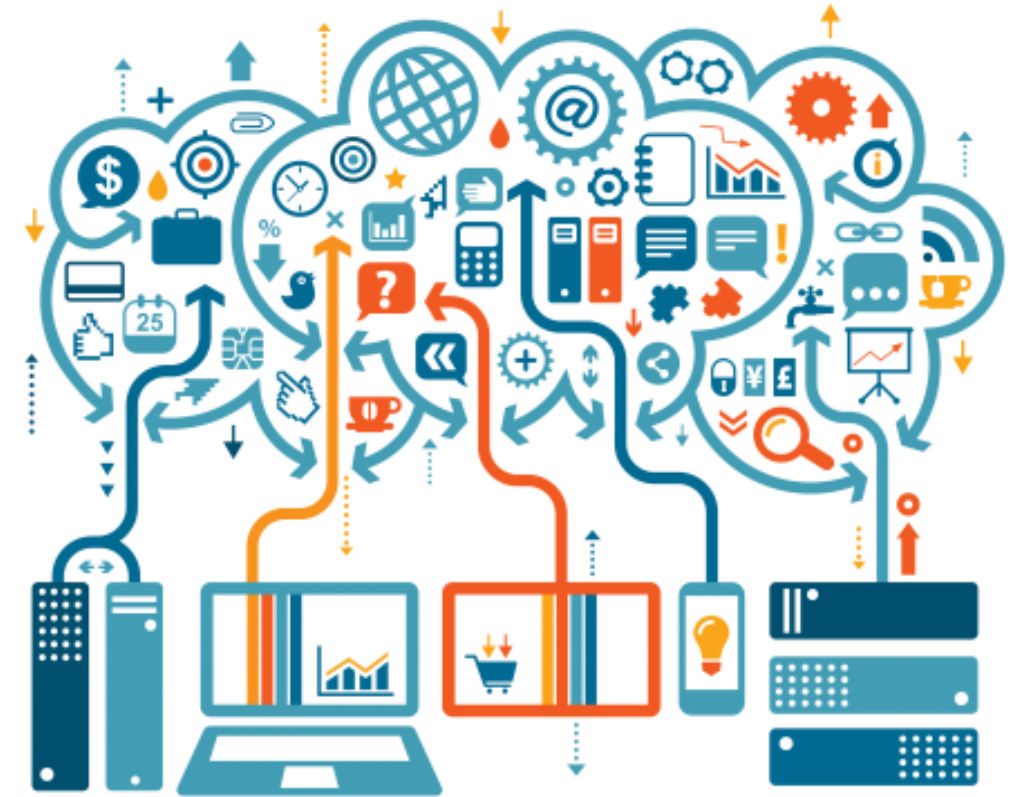
Research Data Management

All activities related to

- Preparation
- Storage
- Archiving &
- Publication

of scientific data.

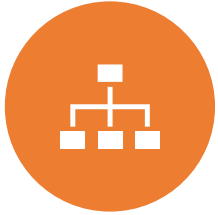
All digital data created during the research process including research results.



<https://wizardtechsolutions.com/wp-content/uploads/2017/03/big-data-cloud.png>

Research Data Management supervises research processes from initial planning to archiving, re-use, or deletion.

Aspects of Research Data Management



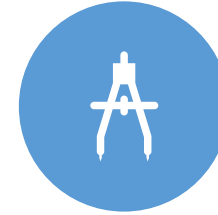
REGULATION AND
STRUCTURING



DOCUMENTATION
AND METADATA



STORING AND
BACKUP



LONG-TERM
ARCHIVING



SECURITY AND
PRIVACY

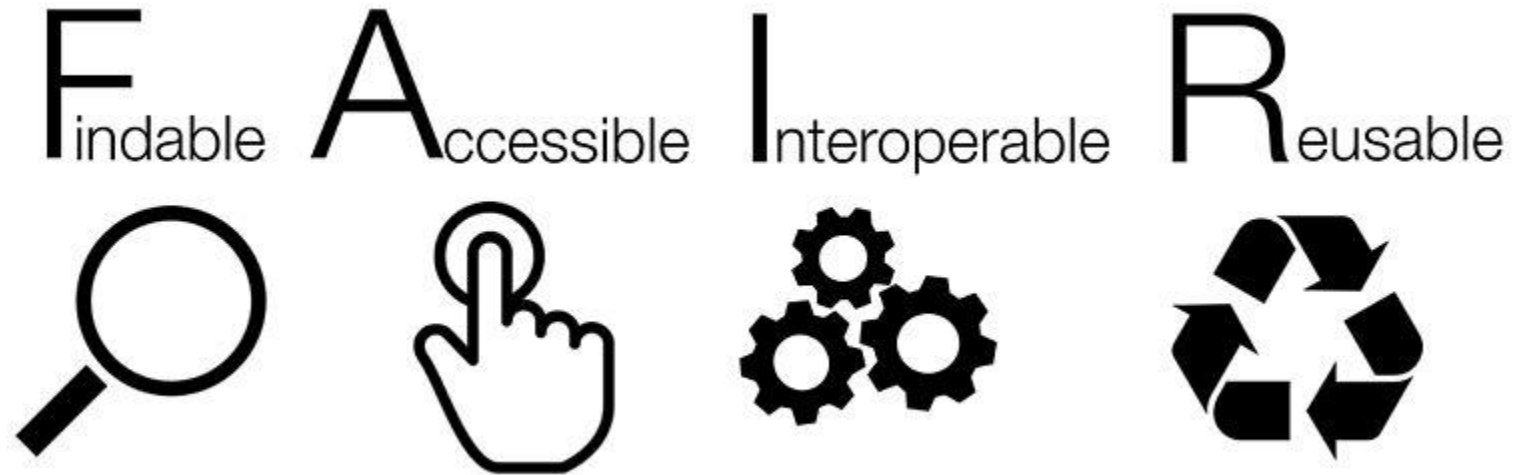


PUBLICATION



LEGAL ASPECTS






G20 countries have committed themselves to guarantee access to publicly funded research data based on the FAIR principles.

