Using software for Research: The Role of Transparency, Reputation and Compliance in Practicing Trust

Judith Hartstein[1,2], Alexander Schniedermann[1], Nathalie Schwichtenberg[1]

[1]German Centre for Higher Education Research and Science Studies, Germany

[2]Humboldt-Universität zu Berlin, Germany

DOI 10.3217/978-3-99161-033-5-001, CC BY 4.0

Abstract. What makes researchers trust in software? We propose that some - but not all - considerations before software-for-research use are disciplinary practices. The representative DZHW Science Survey provided us with n ~ 1,300 complete responses from researchers about their considerations preceding software use. On this data, we performed multiple and simple correspondence analyses to explore patterns concerning trust in research software.

According to multiple correspondence analysis, the relevance of considerations in general is an in-participant characteristic. With simple correspondence analysis, we find that transparency-related considerations are disciplinary practices, while reputation-related considerations are shared across disciplines. We gained mixed results on compliance-related considerations.

We suggest that infrastructure designers should be aware of the pre-established relevance of software-literacy-related and methods-related considerations in some fields as opposed to others.

1. Introduction

In modern science, various domains of research practice involve computational methods or computer-assisted creative work. The range of practices spans from writing with digital tools to empirical research with a completely digital observation-analysis pipeline. While researchers are still divided into a panoply of academic disciplines or fields which differ in their epistemic practices, they share the goal to obtain insights. It is widely acknowledged that the reliability of research results crucially depends on good practices during the research process across fields. But software-use-related epistemic practices have not yet been systematically investigated as disciplinary practices. Our article addresses this research gap with correspondence analysis applied to survey data on considerations about research software.

Software solutions for digital methods, such as cataloguing and searching or statistical computations, are often transferred from one field to the other. On the one hand, research methods embodied as software carry implicit assumptions, sometimes without the users noticing. On the other hand, using software for research is to put faith in it working as expected and to accept its epistemic consequences. Therefrom arises a dissonance which we aim at better understanding with our research: How is trust in software built? And how does this vary by discipline?

We argue that when it comes to research software use, transparency as a 'trust technology' (see Grand et al. 2012) privileges the 'Hard Sciences' over the 'Soft Sciences'. Our study shows that when assessing research software before using it, researchers from the Engineering Sciences and the Natural Sciences consider transparency-related software attributes more often than researchers from the Humanities and the Social Sciences do. Moreover, transparency does not complement but rather adds to reputation-related heuristics for when to trust in research software. Besides, we gained mixed results on compliance-related characteristics of which some are considered in the Life Sciences more than in other fields.

Empirical researchers across disciplines use software in their everyday work: knowledge is produced (as opposed to 'found', see Bonde Thylstrup et al. 2019) through data with software. Therefore, mathematical and computational modelling in research is part of the epistemic work from data to conclusions (Gramelsberger et al. 2024), which is why software practices have recently come into focus for achieving reproducibility with archived data (Davenport et al. 2020).

Software reuse means transferring a fixed configuration of research methods from one research context to another which bears epistemic risks. For computer systems in general, Friedman and Nissenbaum (1996) describe how values embedded throughout systems design travel with the system from the context of production to the context of use, which can lead to biases which are hard to overcome. Fast forward 28 years later, Kinder-Kurlanda and Fahimi (2024) still observe algorithmic fairness is hard to achieve (and even to define) in practice.

Some would argue that automation and artificial intelligence could help to cancel out biases in general, a perspective which is discussed as technological solutionism (see Morozov 2014). But still, research with software is, at its core, human reasoning. According to Bechmann and Bowker (2019), software use is not deterministic on the data which are given, but human intervention is necessary to find meaning in patterns even despite when research relies on machine learning. Another example is computational visualization, where sense making is a joint achievement by computers and humans (see e.g. Börner and Polley 2014).

Researchers trust software with their epistemic work even though reasoning with software is neither objective nor infallible. However, when assessing the suitability of third-party software for their own research goal, researchers can only rely on limited information. To fully grasp epistemic consequences of software use for research, we need to know which characteristics help to build trust and how this is different in different disciplines. To this end, our key concepts are transparency, reputation and compliance.

Open Science enthusiasts demand *transparency* to foster reliability (understood as reproducibility/replicability) of research results to enforce or justify trust in science: the 'FAIR for research software' principles even claim that 'source code is the most reusable form of software' (Chue Hong et al. 2021, p. 11). Moreover, Open Science is discussed as 'new "trust technology"' (Grand et al. 2012) and transparency is considered a decisive factor for trustworthiness, i.e. maintaining scientific integrity (Hardwicke and Vazire 2023; Aczel et al. 2019; Vazire 2017). So, transparency is the most central and most undisputed discursive figure in the Open Science movement.

The role of *reputation* for trust in research software has only rarely been explicitly addressed in the literature. However, Giddens (1990) argued that trust in abstract systems is only possible for lay persons through "access points" (the people involved in their creation or who are "responsible", p.83). Accordingly, scholars have pointed to provenance of data as essential to trust and reproducibility (Viglas 2013; Glavic 2021), more recently also of software (Dhruv and Dubey 2023). In general, researchers use reputation as a heuristic for the reliability of research results (Origgi 2017) up to the extent that quantitative performance indicators of individual researchers determine the perception of their research claims (Müller and de Rijcke 2017).

The *compliance* of research with disciplinary standards and agreed upon methods makes retrieved results acceptable and valid for the scientific peers and identifies a researcher as a member of a certain community or profession (Bowker and Star 1999; Fujimura 1988). While research is generally perceived as creative and uncertain, still standardized methods and (software) tools as well as predefined plans of action are common in modern research (Whitley 1985; Fujimura 1987, 1988; Latour 1987). Disciplinary standards define the rules of formal communication, hence provide 'literary technologies' (Shapin and Schaffer 1985; see also Csiszar 2020), while the clinical sciences have adopted completely standardized and pre-planned research designs (Keating and Cambrosio 2012). Particularly, using distinctive research software can show compliance with agreed upon 'theory-method packages' which solve problems in a way that is accepted within the community (Fujimura 1988). However, software use does not reproduce determined sequences of events and outcomes, but compliance with technical standards always involves creative deviation and skilful tinkering (Suchman 1985; Timmermans and Epstein 2010).

