IJDC | *Conference Paper*

Factors Influencing Perceptions of Trust in Data Infrastructures

Katharina Flicker
TU Wien

Bettina Kern
TU Wien

Andreas Rauber
TU Wien

Fajar J. Ekaputra
WU Wien

Abstract

Trust is an essential pre-condition for the acceptance of digital infrastructures and services. Transparency has been identified as one mechanism for increasing trustworthiness. Yet, it is difficult to assess to which extent and how exactly different aspects of transparency contribute to trust, or potentially impede it in cases of overwhelming complexity of the information provided. To address these issues, we performed two initial studies to help determining the factors that influence or have impact on trust, focusing on transparency across a range of elements associated with data, data infrastructures and virtual research environments. On one hand, we performed a survey among IT experts in the field of data science focusing on quality aspects in the context of re-using and sharing open source software, assessing issues such as the need for documentation, test cases, and accountability. On the other hand, we complemented this with a set of semi-structured interviews with senior researchers to address specific issues of the degree of transparency achievable with different approaches. They include, for example, the amount of transparency we can achieve with approaches from explainable AI, or the usefulness and limitations of data provenance in determining the suitability of data for reuse and others. Specifically, we consider mechanisms on three levels, i.e. technical, process-oriented as well as social mechanisms. Starting from attributes of trust in the "analogue world", we aim to understand which of these can be applied in the digital world, how they differ, and what additional mechanisms need to be established, in order to support trust in complex socio-technological processes and their emergent results when the traditional approaches cannot be applied anymore.

Submitted 26 September 2023 ~ *Accepted* 21 February 2024

Correspondence should be addressed to Katharina Flicker, Research Unit Data Science, TU Wien, Favoritenstraße 9-11/194-04, 1040 Vienna, Austria. Email: katharina.flicker@tuwien.ac.at

This paper was presented at the International Digital Curation Conference IDCC24, 19-21 February 2024

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

Data-driven decisions and technological solutions increasingly permeate all aspects of our life. Science is expected to produce the foundation for these decisions and technologies, whether they come in the form of decision support to tackle societal challenges (e.g. during the COVID pandemic), result in novel technologies (e.g. ChatGPT or self-driving cars), or entire platforms showing emergent behaviour (e.g. social media).

While scientists strive for perfection in all their endeavours, the increasing complexity of scientific processes makes it hard for researchers to fully comprehend each and every component of the process, ensure the absence of subtle errors and thus guarantee the correctness of the final insights gained or solutions developed. In data-driven sciences, data that has been pre-processed in slightly different ways is being integrated from a range of sources. It is being processed by reusing code from Open Source Software repositories that are compiled into increasingly complex analytical processes. Understanding, testing and assuring the quality of all components is next to impossible. Transparency, i.e. extensive and openly available information and documentation on the data objects, code, and processes, is essential to be able to understand and evaluate the quality, i.e. fitness for the respective purpose, of these components, preferably in an automated manner.

However, as data is being integrated from a range of sources, understanding the complete provenance, and thus ultimately the fitness for purpose, is turning into a paramount challenge. Similarly, while (open source) code reuse is speeding up the scientific process, ensuring its correctness becomes increasingly hard, thus affecting the confidence that we as scientists can have in the final insights produced.

It is still not fully understood, which types of information, ranging from technical documentation (such as quality checks applied, software tests, code documentation, digitally signed provenance chains) via process-specific information (actors involved, their qualifications, principles such as pair programming or coding guidelines, or the application of formal review processes, certificates and compliance to standards) to social indicators (such as the number of citations, downloads, or the prominence of an institution) contribute to the perception of trust. Furthermore, initial studies indicate that these criteria differ (at least) across disciplines and seniority levels.

In this paper, we review some of the challenges inherent in determining trust in complex systems with a particular focus on challenges in research infrastructures and the complexity inherent in today's scientific processes. We start with a brief review of related work on key concepts of trust to conceptualize trust in a way that makes it tangible. Second, we describe the study design and its limitations. Third, we provide a discussion of some key insights from the two studies.