https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967

Research data policies



Guidelines for
handling
research data

For example:

Journal Policies

Funding Agency Policies

Subject-specific Policies

Institutional Policies

Journal Policies – Springer Nature



Type 1:

(Photosynthesis Research journal)

„Data sharing and data citation is encouraged“



Type 2:

(Plant and Soil journal)

„Data sharing and evidence of data sharing is encouraged“



Type 3:

(Humanities and Social Science journal)

„Data sharing encouraged and statements of data availability required“



Type 4:

(Scientific Data journal)

„Data sharing, evidence of data sharing and peer review of data required“

Funding Agency Policies

Project Planning

- Applicants should [...] detail in the proposal **what research data will be generated or evaluated** during a scientific research project.

Accessibility

- [...] research data should be made available as soon as possible. **Data should be made accessible at a stage of processing that allows it to be usefully reused by third parties** [...] it must be ensured that access to the data is still guaranteed when, through publication, the rights of use relating to research data are transferred to a third party, usually a publishing house.

Archiving

- [...] **research data should be archived** in the researcher's own institution or an appropriate nationwide infrastructure **for at least 10 years**.

ETHICAL GUIDELINES OF THE GERMAN INFORMATICS SOCIETY

Bonn, June 29, 2018

PREAMBLE

The German Informatics Society (GI) is a registered non-profit organization. With these guidelines, the GI seeks to establish that matters of professional ethics or moral conflicts become the subject of collaborative reflection and action. The guidelines are designed to offer a point of orientation not only to members of the GI association, but to all persons involved in the design, manufacture, operation or use of IT systems.

The ethical guidelines outlined herein express the intent of GI members to conduct themselves in accordance with the values that form the basis of Basic Law for the Federal Republic of Germany and the Charter of Fundamental Rights of the European Union.

The GI and its members are committed to adhering to these guidelines. They also seek to ensure that these guidelines find acknowledgement in public discourse outside the GI.

GI members are especially committed to respecting and protecting human dignity. Whenever norms of the state, society or the private sphere come into conflict with these values, GI members must address the issue.

GI members conduct themselves in such a way as to advocate for the right to self-determination in information and communications technologies, and for the right to guarantee confidentiality and integrity of IT systems.

GI members advocate for discrimination-free organizational structures which take into consideration the divergent needs and diversity of all human beings in the design, manufacture, operation and use of IT systems.

GI members seek to engage and educate the public in discourse concerning the ethical and moral issues pertinent to their individual and institutional conduct. In a networked world, it is imperative that all potential courses of action be subject to interdisciplinary consideration regarding their foreseeable impact and potential consequences. This is the challenge for each of our members.

The fact that the guidelines established here are as open as they are is testimony to the fact that moral conduct cannot be governed by a definitive code of ethics or stringent regulations.

SECTION 1: PROFESSIONAL COMPETENCE

GI members stay abreast of the current state of science and technology in their respective areas of specialization; they take new developments into account and provide constructive criticism. GI members are constantly working to improve their professional competencies.



SECTION 2: EXPERTISE AND COMMUNICATIVE COMPETENCE

GI members are constantly improving their levels of expertise and communicative competencies in order to meet the demands relevant to their duties in the design, manufacture, operation and use of IT systems and to understand the surrounding professional and technical contexts. In order to assess the consequences of IT-systems in the application environment and to propose suitable solutions, there must be a willingness to understand and take into account the rights, needs and interests of those parties who are impacted by them.



SECTION 3: LEGAL EXPERTISE

GI members are familiar with and observant of pertinent legal regulations concerning the design, manufacture, operation and use of IT systems. GI members, in conjunction with their expertise and professional competencies, participate actively in drafting legislative regulations.



SECTION 4: POWERS OF DISCERNMENT

GI members sharpen their powers of discernment to render themselves better equipped to contribute to design processes with individual and collective accountability. This presupposes not only a willingness to call into question and to make judgments about individual and collective actions in public discourse, but also the ability to acknowledge the limits of one's own powers of discernment.



SECTION 5: CONDITIONS OF EMPLOYMENT

GI members are active proponents of socially equitable contractual agreements concerning terms of employment, inclusive of opportunities for professional development and shared governance.



SECTION 6: ORGANIZATIONAL STRUCTURES

GI members advocate for organizational structures which foster and facilitate socially equitable contractual agreements concerning terms of employment.



SECTION 7: TEACHING AND LEARNING

GI members who are computer science instructors foster in their students the capacity for critical thinking; they prepare learners to accept their own individual and collective responsibility, and they act as role models in this regard.



SECTION 8: RESEARCH

GI members who conduct research in the field of computer science adhere to the rules of best practices in scientific research. Of particular importance in this regard is openness and transparency in dealing with criticism and conflicts of interest, the ability to express and to accept criticism as well as the willingness to allow the impact of one's own scientific work in the research process to become the subject of discussion. Scientific research breaches boundaries. These must be clearly articulated.



SECTION 9: COURAGE OF CONVICTIONS

GI members staunchly advocate for the protection and safeguarding of human dignity, even when this is not explicitly mandated by laws, contracts or other norms, or when these stand in direct opposition to the protection and safeguarding of human dignity. This applies even in situations in which GI members' obligations to clients conflict with their responsibility to third-party stakeholders.