Overall, transparency, reputation and compliance differ in their individual-society-science configuration. Transparency refers to the user's ability to assess a given resource directly; reputation refers to the user's perception of a resource by the proxy of the context of production; and compliance refers to the user's perception of a resource by the proxy of its fit with (socially constructed) norms.

We expected to find disciplinary differences in how trust in research software is built, because practices of software use are epistemic practices. We implicitly understand discipline as culture here, therefore drawing from concepts such as 'academic tribes' (Becher 1989), 'epistemic cultures' (Knorr-Cetina 1999), 'regimes of knowledge production' (Marcovich and Shinn 2012) and 'epistemic regimes' (Gläser et al. 2018). Discipline as concept has been criticised as vague (Multrus 2004) and conflicting with organizational structures (Trowler 2014). However, we approach a researcher's discipline as a key demographic information obtained about them via a standardized online survey. A nested categorization helps us to keep in mind that disciplines as cultural categories are, of course, not mutually unrelated.

Within this article, we will elaborate on how we found the continuum explanatory between different disciplines-as-cultures for some but not all practices of trust in research software. We first describe how we operationalize trust practices as ten types of *considerations before use* in a standardized survey. With correspondence analysis, we investigate how these trust practices differ between disciplines.

2. Data & Methods

With Faulkner (2012), we derive 'attitudes of trust' towards software from the 'act of trust' that is using software. Hereinafter, we speak of users' *practices of building trust in software (for research)* as *trust practices*. For our operationalisation of trust practices, we consider the decision to use software as a critical point (in line with Solomon's (2005) 'decision vectors'). Accordingly, we asked participants in an online survey about a specific set of trust-related considerations which they might make before the decision to use software for research with the goal to fan out the different trust practices. We included trust practices regarding transparency, reputation and compliance which have not yet been investigated with a focus on software.

2.1. Data

We developed a survey module on 'Trust within Science' as our contribution to the DZHW Science Survey of 2023 which is a large trend survey among German researchers (Fabian et al. 2024). To investigate how trust in research software is built and how this is different among disciplines, we posed the following question to survey participants: “Which questions do you ask yourself before you use software for your research?” and provided a Likert-5-scale (end-verbalised with ‘always ask myself’ and ‘never ask myself’) for each of the considerations (see Table 1).

Table 1. This table shows the expressions which were provided in the survey (English translation from German original) under the umbrella question ‘Which questions do you ask yourself before you use software for your research?’ and how they map to their short form (‘type of consideration’).

Survey expression	Type of consideration
Is this software established in the field?	establishment (software)
Is the method implemented in the software established in the field?	establishment (method)
Which institution is behind this software?	institution
Where was this software released?	publication venue
Who recommended this software to me?	recommending person
Who made this software?	producing person
Is this software described in a comprehensible manner?	description
Can I check this software myself? Do I have the competence for this?	verifiability
Does this software stand up to my scrutiny?	verification
Does this software follow relevant disciplinary guidelines?	disciplinary guidelines

We obtained 1,702 observations from randomly sampled survey participants. We used 'How often do you use third-party software for your research?' to filter out those never using software from answering the questions on considerations[1]. After filtering, 228 to 317 observations remained for each of the five broader research fields as defined by the German Research Foundation and 12 to 118 observations for each of the twenty-two disciplines corresponding to the nested classification of disciplines.

[1] Participants who articulated that they never use third-party software, were excluded from answering questions on considerations already during the survey.

2.2. Methods

In our study, we used considerations of researchers before their decision to use software for research to better understand when software is perceived as trustworthy in different disciplines. The variables of interest are categorical (disciplines) and ordinal (relevance of considerations) which is why we chose multiple and simple correspondence analysis (Blasius 2001) as quantitative method tailored to a nominal level of measurement. For computation and visualization, we used R (R Core Team 2024) as well as contributed packages, especially FactoMineR (Le et al. 2016), factoextra (Kassambara and Mundt 2020), tidyverse (Wickham et al. 2019) and haven (Wickham et al. 2023). Also, the investigation was preregistered as a project in the Open Science Frameworkwhere we also provide supplementary material[2].

Our correspondence analysis of trust practices had two aspects. Firstly, we used multiple correspondence analysis with disciplines as supporting variables to investigate trust practices as in-person characteristic. Secondly, we conducted ten simple correspondence analyses to investigate whether the individual types of considerations correspond with the twenty-two disciplines.

With the results of simple correspondence analyses at hand, we interpreted a trust practice as disciplinary practice when the relevance of consideration was ordered along the first dimension of the correspondence plot *and* the explained variance was high. We tested the order with Spearman's rho: the Dim1-coordinates of the values {never, 2,3,4, always} were compared to {1,2,3,4,5} with the rank correlation mapping to the interval [-1,1]. If the absolute value of rho was close to 1, then we interpreted the trust practice as ordered in correspondence with disciplines. We considered the explained variance in Dim1 "high" if it was over 50% in simple correspondence plots for contingency tables of size 22x5 (twenty-two disciplines times five value expressions).

In general, distances in correspondence analysis plots must not be interpreted as ordinal or metric, but in our specific case, we only referred to one dimension and interpreted it as ordinal scale only if the Likert scale variable was ordered accordingly. Therefrom we then derived an order for the categorical variable.

[2] Preregistration: https://doi.org/10.17605/OSF.IO/8GM9K, project: https://osf.io/6jvbf/, supplement: https://osf.io/b4fw3/

2.3. Remarks and limitations

We acknowledge path-dependency in our findings as the co-evolution of method application and theory building shaped our analyses and thus, their outcome. The categorization into transparency-related, reputation-related and compliance-related considerations was derived as interpretation from the results of the simple correspondence analysis. Also, our analyses are based on a small (but representative) sample of the German researcher population. As research culture is intertwined with broader cultural contexts, organizational governance and national research policies, further research is needed as to whether our results apply to the global researcher population.

3. Results

Our study suggests that transparency-related trust practices differ between disciplines, that reputation-related trust practices are shared across disciplines, and that compliance-related trust practices fall into two categories: considerations on methods divide and considerations on guidelines unite disciplines. In this section, we firstly provide an overview and show that software use for research is prevalent across disciplines, that trust practices in research software are bimodal, and that multiple correspondence analysis indicates a sceptics-to-believers spectrum of attitudes towards research software. Second, we elaborate on disciplinary differences regarding these trust practices.

