Trust in Research Practices & Infrastructures

Katharina Flicker[1], Stefan Reichmann[2], Susanne Blumesberger[3], Marie Czuray[1], Miguel Rey Mazon[2], Bernd Saurugger[1], Andreas Rauber[1]

[1] TU Wien, Austria

[2] Graz University of Technology, Austria

[3] University of Vienna, Austria

DOI 10.3217/978-3-99161-062-5-003, CC BY 4.0

Abstract. The increasing complexity of digital research workflows raises questions about trust in research processes, results, and infrastructures. This study builds on philosophical concepts of trust to examine their relevance to research practices, particularly in relation to data, tools, services, and open-source software. We explore how trust influences sharing and reuse, the perception of quality indicators, and the development of trustworthy infrastructures. Two exploratory approaches were employed: a survey among data scientists in open-source software, and twelve semi-structured interviews with researchers from various disciplines focusing on trust in data quality. Interviews were transcribed and analysed using inductive coding supported by ATLAS.ti. Findings reveal a consistent gap between research ideals and practice. While researchers recognize the importance of verifying the fitness for purpose for reused resources, time constraints often lead them to rely on proxies such as documentation and source reputation. Trust is closely tied to institutional affiliation, peer review, and ethical standards, indicating that reputation and adherence to ethical codes influence perceptions of trustworthiness. The results also highlight the need for mechanisms to assess and communicate trustworthiness especially in dynamic and interdisciplinary contexts. Questions arise about integrating such mechanisms into research infrastructures, including standards for documentation, compliance monitoring, and responses to violations. This work lays the foundation for future research on institutional and technical processes that can foster trust and trustworthiness in the development and use of digital research infrastructures.

1 Introduction

Research is increasingly based on digital data, collected from a variety of sources, pre-processed and analysed by many stakeholders. The value of such and the need to protect the massive investments in terms of time, money and computing resources having gone into this, led to the building of complex data and compute infrastructures such as encountered in the proliferation of AI research and deployments. These are expected to ensure that research outputs are securely stored, properly documented, available for (re-)use, and compliant with the FAIR Principles[1] (Wilkinson *et al.*, 2016), which purport to streamline research data management practices to make digital research outputs (Wilkinson *et al.*, 2018) reusable by machines and humans (Boeckhout, Zielhuis, and Bredenoord, 2018).

Several scholars have proposed to study the complexity of data processing pipelines in terms of 'data journeys' (Leonelli, 2016; 2020) or 'data distance' (Borgman and Groth, 2025), to conceptually accommodate the amount of invisible labour and actors involved in rendering digital research outputs reusable (Leonelli, 2016). While this discussion is empirically saturated (cf. Leonelli and Tempini, 2020), the quality requirements for digital research outputs for a particular type of research, raise important (yet unanswered) questions concerning the trustworthiness of the underlying processes. More recently, the difficulties in assessing the quality of digital research outputs – ranging from specific findings to complex AI models – have spawned concerted efforts to tease apart the details of (input and result) data quality[2]. It is worth noting that FAIR constitutes a paradigm for guaranteeing the machine-readability of digital research outputs, while saying nothing about quality.

This article discusses the problems of ensuring quality of research outputs in terms of the trustworthiness of underlying processes, starting by classic and recent work on trust and trustworthiness in digital environments in sections 2.1, 2.2 and 2.3. It addresses recent academic crises (e.g., reproducibility) and their effects on trust in research processes, followed by an assessment of the infrastructural turn in STS in sections 2.3 and 2.4. While the research design is described in section 3, section 4 calls attention to empirical qualitative and quantitative data illustrating how researchers evaluate the trustworthiness of digital research outputs. Concluding remarks in section 5 highlight

[1] The FAIR principles mandate that research data be Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2018), to ensure that digital research objects can be discovered and reused.

[2] Both RDA and the EOSC Association have invoked Task Forces (TF) and Working (WG) as well as Interest Groups (IG) to tackle these and related issues. The EOSC Association, for example, established the FAIR Metrics and Data Quality Task Force (2021-2023), and the FAIR Metrics and Digital Objects TF (2024-2025) (https://eosc.eu/eosc-association/eosc-task-forces/), while RDA runs the FAIR Mapping WG (https://www.rd-alliance.org/groups/fair-mappings-wg/activity/), or the FAIR Instrument Data IG (https://www.rd-alliance.org/groups/fair-instrument-data-ig/activity/) among many others.

implications for designing infrastructures and guidelines to enhance trust via mechanisms, processes, and human factors.