SECTION 10: SOCIAL ACCOUNTABILITY

In the design, manufacture, operation and use of IT systems, GI members should contribute to the betterment of local and global living conditions. GI members are responsible for the social and societal consequences of their work. Their influence on positioning, marketing and further development of IT systems should contribute to the socially acceptable and sustainable application of these technologies.



SECTION 11: FACILITATING SELF-DETERMINATION

GI members work toward ensuring that those people impacted by the usage and conditions of use of IT systems are granted adequate opportunity to participate in the design of these systems. This is especially pertinent with regard to systems whose application involves the exerting influence over, monitoring, or surveillance of said populations.



SECTION 12: THE GERMAN INFORMATICS SOCIETY

The German Informatics Society encourages its members to adhere to these guidelines at all times. The GI shall attempt to mediate between parties in situations in which conflicts arise.



Subject-specific policies

SECTION 1: PROFESSIONAL COMPETENCE

GI members stay abreast of the current state of science and technology in their respective areas of specialization; they take new developments into account and provide constructive criticism. GI members are constantly working to improve their professional competencies.



GERMAN
INFORMATICS SOCIETY



CONDITIONS
ENT
proponents
contractual
terms
of
professional



SECTION 9: COURAGE OF CONVICTIONS
GI members staunchly advocate for the protection and safeguarding of human dignity, even when this is not explicitly mandated by laws, contracts or other norms, or when these stand in direct opposition to the protection and safeguarding of human dignity. This applies even in situations in which GI members' obligations to clients conflict with their responsibility to third-party stakeholders.



SECTION 8: RESEARCH

GI members who conduct research in the field of computer science adhere to the rules of best practices in scientific research. Of particular importance in this regard is openness and transparency in dealing with criticism and conflicts of interest, the ability to express and to accept criticism as well as the willingness to allow the impact of one's own scientific work in the research process to become the subject of discussion. Scientific research breaches boundaries. These must be clearly articulated.



The GI and its members are committed to adhering to these guidelines. They also seek to insure that these guidelines find acknowledgement in public discourse outside the GI.

GI members are especially committed to respecting and protecting human dignity. Whenever norms of the state, society or the private sphere come into conflict with these values, GI members must address the issue.

GI members conduct themselves in such a way as to advocate for the right to self-determination in information and communications technologies, and for the right to guarantee confidentiality and integrity of IT systems.

GI members advocate for discrimination-free organizational structures which take into consideration the divergent needs and diversity of all human beings in the design, manufacture, operation and use of IT systems.

GI members seek to engage and educate the public in discourse concerning the ethical and moral issues pertinent to their individual and institutional conduct. In a networked world, it is imperative that all potential courses of action be subject to interdisciplinary consideration regarding their foreseeable impact and potential consequences. This is the challenge for each of our members.

The fact that the guidelines established here are as open as they are is testimony to the fact that moral conduct cannot be governed by a definitive code of ethics or stringent regulations.

SECTION 2: EXPERTISE AND COMMUNICATIVE COMPETENCE

GI members are constantly improving their expertise and communicative competencies in order to be relevant to their duties in the design, manufacture, operation and use of IT systems and to understand the technical contexts. In order to assess

IT-systems in environmental solutions, they must understand the rights, responsibilities and interests of the parties who



SECTION 3: LEGAL EXPERTISE

GI members are familiar with and observe legal regulations concerning the design, manufacture, operation and use of IT systems. In conjunction with their expertise and competencies, participate actively in the development of legal regulations.

SECTION 4: POWER

GI members sharpen their powers of judgment and make themselves better equipped to assess the impact of individual and collective actions. They are willing to acknowledge the limits of one's own powers of judgment.



Institutional Policies

Auszug aus den Amtlichen Bekanntmachungen Nr. 3 vom 18.3.2022

<p>Neufassung der Satzung „Selbstkontrolle in der Wissenschaft – Regeln zur Sicherung guter wissenschaftlicher Praxis an der Universität Potsdam“</p> <p>Vom 16. Februar 2022</p> <p>Der Senat der Universität Potsdam hat auf Grundlage des § 64 Abs. 2 Nr. 2 des Brandenburgischen Hochschulgesetzes (BHG) vom 28. April 2014 (GVBl. I/14, [Nr. 18]), zuletzt geändert durch Gesetz vom 23. September 2020 (GVBl. I/20, [Nr. 26]) in Verbindung mit Art. 14 Abs. 1 Nr. 2 und Nr. 4 der Grundordnung der Universität Potsdam (GrundO) vom 17. Dezember 2009 (AmtBek. UP Nr. 4/2010 S. 60), zuletzt geändert durch die Fünfte Satzung zur Änderung der Grundordnung der Universität Potsdam (GrundO) vom 21. Februar 2018 (AmtBek. UP Nr. 11/2018 S. 63-4) am 16. Februar 2022 folgende Satzung erlassen:¹</p> <p>Inhaltsverzeichnis:</p> <p>Erster Abschnitt: Regeln zur Sicherung guter wissenschaftlicher Praxis</p> <p>§ 1 Allgemeine Grundsätze</p> <p>§ 2 Ausbildung und Betreuung des wissenschaftlichen Nachwuchses</p> <p>§ 3 Verantwortung und Zusammenarbeit</p> <p>§ 4 Forschungsdaten und Forschungsergebnisse</p> <p>§ 5 Publikation und Anerkennung</p> <p>§ 6 Leistungs- und Bewertungskriterien, Begrenzungen</p> <p>§ 7 Wissenschaftliches Fehlverhalten</p> <p>§ 8 Ombudsperson, Beratung</p> <p>Zweiter Abschnitt: Regeln für den Umgang mit wissenschaftlichem Fehlverhalten</p> <p>§ 9 Verfolgung wissenschaftlichen Fehlverhaltens</p> <p>§ 10 Beteiligung der Ombudsperson</p> <p>§ 11 Vorprüfung</p> <p>§ 12 Formelle Untersuchung</p> <p>§ 13 Inkrafttreten, Außerkrafttreten, Übergangsbestimmungen</p> <p>Anhang: Mögliche Entscheidungen und Sanktionen bei wissenschaftlichem Fehlverhalten</p>	<p>Erster Abschnitt: wissenschaftlich</p> <p>§ 1 Allgemeine</p> <p>(1) Unsere Leben haben in der wissenschaftlichen Erkenntnis, von Menschheit getrieben, einen Fortschritt.</p> <p>a) der Korrektur</p> <p>b) der Richtigkeit</p> <p>c) der Redlichkeit</p> <p>d) der Unvoreingenommenheit</p> <p>(2) Jede Wissenschaftler trägt die Verantwortung des Verhaltens des 3. Praxis entspricht rechtlich gewinnungsvoll um und hierbei berücksichtigt cheitersrelevante Aspekte.</p> <p>(3) Die Einhaltung vernichtbar:</p> <p>a) Allgemeine ist, in jeder sind allenfalls (siehe</p> <p>b) Rechte und am gesamten mit</p> <p>c) Bei der Plz haben ist d send zu ber derlich, Gew</p> <p>d) Bei der Plz schungsvor meidung vo der Interpre Verblindun lich, angew</p> <p>e) Wissenschaft prüfen, ob</p>
--	--