3.1. Prevalence of software use and of trust in research software

3.1.1. Software is used across disciplines

Software use is prevalent in all disciplines (see Table 2): only 11% (Engineering and Natural Sciences) up to 29% (Humanities) never use software at all. The non-users were filtered out during the survey and are thus not included in the results on considerations.

Table 2. The table shows the answers to the survey question ‘Do you personally use software for research?’. The totals and percentages of answers in the different answering options (‘Never’, 2, 3, 4, ‘Always’) are given for each of the five broader research fields.

research area	never		2		3		4		always		total	
Humanities	91	29%	61	20%	53	17%	37	12%	67	22%	309	100%
Social and Behavioural Sciences	79	19%	67	17%	54	13%	73	18%	133	33%	406	100%
Life Sciences	31	12%	43	17%	53	21%	58	23%	70	27%	255	100%
Natural Sciences	36	11%	70	21%	67	20%	76	22%	92	27%	341	100%
Engineering Sciences	30	11%	57	21%	59	22%	58	22%	64	24%	268	100%
not assigned	7	18%	7	18%	9	22%	7	18%	10	25%	40	100%

3.1.2. Considerations before use show a bimodal distribution

The individual considerations before software use vary in their importance for the users. Judged by the median[3] answer (see Table 3), the establishment of software and of the implemented method as well as a comprehensive description of the software are most important (each has median = 4), whereas the person producing the software is rather unimportant (median = 2).

However, each of the distribution of considerations before software use among all participants (see figure 1) is bimodal except for ‘the person producing the software’. The answer 'I never consider this.' stands out in all bar charts. These two patterns are unexpected for a Likert-scale-question and suggest further investigation.

[3] Please note that we have not computed arithmetic means, because the median is the appropriate measure of central tendency for ordinal data.

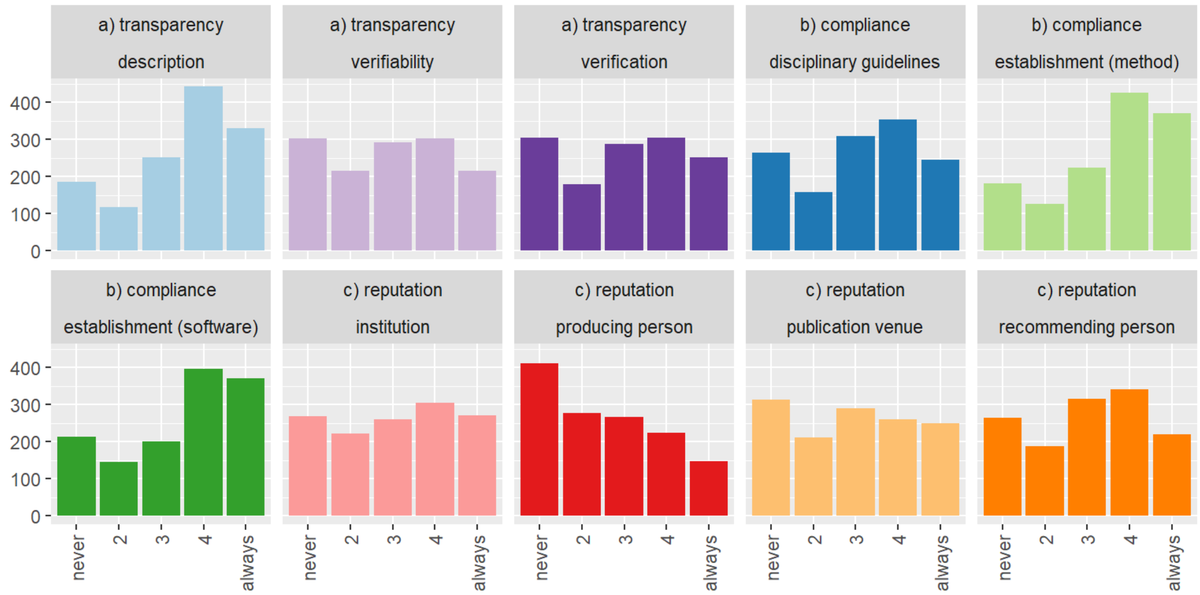

Figure 1. The figure shows the relevance of trust dimensions across disciplines as bar chart. For each of the considerations the category is given, which was assigned during the interpretation of study results, above the name of the consideration. The height of the bars expresses the total number of the answers in the different answering options (‘Never’, 2 3, 4, ‘Always’) to the question whether participants consider the respective attribute/characteristic of software before software use.

3.1.3. Considerations before use show disciplinary differences

The relevance of considerations before using software for research shows a high variance between fields (see figure 2). Overall, researchers from the Humanities consider the provided aspects less often than researchers from other fields. However, the mere counting of relevance does only tell us that disciplines build trust in software for research differently, but not so much how they differ. Thus, to pursue our research goal to find out how disciplines differ in their research practices, we turn to correspondence analysis.