Related Work

There is a sophisticated body of literature regarding trust in e.g. organizations, data, and technologies, as well as on research on open source and the concept of trust itself:

Regarding the increase of (digital) data, questions as to how its value may be further developed through trust arise not only in scientific contexts but also in everyday life (Pink, Lanzeni, & Horst 2018). The relationship between trust in a technology and the resulting use is receiving intensive attention, studying, for example, how trust affects the use of ChatGPT among adults in the US (Choudhury & Shamszare 2023), or how privacy concerns as well as currency-related trust influences the individual's willingness to use Central Bank Digital Currency (Tronnier, Harborth, & Hamm 2022). Others, in turn, explore the nature of trust as a central component of the interaction between people and AI (Jacovi et al., 2021), or in IT in

general (McKnight et al., 2011). A very nice example of the duality of approaches, namely technically as well as human perception based, is provided in the automotive sector, with massive research focusing on technological aspects of safety (mean-time to failure, redundancies, technical edge cases) (Kopetz, 2022) and the equally important consideration of end user perception (Nastjuk et al., 2020).

From the onset, trust has been a crucial characteristic for digital repositories, leading to the emergence of initiatives such as the Data Seal of Approval and the Nestor Catalogue of Criteria for Trusted Digital Repositories[1], which merged to form the now well-established Core Trust Seal[2]. Complementing these are extensive requirements listed in standards such as ISO16363 certification of Trustworthy Digital Repositories[3]. All these criteria catalogues are based on extensive evidence collected over many years of operating repository infrastructures. They are a solid basis for documenting and confirming the solid operations of data repositories. Yet, this constitutes only one of the many factors required to trust code, data, or other resources in R&D projects. Repositories, for example, will frequently not provide any assertions on the (original semantic) quality of the content ingested and disseminated, being concerned with the preservation of the accessibility of the digital objects. Depending on the resources available and the domain-specificity, some repositories will also include semantic quality assurance on data or code hosted by them. Yet, as quality (defined as "fitness for purpose") is inherently dependent on the actual intended use, such assertions are impossible to make on a general level and thus need to be verified (and, hence, verifiable) by the respective users. Thus, understanding the information needs, how to collect the required elements and how to communicate these is essential if we want to extend the concept of trust to insights generated in research environments via complex processes happening therein.

One key component in many research processes is the software used for processing data. Research on open source software has thus emerged in different academic disciplines, focusing on topics such as the motivation of contributors, innovation processes and competitive dynamics (von Krogh & von Hippel, 2006), all of which are essential to understand the trustworthiness of the code to be used in a research project. As for motivations, early research highlights the contributors' motives such as fun, enjoyment, reputation building, learning, private use value as well as private benefits (that is, for example, demonstrating programming skills to potential employers). Additionally, the need for code and improvements drives participation. Once that need is satisfied, however, most leave the project. The few who stay become 'hobbyists'' taking on crucial tasks and are therefore vital for the long-term viability of the software code (Shah, 2006). Newer research broadens its perspective by looking at firms' participation, community participation and the technical design of open source software to see how these impact motivations (von Krogh & von Hippel, 2006). In our study we will take a closer look at the characteristics of such open source software projects that contribute to the perceived trustworthiness of the code. We will use this to determine recommendations on the information to be provided by repositories providing such software for use in research (as opposed to, say, simple "for-fun" projects which might face lesser trustworthiness requirements, or applications to be used outside (relatively closed) lab environments, which might face much higher scrutiny with respect to security requirements).

Researching innovation processes tends to go hand in hand with researching governance and organization in, for example, open source software projects: The development of open source software depends on the contributions of many, who are frequently not being paid for the work they do. Therefore, distinct governance mechanisms and structures emerged that come with specific challenges (von Krogh & von Hippel, 2006). Research may thus focus on how work is organized in distinct governance structures preventing the 'forking' of a software project into many versions of the code base (Kogut & Metiu, 2001), or how products were governed by communities so that they would remain in the commons (O'Mahony, 2003). Understanding the

[1] DIN 31644, https://www.din.de/de/mitwirken/normenausschuesse/nid/veroeffentlichungen/wdc-beuth:din21:147058907

[2] https://www.coretrustseal.org/

[3] https://www.iso.org/standard/56510.html

impact of such governance processes, the balance between more flexible, highly dynamic approaches vs. more controlled and restricted approaches, on the perception of trustworthiness will be essential to assist both the owners of the codebase as well as users in providing and selecting the most suitable components in specific study settings and are hence crucial in designing infrastructure components accordingly.