¹ Die Deutsche Forschungsgemeinschaft (DFG) hat am 4. März 2022 die Umsetzung des Kodex „Leitlinien zur Sicherung guter wissenschaftlicher Praxis“ vom September 2019 bekräftigt.

§ 4 Forschungsdaten und Forschungsergebnisse

(1) Alle Untersuchungen sind in der wissenschaftlichen Einheit vollständig zu dokumentieren.

(2) Die Dokumentation muss eine Überprüfung und Bewertung der erzielten Forschungsergebnisse im jeweiligen Fachgebiet ermöglichen.

(3) Die Dokumentation und die Primärdaten als Grundlage für Veröffentlichungen sind mindestens zehn Jahre bei der Leitung der wissenschaftlichen Einheit, einer etwaigen Nachfolge oder einer anderen ausdrücklich zu bestimmenden Stelle (z.B. einem anerkannten Repositorium) gesichert aufzubewahren (Archivierung). Die Universität Potsdam stellt sicher, dass die hierfür erforderliche Infrastruktur vorhanden ist.

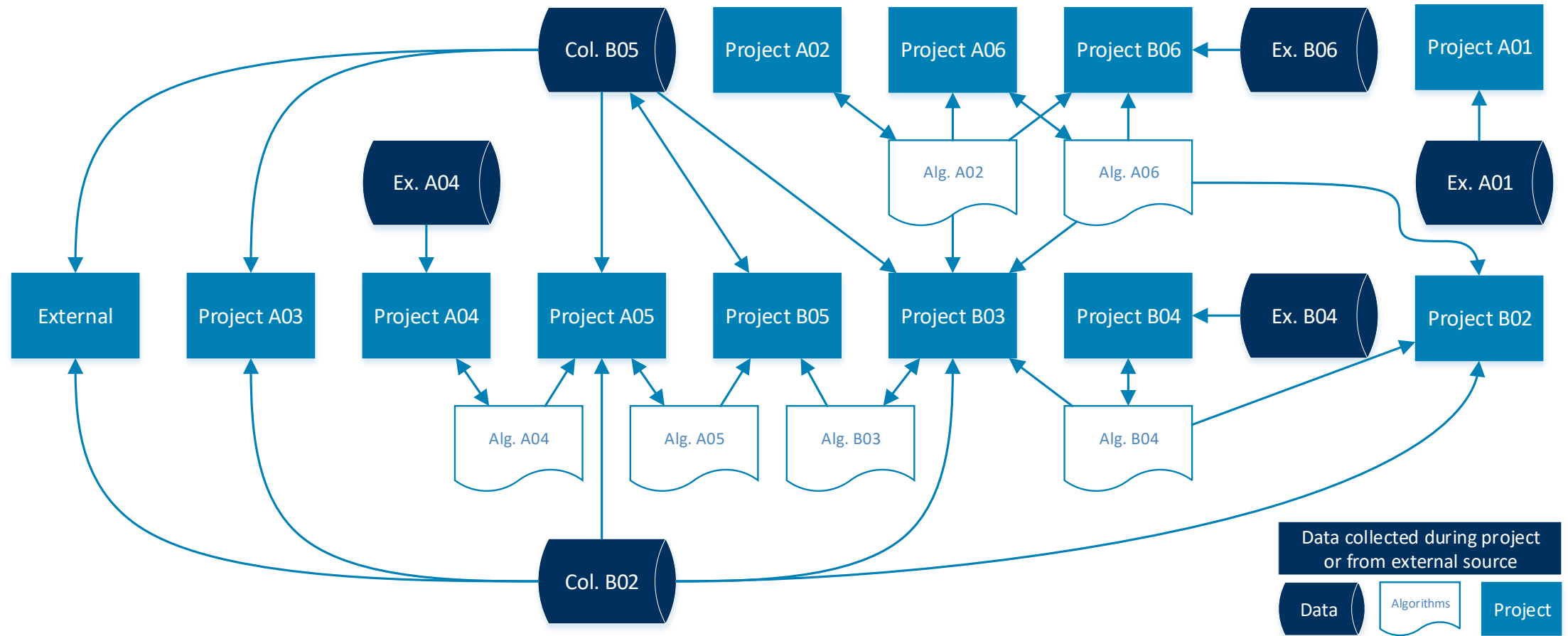
(4) Die einer Veröffentlichung zugrunde liegenden Forschungsdaten und zentralen Materialien (z.B. Quellcode selbst programmierter Software) müssen - soweit möglich und zumutbar - persistent, zitierbar und dokumentiert, den FAIR-Prinzipien („*Findable, Accessible, Interoperable, Re-Usable*“) folgend in einem im jeweiligen Fachgebiet anerkannten Archiv