2 Related Research: Trust, credibility, and (research) infrastructures

This section begins by examining the philosophical foundations of trust and develops a working definition of the concept (2.1). It then illustrates how trust has been conceptualized and studied within the field of information and communication technology (2.2) and explores its role in research processes, outcomes, and infrastructures (2.3). Finally, drawing on insights from Science and Technology Studies, the section defines infrastructures and highlights the critical importance of addressing trust-related issues within the context of research infrastructures (2.4).

2.1 Trust

Trust is elusive. While the concept has been discussed in philosophy and related fields, there are many competing, often contradicting, definitions (McLeod, 2021). In social situations, there tends to be what Walker (2006) dubbed 'default trust' – based on a tacit understanding of what can be expected from one's surroundings. Barber (1983) provides a systematization of different aspects of trust, distinguishing 1) trust in the continuity of natural and social orders, 2) trust in the technical competence of an actor, and 3) trust in an actor's propensity to consider the interests of others (morality). Trust, then, is important across social settings. Social theorists recognize trust as a prerequisite of specifically modern (as opposed to traditional) forms of social interaction, by linking it to the contingency of social interaction: Hardin (2006; 2001), Gambetta (1988), and Luhmann (1979) observe that trust involves a moment of uncertainty about the behaviour of others. Contingency implies that actions can have unintended consequences (Merton, 1936); trusting others solves problems of contingency at the interpersonal, group, and societal levels (Luhmann, 1979). Similarly, Georg Simmel observed that trust is located, epistemically, between complete knowledge and complete ignorance of the other, and is therefore a background assumption of everyday life. In this, trust is essential for social interaction (Lewis and Weigert, 1985), and seems to be irreducibly social (Lewis and Weigert, 2012). Giddens (1996) later observed that trust serves to reduce the complexity of social interactions in that it reduces the 'costs' associated with having to verify others' intentions (Luhmann, 1979). Modern societies inherently rely on trust owing to increasing rationalization, which necessitates reliance on the expertise of others (Collins, 2007). For Luhmann (1979), trust sustains (relatively) stable expectations regarding the natural and social orders.

In philosophy, trust is the subject of theoretical and practical philosophy, where the concern is with finding an evidence-based and rational definition. These discussions are too complex to track here, so the following working definition is proposed: Trust is warranted when it is plausible, well-grounded, and justified. Trust is plausible only if the conditions for trust are obtained, and when one is able to develop trust. Trust is well-grounded only if the trustee is trustworthy. It can be justified even if the trustee is not trustworthy as long as some value can be expected to emerge from giving trust (McLeod, 2021).

Defining trustworthiness is even more elusive, although there is a tendency to define the concept relative to trust. Hawley (2014) offers a helpful working definition based on three assumptions: 1) Trustworthiness is a trait or attribute of someone/something, 2), there is a difference between general and specific trustworthiness, and 3) to be trustworthy means to meet reasonable and appropriate expectations.

2.2 Trust in Information and Communication Technology (ICT)

There is a well-developed body of literature on trust in organisations, data, and technologies. With the growing amount of data, questions about how trust enhances its value arise in scientific and mundane contexts (Pink, Lanzeni and Horst, 2018). The relationship between trust in technology and its subsequent use is gaining significant attention. For example, Tronnier, Harborth and Hamm (2022) explore how privacy concerns and trust in currency affect individuals' willingness to adopt Central Bank Digital Currency; (Jacovi *et al*., 2021) study trust as a crucial component in human-AI interactions and in ICT more broadly (McKnight *et al.*, 2011).