Many of the aspects above can be captured and communicated via specific metadata elements. These may then serve as indicators of quality and are already a familiar sight in the digital world from social media to professional settings: Star-ratings, impact factors, download counters, interface design characteristics, the institution providing the information, etc. are used as indicators of quality or trustworthiness. Yet, in a highly competitive world, such indicators are also prone to become targets of manipulation, with studies and services on "optimization" of such parameters abounding (e.g. manipulated/contracted reviews, downloads via bot-networks, likes-for-sale, citation networks, h-index manipulation via article mergers or planted citations, etc.) (van Bevern et al., 2020; Ertz 2022). Competition is omnipresent in the research environment, complementing cooperation to strive for the best results. We thus need to better understand the correlation between trust, trustworthiness and (non-refutable) indicators supporting these. Crucially, traditional mechanisms of determining the trustworthiness of material, such as personal knowledge of the individual, the research group, or the in-depth inspection of code, data and its provenance, or surrogates such as trusting the (peer) review processes are being eradicated given the increase in data volumes, process complexity, interdisciplinarity, and global diversification, forcing us to identify (and understand) new indicators.

We are looking at trust in a very specific context, namely that of trust in research (including processes, results and infrastructures) as a practice of evidence-based knowledge production. In this context, trust does not need to be complete. It needs to be warranted and evidence-based. Trust, however, is a complex concept that has been the focus of philosophical debates. Still there is no clear definition (McLeod, 2021). In the context of this study, the challenge therefore is to define trust in a way to be properly used with respect to (in a broader sense) research and (more narrowly) for the sharing and re-use of data or open source code. The following explanations are therefore considering two issues: First, can trust be warranted and - if so - under which conditions? Second, can trust be rational and evidence-based?

It appears philosophers are able to agree on when trust is warranted. Warranted, in this case, also refers to justified, well-grounded and plausible: Trust is plausible, if the conditions required for trust exist. That is, for example, at least some optimism about the trustee's competence in one area of expertise. It is also plausible if all needed circumstances that enable a trustor to develop trust in the first place are given (McLeod 2021).

Well-grounded trust is based on the trustee being trustworthy. This is based on the assumption that trust and trustworthiness are not the same thing. Rather trust is an attitude towards people (or things), while trustworthiness is - in this respect - a property of people (or things). To be trustworthy, a trustee has to be competent, willing (to do X if entrusted with X) and reliable. It should also be noted here that the trustor has to accept being vulnerable (McLeod, 2021) (to betrayal (Hawley, 2014)) as well as the risk of being let down (McLeod, 2021).

Trust may be justified because some value might emerge from trust. At times trust itself is valuable. The value of trust relates to both intrinsic as well as instrumental trust. The former can be seen as a sign of respect for others. The latter can be viewed as social and individual "goods" benefiting the trustor, the trustee and society in general. Examples include enhanced cooperation (trust makes cooperation less complicated), meaningful relationships, moral or scientific knowledge because both depend on the testimony of others (we need to trust in the expertise of others because no one can learn all there is to know), autonomy (as it can only be exercised in social environments where people and / or organisations can be trusted), strong social networks and morality as a cooperative activity (as people need to trust to be moral) (McLeod, 2021).

It should also be noted here, however, that unjustified trust may "(...) leave us open to abuse, terror, and deception" (McLeod, 2021, p.17). Therefore, theorizing about distrust is essential not only because of this but also because conceptualizing mistrust facilitates a better understanding of trust (Hawley, 2014). However, due to the wider scope of work that such a perspective would entail, it makes sense to first think about trust and, in potential follow-up studies, also about mistrust, and to take this into account in the data collection.

Whether trust can be rational is a difficult question due to the nature of trust. After all, it implies the risk of being let down. Minimising the risk could be interpreted as an indication of a lack of trust and lead to the elimination of already existing trust. Those who trust might also tend to ignore indications of, for example, abuse or misuse of trust. It is therefore not surprising that there are different views on this issue (McLeod, 2021). As already mentioned earlier, we are looking to define trust for a very specific context - namely that of (data) science as a rational, reasoned knowledge-producing activity. Trust must not be complete in this case, nor should researchers be expected to trust blindly when their daily work is supposed to be based on both reason and evidence. Hence, the following explanations are looking at perspectives that are based on the assumption that trust can indeed be rational, which leads to the question to which extent trust may be rational. There are two main perspectives on this: an internalist and an externalist one. The latter states that trust does not require any particular comprehensible reasons (McLeod, 2021).

The internalist point of view, however, is based on the assumption that for trusting, the trustor must have good reasons that are grounded in evidence of somebody or something being trustworthy. New evidence leads to corrections and updates thereof (McLeod, 2021). This definition of trust is also the kind of trust that (in a best-case scenario) should be merited by research processes, infrastructures and results. We will see, however, that when looking at actual research practices this is, unfortunately, not always the case.