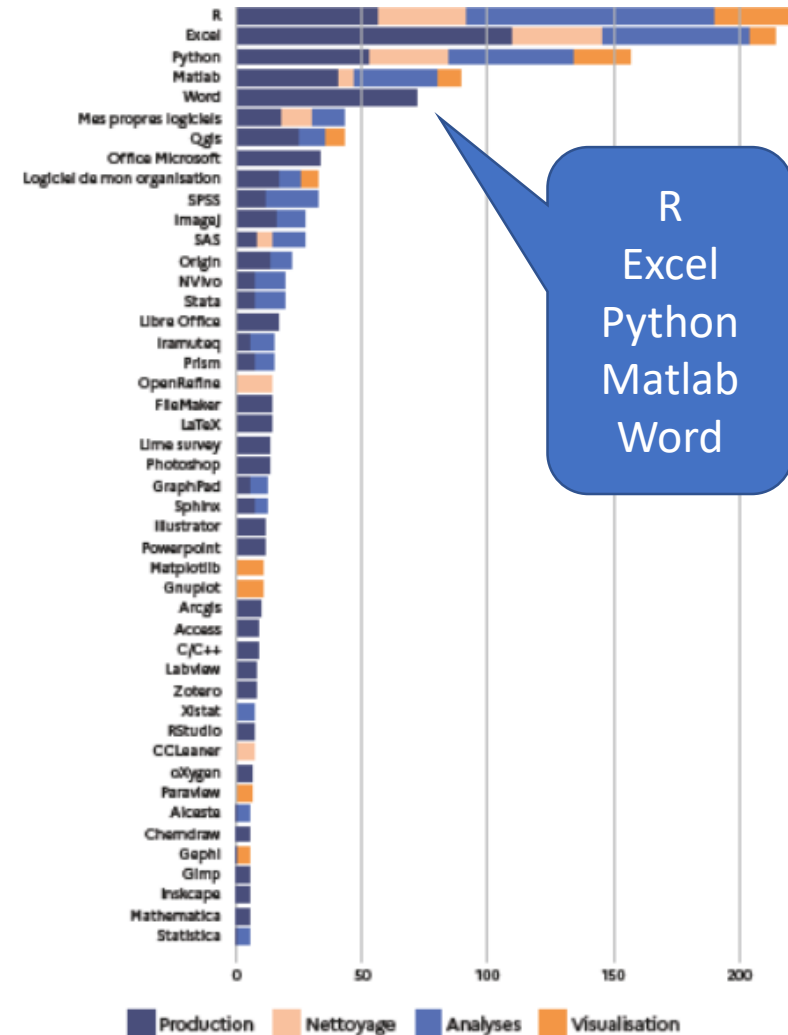
How many of those policies applicable to CRC 1294
did you know?

did you follow?

Why do we care?

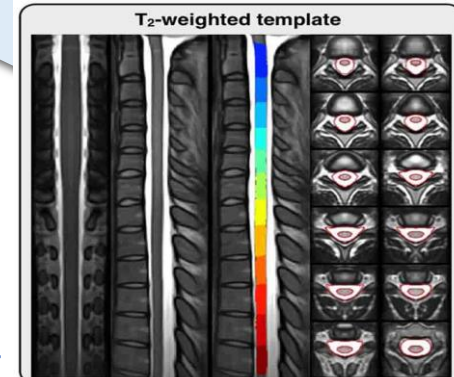
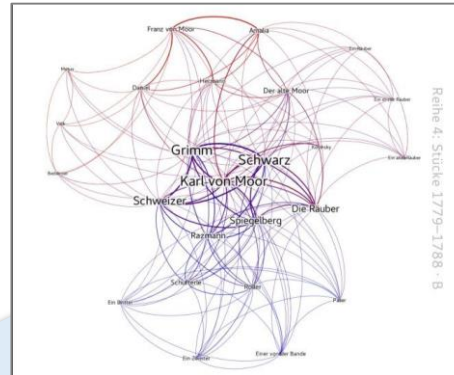


Software in Research

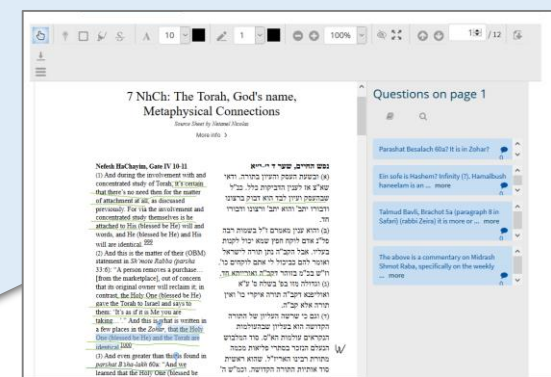
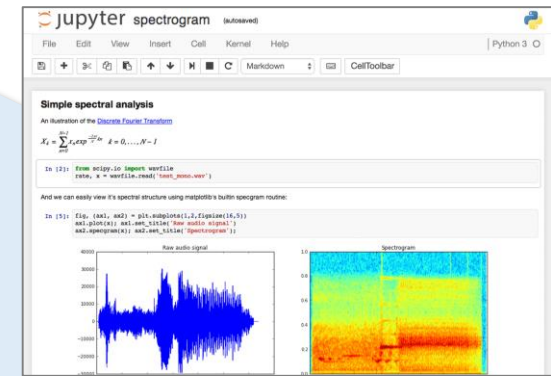