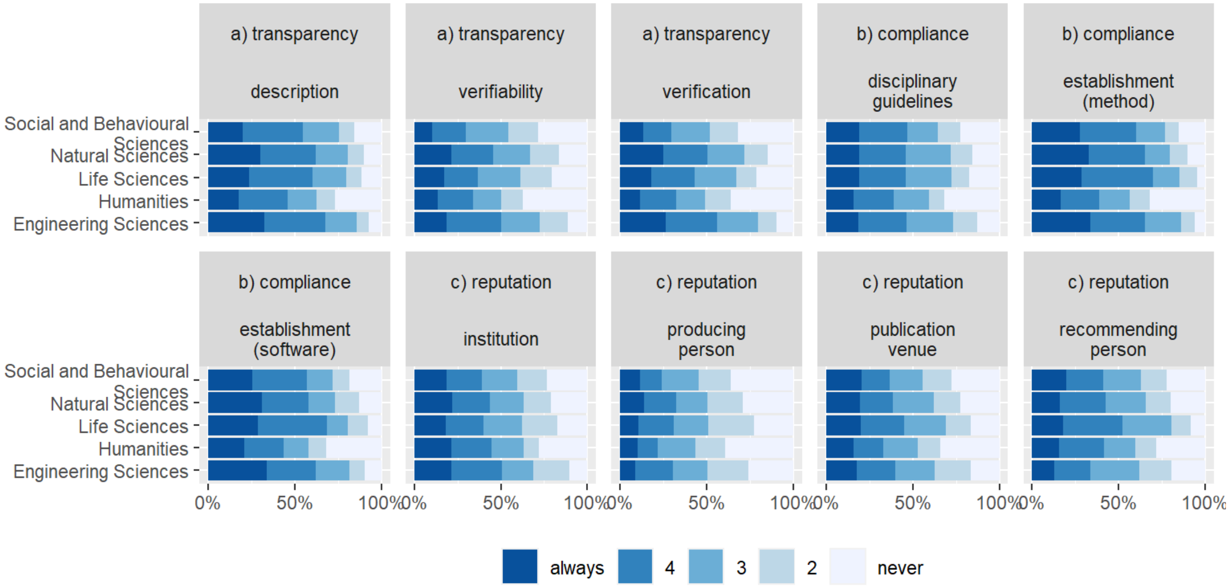

Figure 2. The figure shows the relevance of considerations before use across disciplines as stacked bar chart by fields.

3.1.4. Trust practices arrange on a continuum of sceptics and believers

The multiple correspondence analysis (figure 3) shows that our conceptualization of trust practices in ten dimensions is at least coherent as the barycenters of the relevance expressions of all ten considerations are perfectly ordered (|rho|=1)) along Dim1. This means that participants who never consider one of the software characteristics are also unlikely to consider other ones. We refer to this group as “believers” as they employ more unconditional modes of using research software. In contrast, participants who always consider one characteristic are likely to always consider another. We refer to this group as “skeptics” as they seem to scrutinize software or at least show risk awareness.

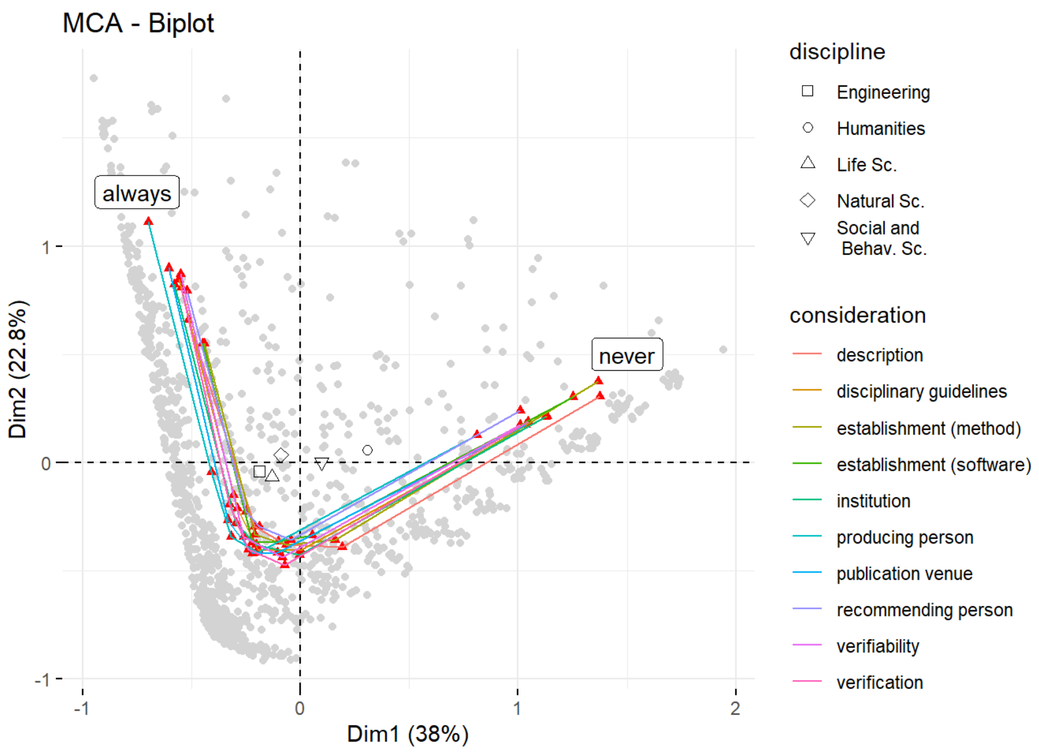

Figure 3. The figure shows the multiple correspondence analysis biplot of all ten trust dimensions together with coarse grained disciplines (dfg5) as supporting variables. Red triangles indicate the barycenters for the different answering options ('Never', 2,3,4, 'Always') to each question. Coloured lines indicate the distribution of said barycenters. Green triangles indicate the barycenters for each discipline. Grey dots indicate the individual observations.

The barycenters of the five broader research fields as supporting variables are all close to the coordinate origin. Whereas we might perceive a weak tendency for researchers in the Humanities of being believers and researchers in Engineering to being sceptics, still the inertia of the different fields is quite low as is the explained variance in Dim1.

3.2. Disciplinarity of considerations before use

The simple correspondence analyses for each of the ten considerations in five expressions with disciplines in twenty-two categories shows that some trust practices correspond more with disciplines than others. Accordingly, we sorted the practices into the following three categories: disciplinary practices, weak disciplinary practices and shared practices. As described above, we considered a trust dimension to be a disciplinary practice if it appeared ordered in the first dimension of the correspondence plot *and* the explained variance was high. We tested that with Spearman's rho which is the rank correlation measure of choice for ordinal variables. An overview of results is given in Table 3, and the correspondence plots in detail are provided with the supplementary material.