Methodology

To maintain an explorative character in the research and to cover aspects that we ourselves might not have thought of, we decided to conduct semi-structured interviews with researchers from different scientific disciplines. Additionally, a survey consisting of both open and closed questions was conducted at the TU Wien among (partly soon-to-be) data scientists to reach a larger number of people.

The semi-structured interviews were based on an interview guide and consisted only of open questions. They were conducted globally via (i) the EOSC Support Office Austria Working Group for Researcher Engagement in Austria, (ii) as part of an interview series with ERC Grantees and Nobel Laureates on future research environments (EOSC Secretariat) and (iii) via a collaboration with the Universiti Teknikal Malaysia Melaka (UTeM) supported by ASEA UNINET, resulting in a total of 25 interviews as well as one focus group interview. These were recorded and transcribed. With the consent of our interviewees, some of them were edited and published on Zenodo (Campiglio et al., 2020). To support the analysis of the large amounts of text, the software ATLAS.ti was used.

The survey was conducted at TU Wien via TUWEL, TU Wien's e-Learning platform over a period of 4 weeks, attracting 90 participants. The quantitative analysis of the survey was carried out automatically via this platform. Answers to open questions were also analysed via ATLAS.ti. The chosen method of (qualitative) analysis is inductive categorization (Mayring, 2013). The anonymized data is shared via Zenodo[4].

The questionnaire consisted of 35 main questions (including both open and closed ones) divided into six thematic blocks. The first was aimed at gathering information about practical experience (4 questions). The second part asked about actual practices related to code sharing (12 questions), while the third block (5 questions) focussed on reuse, the fourth on quality

[4] https://zenodo.org/records/11176945

assessment (7 questions) and the fifth on trust (5 questions). Finally, the questionnaire was completed with a question on accountability and the opportunity to share feedback or other thoughts (2 questions)[5].

Regarding the first topic, we not only asked about the job title and experience in years, but also for a description of the activities in order to be able to better contextualize and interpret subsequent answers (Q1-Q4).

As far as code sharing practices were concerned (Q5-16), information on whether survey participants shared code (snippets) was collected (Q5). In case they did, further questions aimed at eliciting details like the type of code (Q6) and additional information (Q12) or updates they were sharing (Q13, Q14) as well as testing procedures they were applying (Q9, Q10, Q11). Additionally, a few questions inquired into attitudes with respect to such practices - among them motivations for (not) sharing code (Q7), opinions on sufficient testing (Q15), pressure felt when sharing (Q8) and confidence in the quality of one's own code (Q16).

In the context of reuse (Q17-21), we asked for actual practices (whether and how often) and for examples of libraries and forums (Q17) as well as a list or description of the criteria that were decisive for this choice (Q18, Q19). We also wanted to know whether and to what extent the survey participants believed the code to be correct (Q20, Q21).

Eventually, we addressed quality assessment (Q22-28), looking into what survey participants believed to be good quality indicators (Q22, Q23) as well as into why they thought so (Q24, 25). We also explored whether they tested before reuse in order to facilitate quality (Q26). In due course, we asked for quality assessment procedures they did not have time for (Q27, Q28).

The fifth section of the survey tries to culminate the perspectives on trust in one's own code to be shared and trust in the code re-used from others into the amount of trust participants were willing to attribute into the results they produce (Q29-34). We phrased this in a context of accountability, i.e. determining how high a fine (in Euro) would they accept to pay for faulty results (Q31), how they would justify that liability (Q32) and how important the loss of reputation due to erroneous results would be perceived (Q33), and whether there was a difference between their professional employment settings and tasks in a university context (Q34).

At the time of writing this paper a bit more than one third of the survey responses has been analysed qualitatively. Methodologically, the analysis of an initial subset allows us now to formulate hypotheses on which we aim to obtain feedback and extension. We aim to take these into consideration when evaluating the remaining data to test the hypotheses. These insights will be used subsequently to re-design the questionnaire and derive scenarios for further semi–structured interviews to dive deeper into the relationships and trust factors identified.

Applying inductive categorization we iterate through a subset of the questionnaires, identifying codes – such as '(Active) Community' or 'Documentation' – that summarize the statements and assigning these codes to the respective statements. A text passage can be assigned more than one code. That way it is possible to see what kind of codes tend to go together revealing, for example, co-occurrences, the frequency of (co-)occurrences and the distribution of codes. This, in return, may show the extent that different social phenomena – like the idea of what indicates good quality and therefore constitutes trust in open source software – relate with each other.