R
Excel
Python
Matlab
Word

software
edition taberna kritika
Spinal Cord Toolbox



general-purpose
special-purpose

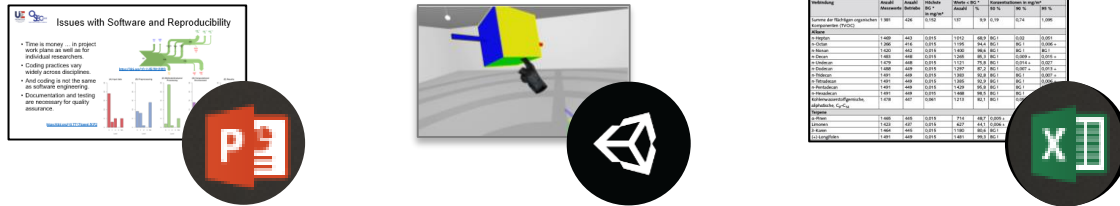


Jupyter Notebook
tools
PDFAnnotator

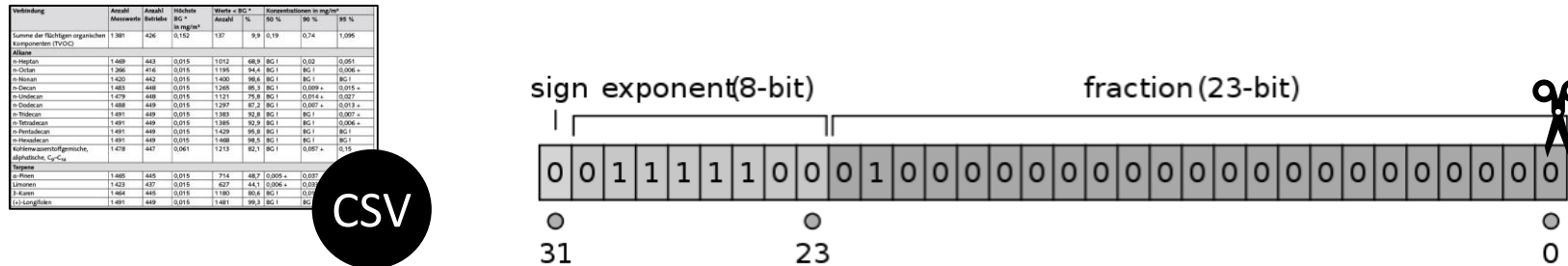
Diversity is necessary for research, but a challenge for sustainability.

Issues with Tools and Reproducibility

Proprietary data formats are often tied to specific software.

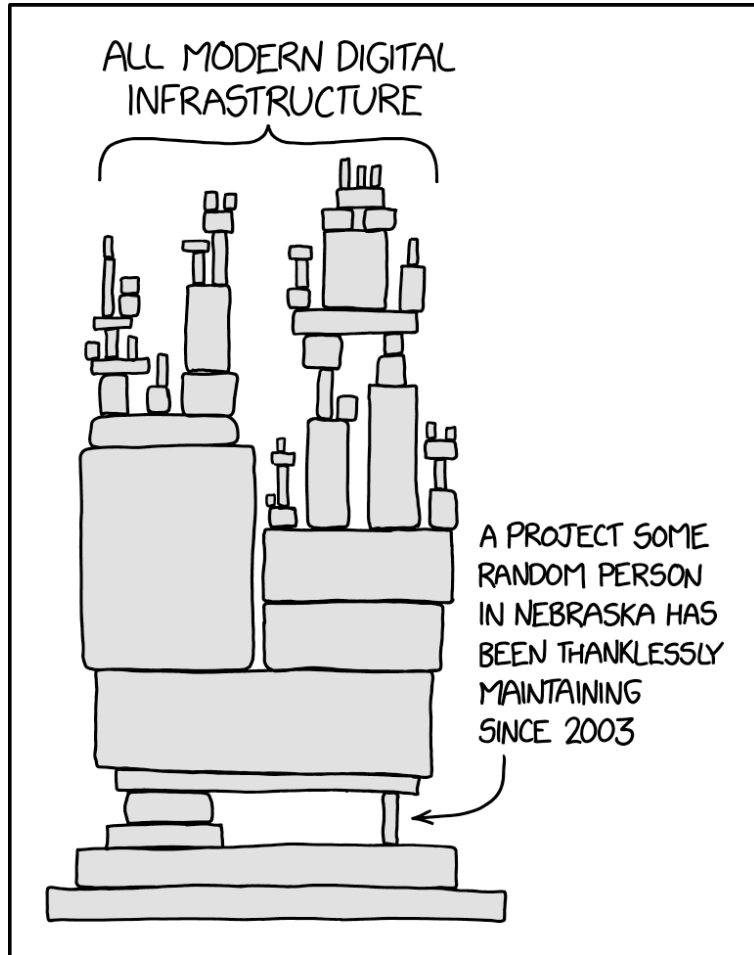


Open data formats can be tied to specific versions of software.



Software is tied to specific execution environments.
Dealing with data, tools etc. requires dedicated literacy.

