

Sascha Safenreider

Good Scientific Practice in Computer and Data Science

How to do science right (“Good Scientific Practice”)

Table of contents

- 1 Good Scientific Practice
- 2 DFG Guidelines
- 3 GSP at Uni Göttingen

About the DFG Code of Conduct (19 Guidelines)

What this course follows:

- The German Research Foundation (**DFG**) publishes the *Guidelines for Safeguarding Good Research Practice (Code of Conduct)* with **19 guidelines**.
- These guidelines set standards for integrity, methods, authorship, and handling misconduct in research.
- We use these guidelines as the backbone of this section.

DFG Code of Conduct (19 Guidelines): <https://www.dfg.de/en/basics-topics/basics-and-principles-of-funding/good-research-practice/code-of-conduct>

Explanation:

- Institutions define and communicate rules of good research practice.
- Researchers must ensure their behaviour complies with these standards.
- Principles: work *lege artis* (follow the rules), honesty, questioning all results, promoting critical discourse.

Example:

6 / 38

9 / 38

Questions:

- Which dilemmas did the groups identify?
- What concrete solutions were suggested?

Guideline 8: Responsibilities and roles in the research process

Explanation:

- Roles and responsibilities must be defined clearly at the start of projects.
- Adjust roles dynamically as projects evolve.
- This ensures accountability and fairness.

Example:

In a collaborative project on cybersecurity, one researcher is responsible for data collection, another for algorithm development, and another for evaluation and documentation. This clarity prevents misunderstandings.

Group Activity – Roles and Fair Assessment

Group Task ⌚ 20 Min

- Develop guidelines for a fictional research team that ensure clear roles and fair recognition.
- Think of an order for these fictional characters and explain why.

Plenary Discussion

Questions:

- Which role distributions and measures did the groups suggest?

Block 3 Research Design, Documentation and Reproducibility

Overview

Guidelines covered:

- Guideline 7: Cross-phase quality assurance
- Guideline 9: Research design
- Guideline 11: Methods and standards
- Guideline 12: Documentation

Plan well, use state-of-the-art methods, and document everything so others can repeat your work.

Guideline 7: Cross-phase quality assurance

Explanation:

- Follow accepted standards at every step.
- When you publish, say how you checked quality.
- If you find errors later, correct or retract.
- Name the source of data, code, and tools; make results repeatable.

Example:

An ML team shares code, data version, random seeds, and the exact environment (e.g., container). After finding a bug in evaluation, they post an erratum and add tests to prevent it again.

Guideline 9: Research design

Explanation:

- Start with a careful review of related work.
- Set clear questions, metrics, and a simple analysis plan.
- Reduce bias (e.g., blinded labels, fixed stopping rules).
- Consider gender/diversity and context in your data and results.

Example:

Before a speech model, the team checks dataset coverage, defines fairness metrics, and preregisters an ablation plan to avoid cherry-picking.

Guideline 11: Methods and standards

Explanation:

- Use sound, appropriate methods; explain new methods clearly.
- Follow community standards (baselines, benchmarks, multiple runs).
- Get missing expertise via training or collaboration.

Example:

A systems paper reports mean & variance over several runs, compares to strong baselines, and releases a Dockerfile plus unit tests to rebuild results.

Guideline 12: Documentation

Explanation:

- Document all steps needed to check the result.
- Include negative or null results; do not cherry-pick.
- Follow field rules; explain any limits in your docs.
- Protect logs and results from changes; cite clearly.

Example:

A data project ships a README, data sheet (provenance, license), versioned configs, and notebooks for preprocessing; non-significant tests go in an appendix.

Group Activity – Research Design, Documentation and Reproducibility

Group Task ⌚ 20 Min

- Make a **1-page checklist** for Quality Assurance and documentation for a small CS project.
- The checklist should cover handling of data, code, experiments, and recording of negative results.

Plenary Discussion

Questions:

- Which Quality Assurance steps did the groups choose and why?
- How does the checklist help others to repeat the work?

Block 4 Publications, Authorship and Openness

Overview

Guidelines covered:

- Guideline 13: Providing public access to research results
- Guideline 14: Authorship
- Guideline 15: Publication medium
- Guideline 17: Archiving

Share results clearly, give proper credit, choose trustworthy venues, and archive data/code so work can be reused.

Guideline 13: Providing public access to research results

Explanation:

- Decide yourself whether, how, and where to publish.
- Describe results fully; where possible share data, materials, methods, and software.
- Follow FAIR (Findable, Accessible, Interoperable, Reusable)

Example:

Release code and dataset in a trusted repo (e.g., GitHub, Zenodo, DOI) with a clear license, README, and metadata. Publish one solid paper instead of several minimal ones that split the same result.