Table 3. The table shows the following for all dimensions of trust (i.e. consideration) under investigation: a) the category which was assigned during our interpretation of results, b) the name of the consideration, c) the median of the relevance of the consideration across all disciplines (in fine grained classification – dfg22), e) the extent to which the relevance of the respective consideration is ordered along Dim1 in the results of simple correspondence analyses with disciplines (|rho| close to 1 implies order, |rho| close to zero implied no order), f) the percentage of variance explained in Dim1 of simple correspondence analyses with disciplines, g) the type of practice which was assigned during interpretation of results, h) for (weak) disciplinary practices the names of the disciplines who most tend to never consider this dimension and i) who most tend to always consider this dimension.

category	**consideration**	**median**	**|rho|**	**% variance**	**type of practice**	**never**	**always**
a) transparency	description	4	1.0	60.19	disciplinary practice	Jurisprudence	Computer Sc., Systems & Electrical Eng.
a) transparency	verification	3	1.0	63.75	disciplinary practice	Literary Studies	Computer Sc., Systems & Electrical Eng.
a) transparency	verifiability	3	1.0	54.68	disciplinary practice	Literary Studies	Computer Sc., Systems & Electrical Eng.
b) compliance	establishment (method)	4	0.9	75.26	weak disciplinary practice	Literary Studies	Agric., Forestry and Vet. Med.
b) compliance	establishment (software)	4	0.8	64.35	weak disciplinary practice	Literary Studies	Agric., Forestry and Vet. Med.
c) reputation	recommending person	3	0.7	44.27	shared practice	NA	NA
c) reputation	producing person	2	0.3	47.14	shared practice	NA	NA
c) reputation	publication venue	3	0.3	47.78	shared practice	NA	NA
c) reputation	institution	3	0.1	47.43	shared practice	NA	NA
b) compliance	disciplinary guidelines	3	0.0	49.25	shared practice	NA	NA

3.2.1. Transparency-related considerations are disciplinary practices

We categorized the transparency-related practices as disciplinary practices, because considerations on a comprehensive description of the software, on verification by the user, and on verifiability by the user are definitely different in different disciplines – all three are clearly ordered along Dim1 (|rho|=1) with 54% to 64% explained variance. Broadly speaking, software-literacy-related considerations discriminate the (Computer) Sciences from the Humanities. Whereas many engineers and natural scientists consider software documentation, their own competence of verification as well as verification itself important, many researchers from the Humanities and the Social Sciences do not include such topics in their decisions to use software.

The consideration of a comprehensive description of the software highly corresponds with disciplines, while being very relevant over all (median=4). The 60.2 % of variance are explained in the first dimension and the relevance of this consideration is ordered along this dimension. The comprehensive description is most important for researchers in Computer Science and the Geosciences, whereas researchers in Jurisprudence and Literary Studies tend to never consider this before using software for research.

The consideration of the estimated competence of verifying the software oneself corresponds with disciplines in a similar way and is of medium relevance (median=3). The explained variance in the first dimension is 54.7 % and the relevance is ordered along the first dimension. This type of consideration is most famous among researchers in Computer Science and Mechanical and Industrial Engineering, whereas researchers in Literary Studies and the Social Sciences least consider this.

A similar picture shows for the actual verification of software which is also of medium relevance (median=3). The explained variance in the first dimension is 63.8 % and the importance of this consideration is ordered along the first dimension. Again, this consideration tells researchers in Computer Sciences and Mechanical and Industrial engineering apart from those in Literary Studies and Social Sciences.

In all three cases, the extreme coordinates for disciplines exceed the extreme coordinates for relevance of consideration. That means that the barycenter of Computer Science is further away from the coordinate origin than the barycenter of „always“ (in the same direction of Dim1) and the barycenter of Literary Studies is further away from the coordinate origin than the barycenter of „never“, accordingly.

3.2.2. Reputation-related considerations are shared practices across disciplines

We categorized reputation-related considerations as shared practices because they only show few, if any, disciplinary ordering ($|rho| \leq 0.7$) and the explained variance in Dim1 is fewer than 50% for all these considerations.

The person recommending a software could be a borderline case for the disciplinarity of trust practices, judged by $|rho|=0.7$, but in this case this means that the barycenter of the neutral expression (“3”) was shifted to an extreme position and the extreme expression “always” was shifted to the middle. Thus, this practice must not be classified as disciplinary practice but as shared practice. Besides, this type of consideration is the least relevant, judged by the median answer (median = 2).

Further reputation-related trust practices, i.e. the person producing the software, the publication venue of the software and the institution behind the software are even less sorted ($|rho| \leq 0.5$) and are thus also to be considered shared practices of trust.

3.2.3. Compliance-related considerations: methods divide, guidelines unite

Compliance-related practices fall into two categories. The establishment of the implemented method ($1 > |rho| >= 0.9$) has over 75% variance explained in Dim1, whereas the establishment of the software itself ($0.9>|rho|>0.8$) still has over 60%, so both are partly ordered close to the maximum and can be considered weak disciplinary practices. However, adherence to disciplinary guidelines is completely unordered ($|rho| = 0$) and is therefore a shared practice.

The correspondence analyses of compliance-related considerations yielded differences between generic and governmental trust practices. On the one hand, trust practices towards discursively emerging phenomena, i.e. (perceived) establishment of method and software, are discipline-specific while also being most relevant (median=4). On the other hand, trust practices towards disciplinary guidelines, which are supposed to be made explicitly consensual through a broader process, are shared practices and of medium relevance (median=3).

The consideration of the establishment of method is almost ordered – only the coordinates of “always” and “4” are switched, while still 75.3% of variance are explained in the first dimension. This consideration is most important in the Life Sciences whereas unimportant in the Humanities. As the relevance is not perfectly but almost ordered, we interpret this as a weak disciplinary practice.

The relevance of the establishment of software shows similar tendencies as the establishment of the method, but here, the two pairs “always”/“4” as well as “2”/”3” are switched within, but still 65.3 % of variance are explained. Again, the Life Sciences are more likely to always consider this, whereas researchers in the Humanities are more likely to never consider this. As the extrema tend to point into different directions, we still categorized this as a weak disciplinary practice, but this is also a borderline case.

In contrast, the relevance of considerations on whether the software meets disciplinary guidelines is very clearly a shared practice. It is totally unordered in correspondence with the twenty-two disciplines.

4. Conclusion

In our introduction, we have pointed out that across all academic disciplines, a) science with software is part of most researchers' everyday work, b) software as a configuration of research methods has epistemic consequences, and c) epistemic risks arise from transferring software from one research context to the other. Thus, researchers across fields must build trust in research software – accordingly, our aim was to understand disciplinary practices of trust in software.