Fragile Constructions

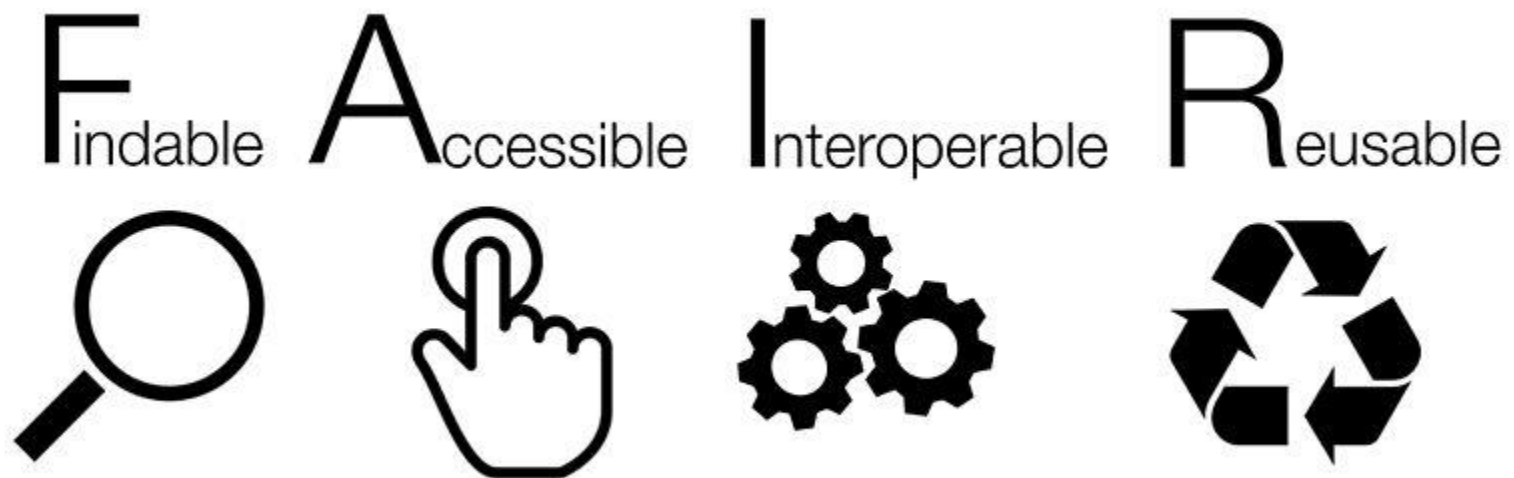


Heartbleed

log4j

fMRI bug

Valuable research funding is wasted
when basic building blocks in software crash!





Why should you care?

Challenges:

- Academic realities
- Different expectations
- Multiple policies
- ...

Advantages (for scientists and the public):

- Impact: More citations and readers
- Budget & Economic Efficiency
- Heuristics (Cumulative Science vs. Duplicate Efforts)
- Reproducibility
- Ethics, Integrity, Transparency

How to nudge such a cultural change?

Confusion
achieved



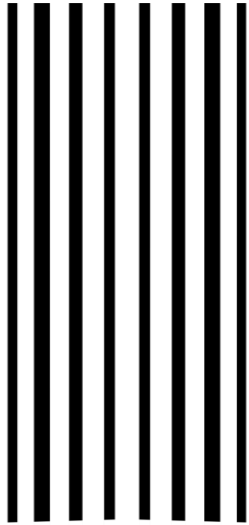


We conduct research data management anyways.
Let's make it professional!

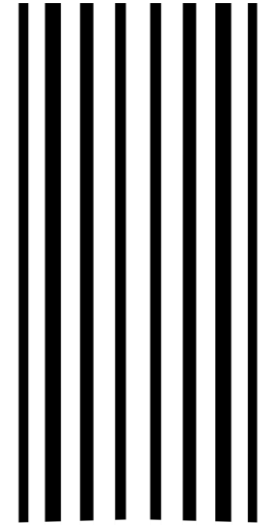
The two pillars of Z03



IT Infrastructure



Knowledge Perpetuation



The two pillars of Z03

IT Infrastructure: Collaborative work & knowledge base

- Wiki
- Overleaf
- Git.UP
- Box.UP
- RDMO.UP
- ...

Shortly: Automated quality feedback &
Container infrastructure

Knowledge Perpetuation: Individual support and collaborations

- Code & software development, Testing
- Good coding practice /
Research software engineering
- Workshops & consultations
- Packaging experts
- Exchange with NFDI initiatives

Shortly: The Carpentries membership

To assist you in making your research FAIR



Planning Research

Project Planning

Applicants should [...] detail in the proposal **what research data will be generated or evaluated** during a scientific research project.

- Required by funding agencies (depends on the programme)
- Tool for quality management in research data management


To record the intended handling of research data includes:

- activities during the research process & subsequent handling
- information about data collection, preparation, archiving, and publication

Research Data Management Organizer

University of Potsdam – Home

Language ▼



University of Potsdam – Research Data Management Organiser

LOGIN

Welcome to RDMO.UP

You can use RDMO.UP – University of Potsdam’s Research Data Management Organizer – for the following tasks:

- Plan and document data management for your research project in a structured way. To that end, you can choose from several questionnaires.
- Share data management plans with other users and edit them collaboratively.
- Create versions of data management plans to make changes and account for changing circumstances during project progression.
- Export different views on your answer to generate documents that fulfill particular funder requirements regarding data management plans.