With the wide-spread use of ICT, the notion of e-trust has been introduced to describe forms of 'trust specifically developed in digital contexts and/or involving artificial agents' (Taddeo and Floridi, 2011). It typically pertains to trust in online environments but has also been used to describe more general issues of digitally mediated social interactions. E-trust is typically analysed along the following three dimensions: 1) Trust in technologies, 2) trust in other users, and 3) trust in technology providers (Taddeo and Floridi, 2011), to which Spiekermann (2016) adds trust in the engineers who built a particular technology. Accounts of e-trust tend to have a cognitive leaning, frequently modelling e-trust as a relationship between a trustor and a trustee, because trust in artefacts cannot be based on either morality (Nickel, Franssen, and Kroes, 2010) or motivations (Hardin, 2006). As Sztompka (1999) points out, there might be a continuum from trust based on rationality to trust based on morality (i.e. rationality begets morality): Betting on virtue is riskier than betting on rationality. Taddeo (2010) suggests explicating trustworthiness not as general reliability, but rather as reliability with respect to performing a specific task. Trust is only achieved when coming from the right reasons: care for the other's interest combined with moral integrity (Nickel, Franssen, and Kroes, 2010). These may be difficult to assess in digital environments. For this reason, Spiekermann (2016) proposes to speak of reliance on technology rather than trust, and correspondingly, reliability instead of trustworthiness.

2.3. Trust in, and credibility of, the research process

The following explores trust in research processes, results, and infrastructures, while not asking why the general public can and should trust science (Oreskes, 2019), but rather, which mechanisms (if any) warrant researchers to trust their own methods, tools, findings, etc. In this context, trust does not need to be absolute but must be justified and supported by evidence.

This understanding of trust is the kind that should be earned by research practices, infrastructures, and outcomes. An important precursor to this discussion is Merton's (Merton, 1973) work on 'The normative structure of science', which claims that in order for science to convey certified knowledge, researchers adhere to a set of four ('Mertonian') norms (Hosseini et al., 2024): universalism, communalism, disinterestedness, and organized scepticism, each with more or less bearing on the emergence of trust. In Merton's view, science is 'disinterested' in the sense that there are no 'external' customers, i.e., the consumers of scientific communication are other scientists. Organised scepticism entails that researchers may trust the system of peer review as a whole without having to check, for each research output, whether it is trustworthy – provided that they trust the system and understand that it has been peer reviewed.

This picture has come under heavy scrutiny (and criticism), under the impression of recent crises of academia, such as the crisis of peer review (Horbach and Halffman, 2018; Daniel, Mittag, and Bornmann, 2007; Smith, 2006), and – connected to the first – the reproducibility crisis (Ioannidis, 2005; Begley and Ioannidis, 2015; Leonelli, 2018; França and Monserrat, 2019), both of which tend to negatively impact the (perceived) trustworthiness of science.

Oreskes (2019) finds that trust in science is based upon consensus. Since there is no single valid scientific method, trust in its outcomes is based on the social character of the scientific process. The trustworthiness of science stems precisely from the social processes by which published results are discussed, reviewed, contested, and ultimately accepted in a 'cycle of credibility' (Latour and Woolgar, 1986).

It should be noted, that the process of allocating credibility in academia is subject to the 'Matthew effect' (Merton, 1968), a dynamic of cumulative advantage (Ross-Hellauer *et al.,* 2022) whereby early success (measured in terms of scholarly impact, i.e. citations) begets later success. Already successful researchers tend to receive rewards and recognition disproportionately, which tends to translate into resources and access to infrastructure, resulting in an extremely stratified distribution of these resources. For Merton, the Matthew effect fulfils an important function at the system level, in that it serves to assess the credibility of sources (similar to trust more generally). At the individual level, the effects of cumulative advantage – 'a general mechanism for inequality across any temporal process (e.g., life course, family generations) in which a favourable relative position becomes a resource that produces further relative gains' (DiPrete and Eirich, 2006) – tend to be detrimental, as 'various aspects of academia are particularly vulnerable to logics of cumulative advantage' (Ross-Hellauer *et al.*, 2022).

2.4. Infrastructuring in Science and Technology Studies (STS) and the Challenge of Trust and Credibility in Research Infrastructures (RIs)

Infrastructure has garnered renewed interest from STS scholars since the 2010s. One of the guiding questions of this revival has been how to address diverse infrastructures (e.g. transportation, information infrastructures, electricity, etc.) under a single rubric despite their heterogeneity. This refocusing has been called 'infrastructure inversion' (Star and Ruhleder, 1994) to emphasize that infrastructure is not a neutral background, but has political consequences: modern nation states could not have coalesced without the expansion of print media (Reicher, 2013). Slota and Bowker (2017, p. 531) remark that 'one of the key insights of STS has been to treat infrastructure relationally: it is not so much a single thing as a bundle of heterogeneous things (standards, technological objects, administrative procedures – in Foucault's term, a *dispositif technique* (Foucault, 1979) – which involves both organizational work as well as technology', and further:

