

CPEC Research Data Management Policy

This policy was enacted in late 2022 and took effect at the start of the second funding period on 01.01.2023.

Preamble

The Center for Perspicuous Computing (CPEC) recognizes the importance of high-quality, reusable, and well-managed research data. Research data management encompasses all phases of research: From the initial, internal processing of data, where integrity, availability, and privacy are key, to publication, where replicability and reproducibility are main goals. CPEC is committed to FAIR (findable, accessible, interoperable, and reusable) research data management.

As a collaborative research center, research at CPEC is conducted by staff from all participating institutions:

- Universität des Saarlandes
- Technische Universität Dresden
- Max-Planck-Institut für Softwaresysteme
- CISP Helmholz Center for Information Security

This Research Data Management Policy is the link between the applicable policies and guidelines of these institutions and the vision of CPEC. All CPEC researchers are expected to act in accordance with this policy and their institution's policies at all times and to bring any conflicts to the attention of CPEC's Research Data Management Officer (<https://dpad.cs.uni-saarland.de/s/cpec-forschungsdatenmanagement-policy-antler-mourner-stuff-immunity#CPEC%E2%80%99s-Research-Data-Management-Officer>) should they arise.

In addition to the provisions below, we refer to the relevant laws and regulations, notably the General Data Protection Regulation (https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en), the DFG's Guidelines for Safeguarding Good Research Practice (<https://doi.org/10.5281/zenodo.6472827>), and the DFG's Guidelines on the Handling of Research Data (https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/guidelines_research_data.pdf).

Internal Data Collection and Processing

1. All relevant research data being collected, processed, or edited must be stored in a manner that ensures data availability and integrity in the event of hardware failure or

human error.

- For source code, we recommend using Git (<https://git-scm.com/>) in conjunction with CPEC's internal GitLab instance (<https://gitlab.perspicuous-computing.science/>).
- For all other data, we recommend CPEC's centralized cloud storage (<https://cloud.perspicuous-computing.science/>).
- Alternatives to these systems may be used provided they offer equivalent guarantees of availability and integrity.

2. If research data contains personal information or other sensitive information, its treatment must comply with all relevant data protection laws. In addition, the following rules apply:

- Prior to collection of personal information as part of CPEC research, the consent of the ERB (Ethical Review Board) of the Faculty of Mathematics and Computer Science at Saarland University (or an equivalent) should be sought. The data handling rules set out by the ERB apply.

3. To the extent possible, this data must be stored on hardware and services owned and operated by CPEC participating entities, e.g., CPEC central cloud storage (<https://cloud.perspicuous-computing.science/>).

- If such data is stored on devices that are not permanently kept in secure facilities, such as an employee's laptop, the data must be encrypted, preferably with full encryption.
- Any cross-site communications involving such data must be conducted via secure, end-to-end encrypted or CPEC owned and operated services, e.g., CPEC's Mattermost chat service (<https://chat.perspicuous-computing.science/>).
- Access rights to this data must be reviewed regularly. Access rights that are no longer required must be revoked.
- When sensitive data no longer needs to be associated with individuals, it should be de-identified.

4. CPEC recommends the use of open, well-documented file formats wherever possible.

Publication of Research Data and Source Code

1. As a guiding principle, CPEC is committed to publishing research data as openly as possible.
2. Any programming code developed within CPEC that is relevant for publication will be published as open source unless prohibited by third-party code license terms. When using third-party libraries or source code, researchers will strive to use resources that

allow open source publication whenever possible. Researchers should use established open source software licenses for this purpose, such as the MIT license.

3. Any other artifacts and research data should be licensed under permissive, open licenses such as Creative Commons Licenses or the Public Domain Dedication (CC0).
4. If research data regarding user studies cannot be published in full because they contain personally identifiable information, data shall be published in pseudonymized or aggregated form to enable reproduction of the published findings wherever possible.

Open Science and Reproducibility

CPEC commits to the principles of Open Science. Researchers will preregister studies (for instance with OSF (<https://osf.io/>) or AsPredicted (<https://aspredicted.org/>)) and publish research materials, codebooks, and source code whenever possible to foster transparency and enable reproducibility of our work. We also strive to make publications openly available to the research community (e.g., in the form of preprints).

Long-Term Archival of Data

1. Any published research data shall be made available for an extended period of time, generally at least 10 years.
2. Researchers may deposit data in repositories operated by their institution where available, or use third-party repositories like Zenodo (<https://zenodo.org/>). When using third-party repositories, researchers must ensure these meet the minimum required retention time.
3. All published research data should be provided with a permanent identifier. DOIs should be used wherever possible.
4. For source code, publishing to code-hosting websites such as GitHub (<https://github.com/>) is generally not considered sufficient. If code published on such a site is referenced in a research publication, a snapshot should be archived long-term in a repository and referenced by its persistent identifier.

CPEC's Research Data Management Officer

When in doubt or if you have additional questions, please contact our internal Research Data Management Officer via CPEC's Mattermost Channel for Research Data Management Questions (<https://chat.perspicuous-computing.science/cpec/channels/research-data-management>).