RDMO.UP is a service of University of Potsdam’s Research Data Management Team, staffed collaboratively by the University Library (UB) and the Centre for Information Technology and Media Management (ZIM). The service currently operates in test mode. It is available for testing by University of Potsdam researchers and their external collaborators. If this does not apply to you, please do not use RDMO.UP for testing and make use of AIP’s public RDMO demo installation instead.

RDMO.UP is not currently integrated with University of Potsdam’s single sign-on. To use it you need to register.

Login

Username


Password

☐ Remember Me

If you have not created an account yet, then please sign up first.

If you forgot your password and want to reset it, click here.

Alternatively, you can login using one of the following third party accounts:




<https://rdmo.uni-potsdam.de/>

Research Data Management Organizer

University of Potsdam – Home

Language ▾



University of Potsdam – Research Data Management Organiser

LOGIN

Verify your e-mail address

We have sent an e-mail to you for verification. Follow the link provided to finalize the signup process. Please contact us if you do not receive it within a few minutes.

Research data at University of Potsdam - Contact

UB/ZIM Research Data Team

Website: <http://www.uni-potsdam.de/forschungsdaten>
E-Mail: forschungsdaten@uni-potsdam.de
Tel.: +49 331 977 2279


RDMO

RDMO is an Open Source Software project.
For more information on RDMO please visit rdmorganiser.github.io.

Research Data Management Organizer

University of Potsdam – Home

Language ▾

 University of Potsdam – Research Data Management Organiser

ULRIKE LUCKE ▾

Create new project

Title
The title for this project.

Description
A description for this project (optional).

Research Data Management Organizer

Catalog
The catalog which will be used for this project.

☐ **DMP-UP**
Questionnaire for data anagement plans (DMP) for research projects at the University of Potsdam.

☐ **DMP-HuWi (beta version)**
This catalogue for the faculty of human sciences is in development.

☐ **DFG**
Fragenkatalog nach den DFG *Leitlinien zum Umgang mit Forschungsdaten (2015)*. Geeignet für Datenmanagementpläne, die Förderanträgen an die Deutsche Forschungsgemeinschaft (DFG) beigelegt werden sollen und keine fachspezifischen Zusatzanforderungen zu erfüllen haben.

☒ **CRC 1294**
Questionnaire for the Collaborative Research Centre 1294 - Data Assimilation (english language only)

Parent project
The parent project of this project.

▼


Create project

Cancel

Research Data Management Organizer

University of Potsdam – Home

Language ▼



University of Potsdam – Research Data Management Organiser

ULRIKE LUCKE ▼

Z03 - Infrastructure project

Description

The information infrastructure project aims at removing barriers that impede an efficient collaboration between researchers in the CRC, as well as between CRC researchers and external collaboration partners. The CRC addresses several variations of data-assimilation and model building problems. Experimental evaluations involve data from different application areas and algorithms in a multitude of programming languages. Researchers in the CRC publish in different fields, and therefore even published ideas and results are not, in all cases, easily accessible by other researchers. Therefore, the goals of this project are (i) to facilitate the sharing of data, algorithms, and results, (ii) to safeguard the sustainable handling of digital artefacts, and (iii) to ensure good publication practice. To achieve these goals, the project provides information technology infrastructure, establishes research-quality tools, trains young scientists regarding research data and code, and fosters a healthy cultural change towards more FAIRness.

Catalog

CRC 1294
Questionnaire for the Collaborative Research Centre 1294 - Data Assimilation (english language only)

Options

[Answer questions](#)

Update project information
Update project catalog
Update parent project
Update project tasks
Update project views
Delete project
Add member
Create snapshot
Back to projects overview

Export




Today's homework:

1. Create a DMP for your project.
2. Answer the questions.

Research Data Management Organizer

University of Potsdam – Home

Language ▾



University of Potsdam – Research Data Management Organiser

LOGIN

Welcome to RDMO.UP

You can use RDMO.UP – University of Potsdam’s Research Data Management Organizer – for the following tasks:

- Plan and document data management for your research project in a structured way. To that end, you can choose from several questionnaires.
- Share data management plans with other users and edit them collaboratively.
- Create versions of data management plans to make changes and document changing circumstances during project progression.
- Export different views on your answer to generate documents that fulfill particular funder requirements regarding data management plans.

RDMO.UP is a service of University of Potsdam’s Research Data Management Team, staffed collaboratively by the University Library (UB) and the Centre for Information Technology and Media Management (ZIM). The service currently operates in test mode. It is available for testing by University of Potsdam researchers and their external collaborators. If this does not apply to you, please do not use RDMO.UP for testing and make use of UP’s public RDMO demo installation instead.

RDMO.UP is currently integrated with University of Potsdam’s single sign-on. To use it you need to register.

login

Username

christian.riedel

Password

.....


☐ Remember Me

Login

If you have not created an account yet, then please sign up first.

If you forgot your password and want to reset it, click here.

Alternatively, you can login using one of the following third party accounts:

 SIGN IN with ORCID ID

Knowledge Perpetuation: The Carpentries

- Join the Carpentries Network
- Z03 members give workshops at other institutions
- SFB members can choose coding and data workshop from the Carpentries network in return

Let us know about your workshop requests. We will take care of it.



Knowledge Perpetuation: Packaging experts

Shafayet Hossen Chowdhury & Safial Islam Ayon

M.Sc. Students in Data Science

Assist with:

- Reproducibility of your research
- Investigation of Data/Code
- Investigation of Documentation

Currently investigating the reproducibility of published research. Spoiler: Has improved. Stay tuned!

Safial Islam Ayon is at the Spring School today. Get in touch! 😊

A close-up photograph of an orange cat looking upwards. The cat's mouth is slightly open, and its long white whiskers are prominent. The background is dark and out of focus.

Thank you!