> 'The centrality of the material in anthropological infrastructure studies engenders a discussion of 'embodied experience governed by the ways infrastructures produce the ambient conditions of everyday life: our sense of temperature, speed, fl[u]orescence, and the ideas we have associated with these conditions' (Larkin 2013, 335). Interactions with infrastructure govern not just the aesthetic experience of the world, they define imaginaries of what is possible and potentially possible and are presented politically as a pathway to those potentials.' (Slota and Bowker, 2017, p. 535)

STS scholars have increasingly been using the retronym 'built infrastructure' to describe what infrastructures used to be before the advent of information infrastructures. Similiarly, David (1990) found a cultural lag between innovation and usability (Slota and Bowker 2017, p. 536), before the invention of the appropriate 'infrastructural imaginary'. For Research Infrastructures, this would be the function of the Open Science discourse, respectively the FAIR Principles (Wilkinson *et al.*, 2016). Further, many have pointed to the network effects of information/communication infrastructures (where additional users do not increase individual costs associated with using a network, but entail additional benefits).

With 'The Ethnography of Infrastructure', Susan Leigh Star (1999) presents a methodological treatise on how to study infrastructure. According to Star, infrastructure is both ecological and relational, part of actions, tools, and the built environment. The ecology of the distributed high-tech workplace is impacted by infrastructure that permeates all its functions. In order to fully appreciate how an infrastructure works, one needs to examine those who are *not* served by it, and to examine technological systems in the making – what counts as infrastructure is a matter of perspective. Infrastructure is fundamentally relational (Star and Ruhleder 1996, p. 113) and encodes work (Star 1999, p. 385), from which follows the methodological imperative to find the invisible work needed to sustain a given infrastructure (cf. Leonelli, 2016).

Infrastructure tends to mean different things to different groups. Star and Ruhleder were arguably the first to formalize many of the preceding concepts of infrastructure, influencing future methodological work in infrastructure studies. Their 'Steps towards an Ecology of Infrastructure' (1996) was an ethnographic study of a collaboratory of biologists and computer scientists working with the Worm Community System (WCS), a digitized library of C. elegans flatworm specimens and technologically mediated paths for collaboration among the biologists working with them, which was in many ways an ideal site for the implementation of new computing infrastructure (social expectation of collaboration and a well-established network of biologists sharing specimens). WCS was a failed project with little uptake among the studied communities. In their accounting for that failure, Star and Ruhleder propose infrastructural issues as major factors influencing that outcome (Slota and Bowker, 2017, p. 537): 'As Star and Ruhleder (1996) have argued so eloquently, one person's invisible infrastructure is another person's job, to be faced materially and directly every day. Infrastructure, as they argue, is inherently relational – a given system, technology, or organization is infrastructural to a particular activity at a particular time' (Slota and Bowker 2017, p. 531).

The upshot of this discussion is as follows. Infrastructure has multiple dimensions: it is both embedded and transparent; it exists (metaphorically) underneath other social, technical (built) worlds and does not need to be reconsidered every time it is 'used' to enable a task. It tends to become visible only upon breakdown; it embodies standards and practices that are learned as part of enculturation processes into a given user community; it is rarely built de novo and tends to be fixed in modular increments (Star, 1999, p. 381 ff.).