Key Findings and Discussion

Both the interviews and the survey provide interesting insights into the factors that contribute to trust, or the limits of what can be achieved. Some of the interim key findings, forming the basis for subsequent analysis, are summarised below.

[5] The questionnaire on which this paper is based as well as a revised version from 2023/24 are available at: https://zenodo.org/records/10626345

An uneasiness concerning mechanisms to determine data quality was expressed especially when the data does not come from a standardized and highly recognized source, an established repository, or from one's own collection. Nevertheless, about two-thirds of participants do not test code that they want to reuse. This becomes particularly interesting when compared with answers to the question of whether penalties are accepted (Q31, Q32). In this context, it is repeatedly pointed out that code can always be buggy, and that the responsibility for testing and quality checks therefore lies with those who reuse code.

However, it should be mentioned here that no clear result is yet available. First, the remaining questionnaires still need to be analysed to verify this hypothesis. Second, it still needs to be determined whether the two (non-testing) thirds are also the people who reject fines (Q31, Q32) on the basis of the argument just mentioned, or if they are other participants.

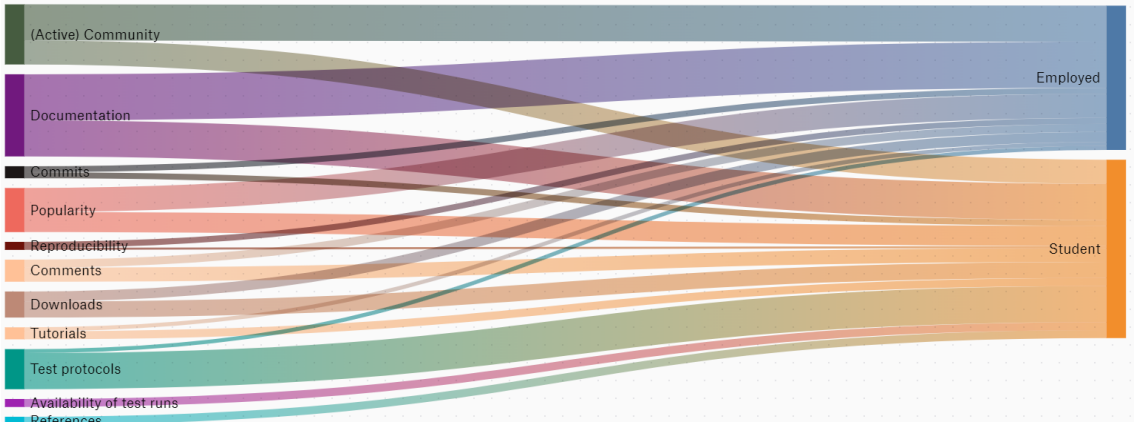

Figure 1. Sankey Diagram. On the left-hand side, from top to bottom: (Active) Community, Documentation, Popularity, Reproducibility, Commits, Comments, Downloads, Tutorials, Test Protocols, Availability of test runs, References. On the right-hand side, Employed (top) vs. (only) Student (bottom).

Against this background, the survey also revealed a strong cognitive disconnect between the trust participants had in their own code and their behavior when re-using components contributed by others: Only a minority is sure (1.11%) or very confident (17,78%) that their code is bug-free. Likewise, they are not confident about code they are re-using (it should be noted that 94,45% of the participants re-use code): The majority of 35,56% hopes the code they re-use is bug-free, while 33,33% are somewhat confident and only 32,22% are very confident; 4,44% are not confident at all. As mentioned before, 70% do not test code before reusing it. This resulted in a rather low confidence in the correctness of the results produced (mean 67 on a scale of 1-100) - but still low willingness to accept liability for the results they produced. When asked about the level of accountability they were willing to accept, a significant fraction of the participants indicated that this was dependent on the circumstances, linking the accountability to potential remuneration received. Research and data analytics as part of the academic programme and code contributions made "voluntarily" were not perceived to merit much accountability, whereas a commercial setting would lead to higher levels of accountability accepted. However, this could also be due to a misunderstanding of the question: while the actual question is aimed at research in general, the participants in the study probably understood this more in terms of course assignments.