Our study has shown that transparency, reputation and compliance differ in their relevance for building trust in research software. While transparency-related considerations divide the researcher population into 'Hard Sciences' and 'Soft Sciences' (see Snow 1959), reputation-related considerations unite researchers across disciplines. We gained mixed results on compliance-related considerations: while compliance with disciplinary standards in general is important across fields, the relevance of establishment of a software or the implemented method weakly distinguishes between fields.

Our multiple correspondence analysis revealed that researchers across disciplines arrange on a spectrum between always and never considering the ten trust dimensions when assessing software. With Giddens (1990) we know that trust is related to lack of information and risk awareness. When we relate this to the spectrum we found of sceptics and believers, we may conclude that the believers are rather unaware of (epistemic) risks (or consequences) of software use. While Giddens (1990) stated that "[r]espect for technical knowledge usually exists in conjunction with a pragmatic attitude towards abstract systems based upon attitudes of scepticism or reserve." (p. 90), nowadays it seems that a relevant portion of people are not sceptical towards software at all, judged by the fact that they never ask themselves any questions about software before use.

We found that the importance of transparency-related considerations is field specific: according to simple correspondence analyses, researchers from the Engineering Sciences tend to always assess research software before use based on its description, verification and verifiability whereas researchers from the Humanities tend to never consider these attributes – and Natural Sciences, Life Science as well as Behavioral and Social Sciences arrange between the extrema.

Thus, although transparency is called for by the Open Science movement in general (Grand et al. 2012; Hardwicke and Vazire 2023; Aczel et al. 2020; Vazire 2017) as well as specifically for software (Chue Hong et al. 2022; Barker et al. 2022; Lamprecht et al. 2020), we found that not all researchers from all disciplines can benefit from it, when it comes to software. This parallels earlier results about software users outside science by Zenkl and Griesbacher (2020) who found that "technology affinity" is an important factor for having trust in automated driving: in automated driving "security" is of central concern, and those who are enthusiastic about the opportunities which are opened up by technology are more likely to trust than others, despite knowing about the risks. Accordingly, we conclude that researchers can only benefit from transparency when they are able to use the given information for their assessment, which varies by field.

In contrast, the relevance of reputation-related considerations does not correspond with disciplines according to our analyses. Whether researchers assess software based on the recommending person, producing person, publication venue or institution is not related to the user's disciplinary background. Thus, across disciplines, researchers assess software by its context of production.

The widespread use of reputation as heuristic for whether to trust in others' research is not surprising per se as related mechanisms have been discussed in the literature (see Merton 1968; Origgi 2017; Müller and de Rijcke 17) for decades. To this background, contextual information is called for regarding research data and software (Viglas 2013; Glavic 2021; Dhruv and Dubey 2023).

However, reputation does not complement transparency according to our findings. It would have been plausible if those who can directly assess software (by description or verification) did not use reputation as a heuristic so much. That would mean that researchers from Engineering would use reputation less than researchers from the Humanities. But instead, from the fact that reputation-related trust practices did not arrange with disciplines, we deduce that reputation adds to transparency as a trust technology rather than complementing it.

Regarding compliance, we found two different types of correspondence with disciplines. Considerations on establishment are weak disciplinary practices and tend to be more relevant for researchers from the Life Sciences than the Humanities while the consideration of compliance with disciplinary guidelines is a shared practice.

It is not surprising that the assessment of software based on its compliance with disciplinary guidelines is independent from disciplines, as explicitly agreed upon ways of doing science permeate all scientific fields. In that sense, our results confirm the perspective of discipline-as-culture (see Becher 1989; Knorr-Cetina 1999). Here, disciplinary guidelines serve as codified norms that establish a many-to-one power dynamic and social 'regimes' that judge on community membership (see Marcovich and Shinn 2012; Gläser et al. 2018). However, disciplinary guidelines are not fixed but the emergence of new 'theory-method packages' initiates new research trends (Fujimura 1988; Galison 2010) with new guidelines to adopt. For example, the Life Sciences communities reacted to the problematization of the published record as part of the replication crisis with calls for reevaluating the dominant epistemic regime and its modes and standards for research practices (see Ioannidis 2005; Hosseini et al. 2022). These calls, however, led not only to the development of new formal standards or even bureaucracies such as preregistration or mandatory data publishing (Penders 2022), but also turned into distinct social movements, yet new cultures within research (see Peterson and Panofsky 2023).

However, the establishment of a particular software or method has shown to be more important in the Life Sciences than in the Humanities. We know that epistemic reforms can result in very specific standards and tools that become mandatory aspects of proper research, i.e. for writing biomedical reports (Altman and Simera 2016), and assume that software use follows this trend: using established software is then not only an act of finding 'the right tools for the job' (Clarke and Fujimura 1992), but signals being a skilled expert who has access to professional resources. This is even irrespective of the actual practice of software use during the actual research process which often remains inaccessible to outsiders such as readers of a paper, but in turn, exact reference and mentioning of the used software become crucial. Not surprisingly, disciplinary assignment has been found as a factor that explains the co-citation of software packages (Li and Yan 2018). A similar signifying role has been found in other accounts of standardization and the usage of procedural techniques, i.e. medical treatment protocols (Timmermans and Berg 2003; Timmermans and Epstein 2010).

Overall, the difference in individual-science-society configurations of trust practices plays out differently in different fields. Researchers from the Humanities do not benefit from transparency when assessing software and can only derive their trust in the software from its context of production (i.e. reputation-related characteristics). At the same time, derivative trust is open to all – also to software literate researchers – and empirically all disciplines alike use reputation-related considerations for software assessment. This makes affordances for derivative trust seem democratic whereas affordances for direct trust distribute support unevenly among disciplines. Being aware of this fact could prevent the research community from reinforcing the divide between the “two cultures” (Snow 1959), now, when manifold overarching research (data) infrastructures are developed. We recommend putting emphasis on shared practices when designing new research infrastructures.

We conclude that trust practices as epistemic practices are a promising research topic which could not be covered in full in our study. While we contextualized our findings with theoretical and empirical works of others, what drives disciplinary differences in practicing trust remains an unsolved puzzle. We intend to address this follow-up research question in the future and with additional data.