Infrastructure is inherently technical, social, organisational, political, as well as (in the case of research infrastructures (RIs)), epistemic (Edwards, 2010). STS applies sociotechnical concepts to RIs (Slota and Bowker, 2017, p. 537): 'For Hughes and Latour, infrastructure was not just technology: it was always already braided with social, cultural, and political actors and their values' (Slota and Bowker, 2017, p. 532). Actor-Network Theory was developed largely in an ethnomethodological tradition, but with the important addition of stressing the interchangeability of human and non-human actors (Latour, 1993): Technology is politics by other means (Latour, 1987). There has been a return to physical infrastructures in STS after 2010, with increased interest in the material aspects of knowledge production (Edwards, 2010), even before the advent of the Open Science movement and its renewed interest in knowledge infrastructures and research practices (across STS, CSCW, Information Science, etc.). Another important strand of literature discusses information infrastructures and changes to the profession of the librarian over the last four decades. The advent of Open Science, with the establishment of data professionals (data stewards) changes the profession yet again. Since then, STS developed an interest in the epistemic affordances of infrastructures. Edwards (2010) details how climate is established as a global phenomenon through the construction of a global climate observational infrastructure. With the dominance of data, STS focuses increasingly on the dominance of computer and data science over 'domain science' (Ribes, 2017; Ribes *et al.*, 2019). The situation resembles the displacement of earlier infrastructures (e.g. Xerox machines) in a process of colonisation which tends to remodel the colonised practices in its own image. EOSC as a federation of RIs seems particularly amenable to STS discussions of infrastructure via ANT's contention that agency is distributed between humans and non-humans (Latour, 1987). Latour and Woolgar (1986) had already drawn attention to the material substrates of intellectual life ('immutable mobiles'), as had Knorr Cetina (1981). Edwards *et al.* (2013) postulate a continuity from the study of physical infrastructure to the study of cyberinfrastructures (e.g., in terms of path dependence). The need for standardization (Busch, 2011) tends to create problems of its own, e.g. in defining a 'standard human', or in defining a 'standard research process'). In this, STS work has also had its 'infrastructural inversion', in beginning to describe the history of large-scale systems as part of human organization (Edwards, 2010).

3 Research Design & Methodology

The research described below was conceived against the backdrop of the EOSC Focus project[3] and activities of the EOSC Support Office Austria[4] Working Group on Researcher Engagement in Austria. Both initiatives are geared towards shaping the European Open Science Cloud (EOSC). A survey as described in section 3.1 was launched as part of EOSC Focus, while the WG conducted semi-structured interviews as outlined in section 3.2. The survey was used to enable quantitative analysis, offering the typical advantages associated with standardized surveys, such as broad reach, anonymization, efficiency in data collection and analysis, and relatively low demands on personnel resources. In parallel, semi-structured interviews allowed participants to raise relevant issues that may not have been anticipated in the survey. The design and implementation of both the survey and interviews are described below concluding with limitations of the research design in section 3.3.

[3] https://eosc.eu/eosc-focus-project/

[4] https://eosc-austria.at/

3.1 Survey Design and Methodology

The survey was sent to an entire cohort of students of Data Science (N=120, response rate=75%). 44,33% of participants were also working professionally in the field, with an average of 2.86 years of experience. It was designed to understand factors affecting trust as well as quality indicators used by respondents with respect to sharing and reusing open-source software, data, tools and services. The survey consisted of 35 open and closed questions, organised into six sections. Section 1 (4 questions) collected information on the participants' practical experience. Section 2 (12 questions) addressed respondents' code-sharing practices. Section 3 (5 questions) focused on code reuse, and section 4 (7 questions) explored aspects of quality. Trust was addressed in section 5 (5 questions). The survey concluded with a question on accountability, and another asking for additional comments. The anonymised data set[5] and the survey[6] are available on Zenodo, while the interim results were presented and discussed at the 18th edition of the International Digital Curation Conference 2024 (IDCC24) and published (Flicker *et al.*, 2024).

The survey was circulated via TUWEL, TU Wien's e-learning platform. The standardized questions were thus analysed automatically; the open-ended questions were analysed using inductive categorisation to derive categories ('codes') from the text responses. These 'codes' are intended to represent the material and thus allow statements and interpretations without distorting the core content of the material (Mayring, 2013). The qualitative analysis was supported by the software ATLAS.ti (Kelle, 2013).

3.2 Semi-structured interviews

Additionally, twelve semi-structured interviews were conducted by the EOSC Support Office Austria's (EOSC SOA) Working Group Researcher Engagement in Austria (WG REA). The interview guide was developed iteratively by all members of the WG REA, and structured in three parts: *Part 1* collected interviewees' demographic information (disciplines, institutional and departmental affiliation, seniority level, position and career stage, and gender). Two introductory questions related to the interview partners' research and their motivations aiming at more information to contextualize all information given on research and data practices, trust and data quality. *Part 2* focused on actual data practices. *Part 3* dealt with trust in data and data quality. The interview guideline is available on Zenodo[7].

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

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

[7] https://zenodo.org/records/15295668

Interview participants were recruited from Austrian public universities, to ensure the collection of nationally relevant perspectives and requirements, and to accommodate for the limited resources available to the WG members[8].