We also identified unexpected quality indicators used by the participants. While some of the most widely used code for running scientific experiments in professional settings has not been updated for years (with researchers requesting continued operations of old compute environments to ensure identical operational settings that are fully documented and well-understood), study participants from the computer sciences domain indicated that the frequency

of updates and the recency of the latest update were indicators of highly active, and thus high-quality projects. This could be due to a lack of seniority (study participants were mostly students of the computer sciences with an average of 2.86 years of professional experience) as well as to peculiarities of the discipline. The latter relates to code / algorithms being mainly "tools to be used" outside the computer sciences, whereas within the computer sciences they are a key element of the research in and by itself.

Apart from that, quality indicators were ranked as follows in the quantitative survey (the number and percentage in brackets indicates how many participants consider this indicator relevant): Documentation and tutorials (75: 83,33%), number of downloads (59: 65,56%), comments in the forum (46: 51,11%), number of contributors (37: 41,11%), availability of test runs and test protocols (36: 40%), frequency of commits (32: 35,56%), number of types of references to the project / code (28: 31,11%), age of the project (23: 25,56%), type of commits / commit messages (22: 24,44%), most recent commits (19: 21,11%) and others (4: 4,44%). These quality indicators were suggested as possible answers in the survey, whereby participants could select more than one indicator.

The qualitative analysis also identified some - similar, but not identical - quality indicators. These are: an active community, the popularity of libraries / repositories / fora, the availability of test protocols, comments of users, the number of downloads, the frequency of commits, tutorials, the availability of test runs, references and reproducibility.

The Sankey diagram in Figure 1 shows these indicators on the left-hand side. On the right-hand side, the participants are divided into those students who also work in a subject-relevant area (Employed) and those who do not (Student). The width of the bars shows how frequently certain codes / quality indicators or groups of people (i.e. employees and students) occur: the wider the bar, the more frequently. For example, documentation is mentioned more frequently as an indicator of quality than reproducibility.

The stream-shaped connections between the indicators on the left and the groups of people on the right allow a rough estimate of which group has named which indicators and how often. For example, the availability of test protocols was mentioned more often by students. Surprisingly, the availability of test runs and references have so far only been mentioned by the student group and not by those also working in data science related domains. This might hint at a strong academic focus on these quality assurance mechanisms that is not mirrored in the enterprise sector (although it should be kept in mind that only one third of the questionnaires have been analysed qualitatively so far to formulate the hypotheses).

Documentation is the most important quality indicator in terms of both quantitative and qualitative analyses that leads the survey participants to trust whatever they are re-using. In principle, this corresponds to the idea of evidence-based trust. However, when taking a closer look at what they

what they consider when referring to „documentation": As Figure 2 reveals, documentation is strongly associated with evidence for „good quality", „popularity" and „transparency". It also revelas the broad range of documentation types

This, however, might be due to how we asked the questions. For future research, survey questions might therefore need to be rephrased. Additionally - if we want to make sense of why documentation is deemed crucial for quality - we need to put a focus on documentation in terms of questions before launching the survey again.