Guideline 14: Authorship

Explanation:

- Author = genuine, identifiable contribution to content (text, data, or software).
- All authors approve the final version; responsibility is shared.
- No honorary/guest authors; order based on clear criteria; use acknowledgments for non-author help.

Example:

Team roles: dataset curation, model design, experiments, writing. Decide order early (e.g., first = main analysis, last = PI). A colleague who only ran a few plots is acknowledged, not added as author.

Guideline 15: Publication medium

Explanation:

- Select venues carefully (journals, conferences, repositories).
- Check peer-review, ethics/data policies, and editorial board; beware predatory venues.
- Quality of the work does not depend on the medium.

Example:

Before submitting to a new journal, check indexing, review process, APC (Article Processing Charges) transparency, and policies on data/code. For CS, consider a reputable conference and an open repository for artifacts.

Guideline 17: Archiving

Explanation:

- Archive raw data, code, key materials, and (if used) research software for an appropriate period (often ~ 10 years, field-dependent).
- Ensure access and identifiers (e.g., DOIs); explain if some data cannot be stored.
- Institutions (e.g. GWDG) provide infrastructure to enable archiving.

Example:

Store raw logs, training data version, configs, and container image in a repository with a DOI. Sensitive data: keep encrypted with controlled access; add a data-use agreement and a retention plan.

Group Activity – Publishing, Authorship, Archiving

Group Task ⌚ 30 Min

- **Authorship:** Evaluate a fictional author list (Professor who provided funding, PhD who built the model, Research Assistant who labeled data, engineer who set up infrastructure). Propose an order and justify it.
- **Archiving:** Draft a 6-point archiving plan (what, where, how long, access, identifiers, licenses) for a small dataset + codebase.
- **Venue check:** Pick a target journal/conference and list 5 checks to avoid predatory venues.

Plenary Discussion

Questions:

- What criteria did you use for authorship order and why?
- How does your archieving plan look like?
- Which checks best detect unreliable publication venues?

Block 5 Misconduct, Conflicts and Peer Review

Overview

Guidelines covered:

- Guideline 6: Ombudspersons
- Guideline 16: Confidentiality and neutrality of reviews
- Guideline 18: Complainants and respondents
- Guideline 19: Procedures for alleged misconduct

Handle conflicts and suspected misconduct fairly: protect people, keep reviews confidential, and follow clear procedures.

Guideline 6: Ombudspersons

Explanation:

- Each institution appoints independent ombudspersons ([GAUSS](#)).
- They advise confidentially on good practice and conflicts.
- They can forward cases to the proper committee when needed.

Example:

A PhD fears unfair authorship. They contact the local ombudsperson, who keeps it confidential, mediates a meeting, and, if needed, refers the case to the investigation committee.

Guideline 16: Confidentiality and neutrality of reviews

Explanation:

- Keep manuscripts/proposals strictly confidential; do not reuse content.
- Declare conflicts of interest and recuse when needed.
- Members of panels must follow the same rules.

Example:

A Committee member recognizes a submission from a close collaborator. They declare the conflict and step out. They never use the code ideas seen during review.

Guideline 18: Complainants and respondents

Explanation:

- Protect both sides; apply presumption of innocence.
- Complaints must be in good faith; malicious claims are misconduct.
- Anonymity may be allowed if facts are concrete; careers must not be harmed.

Example:

An anonymous tip alleges fabricated benchmarks. The committee accepts detailed evidence, keeps names confidential, and ensures the accused faces no penalties unless misconduct is proven.

Guideline 19: Procedures for alleged misconduct

Explanation:

- Institutions define what counts as misconduct (e.g., fabrication, falsification, plagiarism).
- Clear, lawful steps: intake, assessment, investigation, hearing, decision, measures.
- Process is timely, fair, and confidential; sanctions match severity.

Example:

A study merges training and test data. After investigation and hearing, the paper is retracted, the lab updates Quality Assurance rules, and the institution issues measures per policy.

37 / 38

Questions:

- How do your plans protect both the complainant and the respondent?
- What steps ensure a fair procedure?

GSP at Göttingen University

- Orientation Framework for Good Research Practice
- "Ordnung [...] zur Sicherung guter wissenschaftlicher Praxis"
(Regulations for ensuring good scientific practice)
- <https://www.uni-goettingen.de/en/567647.html>

Group Task – How does this work in Göttingen?

Group Task ⌚ 60 Min

- 1 Get an overview over the orientation framework (5min)
- 2 Each group gets to answer 1-2 questions (30min)
 - ▶ What are our rules regarding using AI in research?
 - ▶ When and how are publications withdrawn?
 - ▶ When and how should you ask an ethics committee?
 - ▶ What do you have to do to collect personal data e.g. for a questionnaire?
 - ▶ What should you do with raw data from your research?
 - ▶ What should you do with source code from your research?
 - ▶ What to do when a superior suggests changing authors (unjustified)?
- 3 Present your findings to everyone (25min)