The participants were recruited through personal contacts, additional researchers were approached via email using a purposive, though not fully systematic, strategy, with a particular focus on department heads and recipients of ERC grants, individuals who are typically well-established within their disciplines and thus possess substantial familiarity with research infrastructures, environments, and practices. Furthermore, it could be expected that their professional standing and credibility within the research community would increase the likelihood that their views were influential at both peer and institutional levels. These individuals were also well positioned to disseminate the interview request within their teams. Respondents were recruited from fields as diverse as Computer Sciences, Life Sciences, Social Sciences and Humanities, covering ten Austrian public universities (out of 23). The interviews were conducted, recorded and transcribed by five members of the WG REA. The transcripts were later edited for legibility.

Para-linguistic features of spoken language (laughter, sighs or breathing) or extra-linguistic features (eye movements, gestures) were not transcribed, while word repetitions and interruptions or slips of the tongue were written down (Kowal and O'Connell, 2013).

Interviews were edited for publication on Zenodo[9], in consultation with the interviewees. It is crucial to note, that the transcribed interviews – not the interviews edited for publication – were used for the analysis. Interview analysis followed the inductive categorization approach (Mayring, 2013) and was supported by ATLAS.ti (Kelle, 2013). After a first round of analysis, applied codes as well as interim results were discussed in small groups.

3.3 Limitations of the research design

Four factors limit the interpretability of the results. First, the original purpose of the WG REA was not to improve the understanding of research practices, nor to develop indicators that facilitate trust in data and data quality, services and tools, but to collect requirements for research environments and infrastructures from Austrian researchers, and feed them into strategic (inter-) national policy papers to inform the development of EOSC. Consequently, the study prioritized research institutions located in Austria, while

[8] This approach was also chosen for pragmatic reasons: public universities are officially recognized research institutions and are listed by oesterreich.gv.at, a platform across public authorities. The Austrian Federal Chancellery is responsible for the content of this platform.

[9] https://zenodo.org/communities/wgreaeoscsoa/records?q=&l=list&p=1&s=10&sort=newest

simultaneously striving for an inclusive survey approach that encompassed a broad range of scientific disciplines.

Second, the WG REA members come from diverse professional backgrounds, including the library sector, social sciences, data science, physics, and business informatics. As a result, their familiarity with interviewing as a method of data collection and transcription and as a theoretically informed - but not neutral - practice varied considerably. While many issues relating to transcription were addressed through group discussions (e.g., regarding what should be transcribed), the diversity in experience contributed to inconsistent interviewing styles and occasional errors during the interviews (Hopf, 2023). Despite all interviewers being well-acquainted with the project, challenges remained in determining when to steer conversations back to the central topic or when to probe further with follow-up questions. These difficulties were compounded by disciplinary differences: interviewers often came from fields different from those of the interviewees, which sometimes hindered their ability to recognise opportunities for content-relevant follow-up questions. In several cases, such gaps were only identified retrospectively during transcript review and subsequent research.

Third, due to limited time and financial resources, only a single round of analysis was conducted for the open-ended survey responses and interviews. Coding and results were discussed, reviewed, and revised within small groups, but not by the entire research team. As such, the findings should be interpreted with appropriate critical consideration.

Last, regarding representativeness of the survey, it is important to note the implications of the sampling strategy and data collection procedure as described in '3.1 Survey Design and Methodology'. The findings are generalizable to the entire cohort of Data Science students at TU Wien but do not necessarily constitute a representative sample of data scientists more broadly. A more fundamental limitation of these responses concerns the fact that these likely represent ideal (as opposed to actual) data practices, in the same way that research papers model ideal, not actual, laboratory practices (Knorr Cetina, 1981).

4 Results and Discussion

This section presents and discusses the results jointly. Instead of separating findings by how the data was collected (the survey and the semi-structured interview series) the chapter is organized into three thematic sections reflecting the key findings: Perceiving Trustworthiness (4.1), Indicators of Quality (4.2), and Documentation (4.3). This structure was chosen because insights from both the survey and the interviews contributed to each of these themes.

4.1 Perceiving Trustworthiness

In the context of scientific knowledge production, trust needs to be rational as well as evidence based. In other words, there needs to be a reason grounded in evidence that someone or something is trustworthy. Reasoning must be revised in case new evidence shows up (McLeod, 2021). It is thus linked to quality management, or rather to methods to check quality that differ depending on the actual discipline and related research questions.