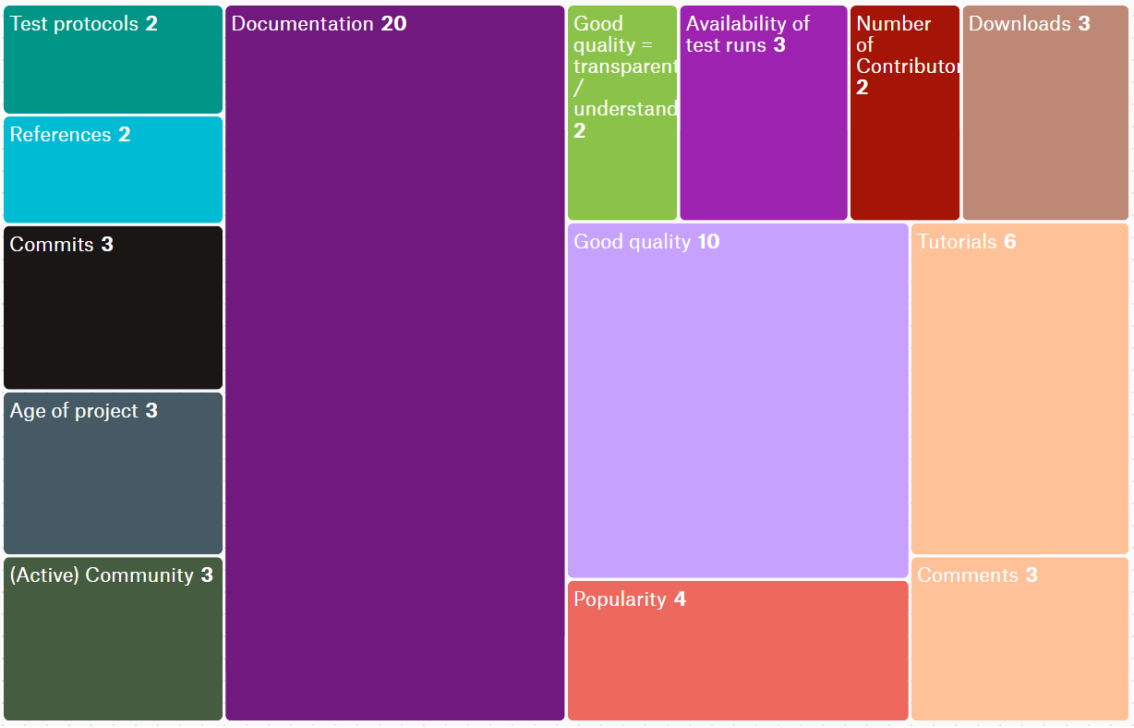

Figure 2. This diagram shows which codes were named together with the "Documentation" code. For the sake of a clear illustration, 27 further codes that were co-occuring only once have been removed.

Limitations of the degree of transparency that is achievable via explainable AI (XAI) approaches were revealed in some of the interviews. Given the complexity of any reasonably sophisticated model, fully understanding, comprehending the inner workings and predicting the behaviour of any such system is an illusion. Being able to inspect such systems, however, may contribute to a higher level of trust.

Yet, trust is more likely to be attributed based on other forms of transparency, likely focusing more on the process that led to the training and release of any such model, the qualification and authority of the team that built it. This is highly similar to the forms of trust being awarded in the "analogue" world, where experts such as medical doctors, car mechanics, need to be trusted based on their qualifications, the process leading to them being allowed to perform certain activities, rather than the patient / customer fully understanding the decision process that led to a certain diagnosis.

Conclusions and Future Work

To better understand how we can design solid, trustworthy data infrastructures, what information to collect to allow humans or machines to decide which data and tools to use for high-quality research, what level of transparency to provide, a joint effort may be required to obtain a comprehensive picture across domains, seniority levels and other factors. This will be essential to design solid systems that provide the type of transparency needed to support trustworthy science.

Based on the preliminary review of the data, a few (technical) recommendations can already be derived with respect to metadata and metrics. For one thing, it would be good to consider (information on) tests performed as well as test cases part of the documentation process. It is also crucial to know the purposes of the original data collection / code. In other words, re-users might want to know whether they are working with a student assignment or reviewed code.

Last, it might be of interest to give information on what has already been done with the data / code (including results, or things that went wrong) and who worked with it.
Keeping interim results in mind, two aspects are striking and should therefore be put to test. First, two thirds of the survey participants are not testing whatever it is they decide to reuse. Yet, many claim that penalties are unacceptable. Programmers cannot be responsible for any bugs because it is simply not possible to guarantee bug-free code. The responsibility of testing, they say, lies with anyone who reuses code. Obviously, that does not go together. Therefore, we need to evaluate if the people stating the first are the same group who are propagating the latter. In addition, we need to complete the analysis of all data collected to see if a majority of survey participants thinks so, or if it currently only appears to be so.

Second, we should take a closer look at documentation being the number one quality indicator according to the quantitative analysis and the second-best indicator according to the qualitative one. For the time being, it seems that survey participants associate many different things with documentation. To derive further recommendations, however, we need to learn more about common features.

The questionnaire should also be revised for further steps. Restructuring and focusing on certain topics — now recognized as important — allows for a shorter questionnaire. It should also now be possible to ask mostly closed questions and to supplement the survey with some (open) expert interviews or discussions with focus groups to discuss results.

In follow-up studies, we aim to explore in more detail the actual elements that, if transparently shared, contribute to the perception of trust and correlate these with actual quality indicators in research and development. We further plan to investigate in more detail differences between communities of practice and seniority levels.

Specifically, we are looking forward to receiving feedback on the questionnaires[6] as well as collaborating to extend the study into specific disciplines and infrastructure settings. A first step in this direction was initiated last October during a workshop organized jointly by the European Research Consortium on Informatics and Mathematics (ERCIM) and the Japan Science and Technology Agency (JST) (Rauber et al., 2024). This should allow us to jointly derive recommendations on how to improve the design and operations of research infrastructures on a technical, (governance) process and communication level for them to be both trustworthy and trusted.