CRediT (https://credit.niso.org/)

JH: Conceptualization; Investigation; Methodology; Formal Analysis; Writing – original draft; Writing – review & editing; Visualization

AS: Conceptualization; Writing – review & editing; Methodology; Writing – original draft

NS: Conceptualization; Writing – review & editing; Methodology; Writing – original draft

Acknowledgements

The authors would like to acknowledge the team of the DZHW Science Survey, Gregor Fabian and Christophe Heger for conducting the survey and for helping to construct the module. The authors also thank Martin Reinhart for critical comments on the draft manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

Aczel, Balazs, Barnabas Szaszi, Alexandra Sarafoglou, Zoltan Kekecs, Šimon Kucharský, Daniel Benjamin, Christopher D. Chambers, et al. 2019. “A consensus-based transparency checklist.” Nature Human Behaviour 4, no. 1 (December): 4–6. https: //doi.org/10.1038/s41562-019-0772-6.

Altman, Douglas G, and Iveta Simera. 2016. “A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network.” Journal of the Royal Society of Medicine 109, no. 2 (February): 67–77. https://doi.org/10.1177/0141076815625599.

Barker, Michelle, Neil P. Chue Hong, Daniel S. Katz, Anna-Lena Lamprecht, Carlos Martinez-Ortiz, Fotis Psomopoulos, Jennifer Harrow, et al. 2022. “Introducing the FAIR Principles for research software.” Scientific Data 9, no. 1 (October). https://doi.org/10.1038/s41597-022-01710-x.

Becher, Tony. 1989. Academic tribes and territories: Intellectual enquiry and the culture of disciplines. Buckingham: Society for Research into Higher Education & Open Univ. Press.

Bechmann, Anja, and Geoffrey C. Bowker. 2019. “Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media.” 6:205395171881956. https://doi.org/10.1177/2053951718819569.

Blasius, Jörg. 2001. Korrespondenzanalyse. DE GRUYTER, December. https://doi.org/10.1515/9783486808346.

Bonde Thylstrup, Nanna, Mikkel Flyverbom, and Rasmus Helles. 2019. “Datafied knowledge production: Introduction to the special theme.” Big Data & Society 6, no. 2 (July): 205395171987598. https://doi.org/10.1177/2053951719875985.

Börner, Katy, and David E. Polley. 2014. Visual Insights: A Practical Guide to Making Sense of Data. https://doi.org/10.1108/oir-10-2014-0230.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. Sorting things out: Classification and its consequences. Cambridge, Mass. [u.a.]: MIT Press.

Chue Hong, Neil P., Daniel S. Katz, Michelle Barker, Anna-Lena Lamprecht, CarlosMartinez, Fotis E. Psomopoulos, Jen Harrow, et al. 2021. “FAIR Principles for Research Software (FAIR4RS Principles)” [in en], https://doi.org/10.15497/RDA00068.

Clarke, Adele E., and Joan H. Fujimura, eds. 1992. The right tools for the job: At work in twentieth-century life sciences. Includes bibliographical references and index. Princeton, New Jersey.

Csiszar, Alex. 2020. The scientific journal: Authorship and the politics of knowledge in the nineteenth century. Paperback ed. Literaturangaben. Chicago: The University of Chicago Press.

Davenport, James Harold, James Grant, and Catherine Mary Jones. 2020. “Data Without Software Are Just Numbers.” Data Science Journal 19. https://doi.org/10. 5334/dsj-2020-003.

Dhruv, Akash, and Anshu Dubey. 2023. “Managing Software Provenance to Enhance Reproducibility in Computational Research.” Computing in Science & Engineering 25, no. 3 (May): 60–65. https://doi.org/10.1109/mcse.2023.3314288.

Fabian, Gregor, Christophe Heger, and Merritt Fedzin. 2024. “Barometer für die Wissenschaft - Ergebnisse der Wissenschaftsbefragung 2023” (March). https://doi.org/10.31235/osf.io/ea5kr.

Faulkner, Paul. 2012. “The practical rationality of trust.” Synthese 191, no. 9 (May): 1975–1989. https://doi.org/10.1007/s11229-012-0103-1.

Friedman, Batya, and Helen Nissenbaum. 1996. “Bias in computer systems.” ACM Transactions on Information Systems 14, no. 3 (July): 330–347. https://doi.org/10.1145/230538.230561.

Fujimura, Joan H. 1987. “Constructing ‘Do-able’ Problems in Cancer Research: Articulating Alignment.” Social Studies of Science 17, no. 2 (May): 257–293. https://doi.org/10.1177/030631287017002003.

Fujimura, Joan H. 1988. “The Molecular Biological Bandwagon in Cancer Research: Where Social Worlds Meet.” Social Problems 35, no. 3 (June): 261–283. https://doi.org/10.2307/ 800622.

Galison, Peter. 2010. “Trading zones and interactional expertise: Creating new kinds of collaboration.” Chap. Trading with the enemy, edited by Michael E. Gorman, 25–52. Inside technology. Cambridge, Mass.: MIT Press.

Giddens, Anthony. 1990 [2015]. The consequences of modernity. Repr. Cambridge: Polity Pr.

Gläser, Jochen, Grit Laudel, Christopher Grieser, and Uli Meyer. 2018. Scientific fields as epistemic regimes: new opportunities for comparative science studies. 3-2018:35. TUTS - Working Papers. Berlin: Technische Universität Berlin, Fak. VI Planen, Bauen, Umwelt, Institut für Soziologie Fachgebiet Technik- und Innovationssoziologie.

Glavic, Boris. 2021. “Data Provenance.” Foundations and Trends® in Databases 9 (3–4): 209–441. https://doi.org/10.1561/1900000068.

Gramelsberger, Gabriele, Daniel Wenz, and Dawid Kasprowicz. 2024. “Understanding and Analysing Science’s Algorithmic Regimes: A Primer in Computational Science Code Studies.” In Algorithmic Regimes, 57–78. Amsterdam University Press, December. https://doi.org/10.1515/9789048556908-003.

Grand, Ann, Clare Wilkinson, Karen Bultitude, and Alan F. T. Winfield. 2012. “Open Science: A New “Trust Technology”?” Science Communication 34, no. 5 (September): 679–689. https://doi.org/10.1177/1075547012443021.

Hardwicke, Tom E., and Simine Vazire. 2023. “Transparency Is Now the Default at Psychological Science.” Psychological Science (December). https://doi.org/10.1177/09567976231221573.

Hosseini, Mohammad, Enric Senabre Hidalgo, Serge P.J.M. Horbach, Stephan Güttinger, and Bart Penders. 2022. “Messing with Merton: The intersection between open science practices and Mertonian values.” Accountability in Research 31, no. 5 (November): 428–455. https://doi.org/10.1080/08989621.2022.2141625.

Ioannidis, John P. A. 2005. “Why Most Published Research Findings Are False.” PLoS Medicine 2, no. 8 (August): e124. https://doi.org/10.1371/journal.pmed.0020124.

Kassambara, Alboukadel, and Fabian Mundt. 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7.

Keating, Peter, and Alberto Cambrosio, eds. 2012. Cancer on trial: Oncology as a new style of practice. Includes bibliographical references and index. - Print version record. Chicago.

Kinder-Kurlanda, Katharina, and Miriam Fahimi. 2024. “15. Making Algorithms Fair: Ethnographic Insights from Machine Learning Interventions.” In Algorithmic Regimes, 309–330. Amsterdam University Press, December. https://doi.org/10.1515/ 9789048556908-015.

Knorr Cetina, Karin. 1999. Epistemic Cultures: How the Sciences Make Knowledge. Harvard University Press, December. https://doi.org/10.4159/9780674039681.

Lamprecht, Anna-Lena, Leyla Garcia, Mateusz Kuzak, Carlos Martinez, Ricardo Arcila, Eva Martin Del Pico, Victoria Dominguez Del Angel, et al. 2020. “Towards FAIR principles for research software.” Edited by Paul Groth, Paul Groth, and Michel Dumontier. Data Science 3, no. 1 (June): 37–59. https://doi.org/10.3233/ ds-190026.

Latour, Bruno. 1987. Science in action: How to follow scientists and engineers through society. Hier auch später erschienene, unveränderte Nachdrucke. Cambridge, Massachusetts: Harvard University Press.

Le, Sebastien, Julie Josse, and Francois Husson. 2008. “FactoMineR: A Package for Multivariate Analysis.” Journal of Statistical Software 25 (1): 1–18. https://doi.org/10.18637/jss.v025.i01.

Li, Kai, and Erjia Yan. 2018. “Co-mention network of R packages: Scientific impact and clustering structure.” Journal of Informetrics 12, no. 1 (February): 87–100. https://doi.org/10.1016/j.joi.2017.12.001.

Marcovich, Anne, and Terry Shinn. 2012. “Regimes of science production and diffusion: towards a transverse organization of knowledge.” Scientiae Studia 10 (spe): 33–64. https://doi.org/10.1590/s1678-31662012000500003.

Merton, Robert K. 1968. “The Matthew Effect in Science.” Science 159 (3810): 56–63.

Morozov, Evgeny. 2014. To save everything, click here: The folly of technological solutionism. Paperback 1. publ. New York, NY: PublicAffairs.

Müller, Ruth, and Sarah de Rijcke. 2017. “Thinking with indicators. Exploring the epistemic impacts of academic performance indicators in the life sciences.” Research Evaluation 26, no. 3 (July): 157–168. https://doi.org/10.1093/reseval/rvx023.

Multrus, Frank. 2004. Fachkulturen.

Origgi, Gloria. 2017. Reputation: What It Is and Why It Matters. Edited by Stephen Holmes and Noga Arikha. Princeton University Press, November. https://doi.org/10.2307/j.ctvc77bzk.

Penders, Bart. 2022. “Process and Bureaucracy: Scientific Reform as Civilisation.” Bulletin of Science, Technology & Society 42, no. 4 (September): 107–116. https://doi.org/10.1177/02704676221126388.

Peterson, David, and Aaron Panofsky. 2023. “Metascience as a Scientific Social Movement.” Minerva 61, no. 2 (April): 147–174. https://doi.org/10.1007/s11024-02309490-3.

R Core Team. 2024. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Shapin, Steven, and Simon Schaffer. 1985 [2011]. Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life. Edited by Simon Schaffer. Princeton Classics Ser. v.109. Princeton: Princeton University Press.

Snow, C.P. 1959. THE REDE LECTURE, 1959. Cambridge.

Solomon, Miriam. 2005. Social Empiricism. The MIT Press.

Suchman, Lucy A. 1985 [1999]. Plans and situated actions: The problem of human-machine communication. Learning in doing: social, cognitive, and computational perspectives. Cambridge [u.a.]: Cambridge Univ. Press.

Timmermans, Stefan, and Marc Berg. 2003. The Gold Standard: The Challenge Of Evidence-Based Medicine. Temple University Press.

Timmermans, Stefan, and Steven Epstein. 2010. “A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization.” Annual Review of Sociology 36, no. 1 (June): 69–89. https://doi.org/10.1146/annurev.soc.012809.102629.

Trowler, Paul. 2014. “Academic Tribes and Territories: the theoretical trajectory” [in en]. Austrian Journal of Historical Studies, Bd. 25 Nr. 3 (2014): Die ”Stämme” der Akademie. https://doi.org/10.25365/OEZG-2014-25-3-2.

Vazire, Simine. 2017. “Quality Uncertainty Erodes Trust in Science.” Collabra: Psychology 3, no. 1 (January). https://doi.org/10.1525/collabra.74.

Viglas, Stratis D. 2013. “Data Provenance and Trust.” Data Science Journal 12 (0): GRDI58–GRDI64. https://doi.org/10.2481/dsj.grdi-010.

Whitley, Richard. 1985 [2006]. The intellectual and social organization of the sciences. 2. ed., reprinted. Includes bibliographical references and index. - Previous ed.: Oxford : Clarendon, 1984. Oxford [u.a.]: Oxford Univ. Press.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” Journal of Open Source Software 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Evan Miller, and Danny Smith. 2023. haven: Import and Export ’SPSS’, ’Stata’ and ’SAS’ Files. R package version 2.5.2.

Zenkl, Thomas, and Martin Griesbacher. 2020. “Trust in Automated Vehicles from a Sociological Perspective.” Jusletter-IT, no. 27-Mai-2020, https://doi.org/10.38023/962185e1-c563-4193-a56a-c73280c86461.