Acknowledgements

Part of this work has been funded by ASEA UNINET (project ID TUWIEN-2022/2023), EOSC Secretariat (Horizon 2020, grant agreement number 831644), and EOSC Focus (Horizon Europe, grant agreement number 101058432). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

Campiglio, E., Carver, M., Esposito, E., Flicker, K., Frank, J., Hodges, R., Holm, P., Karlson, K. B., Krug, I., Muir, E., Pulignano, V., Quirico, O., Rauber, A., Saurabh, S., Saurugger, B., Schiffels, S., Schuck, N., Susi, T., Wagner, W., & Wolkers, M. (2020) *Report on "Visions, requirements and needs for Future Research Environments: An Exploration Series with Researchers"*. Zenodo. https://doi.org/10.5281/zenodo.4336705

[6] https://zenodo.org/records/10626345

Choudhury, A., & Shamszare, H. (2023). Investigating the impact of user trust on the adoption and use of ChatGPT: survey analysis. *Journal of Medical Internet Research*, 25:e4718. https://doi.org/10.2196/47184

Ertz, M. (2022). New scam: Do you want to be paid to cite others' research? *ResearchGate.* https://www.researchgate.net/post/New_scam_Do_you_want_to_be_paid_to_cite_others_research

Hawley, K. (2014). Trust, Distrust, Commitment. *Noûs*, 48(1): 1-20. https://doi.org/10.1111/nous.12000

Jacovi, A. Marasović, A. Miller, T., & Goldberg, Y. (2021). Formalizing trust in Artificial Intelligence: prerequisites, causes and goals of human trust in AI. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 624–635. https://doi.org/10.1145/3442188.3445923

Kogut, B., & Metiu, A. (2002). Open-source software development and distributed innovation. *Oxford Review of Economic Policy*, 17(2): 248–264. https://www.jstor.org/stable/23606809

Kopetz, H. (2022). An Architecture for Safe Driving Automation. *Principles of Systems Design*, Springer. https://doi.org/10.1007/978-3-031-22337-2_4

Mayring, P. (2013). Qualitative Inhaltsanalyse. *Qualitative Forschung. Ein Handbuch:*. Rowohlt Taschenbuch Verlag: 468-475. ISBN 978-3-499-55628-9

McLeod, C. (2021). Trust. *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/trust/

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on management information systems (TMIS), 2*(2), 1-25. https://doi.org/10.1145/1985347.1985353

Nastjuk I., Herrenkind, B. Marrone, M., Brendel A.B., & Kolbe, L.M. (2020). What drives the acceptance of autonomous driving? An investigation of acceptance factors from an end-user's perspective. *Technological Forecasting & Social Change,* 161. https://doi.org/10.1016/j.techfore.2020.120319

O'Mahony, S. (2003). Guarding the commons: How community managed software projects protect their work. *Research Policy, 32(7):* 1179–1198. https://doi.org/10.1016/S0048-7333(03)00048-9

Pink, S., Lanzeni, D., & Horst, H. (2018). Data anxieties: Finding trust in everyday digital mess. *Big Data & Society, 5*(1). https://doi.org/10.1177/2053951718756685

Rauber, A. Oyama, S., Kashima, H., Yanai, N., Li, J., Takeuchi, K., Aizawa, A., Plexousakis, D., & Flicker, K. (2024). Theme 3: Trust in Data-Driven Research. *Report on the 4th Joint JST/ERCIM Workshop*: 9-10. *ERCIM News, 136.* https://ercim-news.ercim.eu/images/stories/EN136/EN136-web.pdf

Shah, S. K. (2006). Motivation, Governance, and the Viability of Hybrid Forms in Open Source Software Development. *Management Science, 52(7):* 1000–1014. https://doi.org/10.1287/mnsc.1060.0553

Tronnier, F., Harborth, D., & Hamm, P. (2022). Investigating privacy concerns and trust in the digital Euro in Germany. *Electronic Commerce Research and Applications*, *53*, 101158. https://doi.org/10.1016/j.elerap.2022.101158

van Bevern, R., Komusiewicz, C., Molter, H., Niedermeier, R., Sorge, M., Walsh, T. (2020). *h-Index manipulation by undoing merges. Quantitative Science Studies, 1(4):* 1529–1552. https://doi.org/10.1162/qss_a_00093

von Krogh, G., & von Hippel, E. (2006). The Promise of Research on Open Source Software. *Management Science, 52(7):* 975–983. https://doi.org/10.1287/mnsc.1060.0560