Additionally, it is – and that across disciplines – also linked to context, or more specifically to organisations, institutions, and (peer-reviewed) journals, although the latter has been criticised for being flawed (lack of time, highly subjective perspectives). In other words, reputation matters. This leads to further questions about how organisations, institutions and journals can both build and maintain a reputation for being trustworthy, or about how reputation can be communicated in dynamic environments or in case potential recipients come from outside the core community or different geographic regions. While there is currently no answer to the second question, there are certainly hints on how reputation can be maintained in a scientific context. As one participant in the semi-structured interview series noted: 'In my opinion, trust in this context is only possible if I can assume that my colleagues have the same medical ethics or scientific ethics.' (Schmutzhard and Flicker, 2023)

4.2 Indicators of Quality

Although previous studies have partially explored the discrepancies between ideal research practices (Flicker *et al.,* 2024) or the motivations for data sharing and reuse (Reichmann *et al.*, 2021) and actual research conduct, the combination of the survey and the interview series offers novel insights: In the context of sharing data, tools, services and open-source software, the survey suggested a strong discrepancy between the ideal of quality checks and their implementation. While many stated that quality checks should be conducted before re-use to determine fitness for purpose, they failed to live up to that (mostly due to lack of time). Rather, they decided to trust whatever they intended to reuse based on specific indicators.

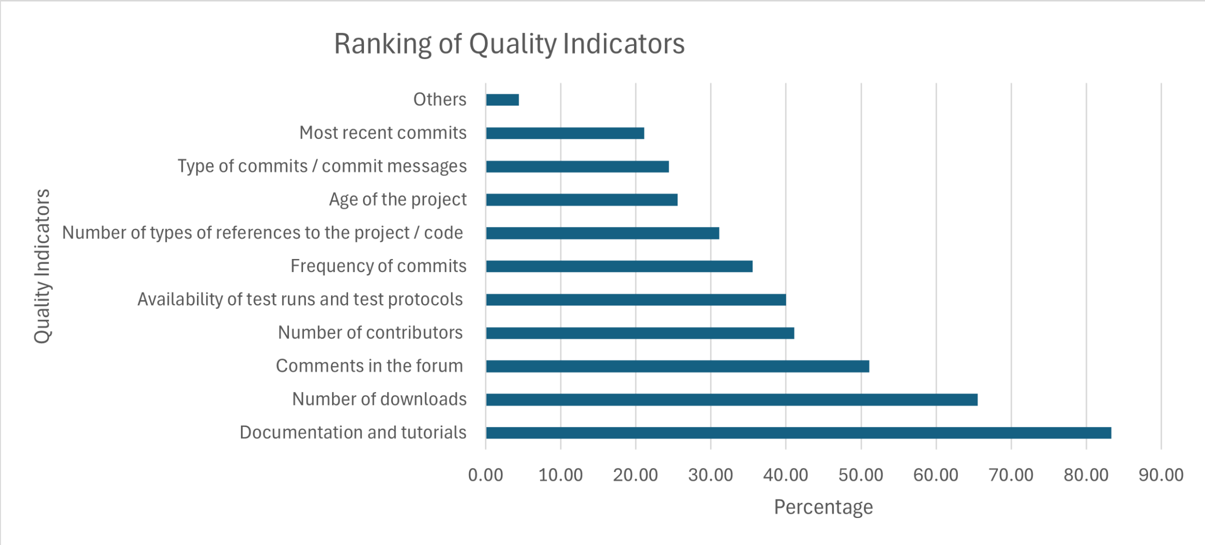

Figure 1: Quality Indicators as ranked by students and professionals of Data Science.

Fig. 1 shows the quality indicators believed to be most crucial based on the survey's quantitative analysis. The top ranked quality indicator was 'Documentation', followed by the 'Number of downloads' and 'Comments in the Forum'. While 83.33% considered 'Documentation and tutorials' to be the most important, the 'Number of downloads' was almost 20% lower at 65.56%. Other quality indicators were 'Number of contributors' (41.11%), 'Availability of test runs and test protocols' (40%), 'Frequency of commits' (35.56%), 'Number of types of references to the project / code' (31.11%) and 'Age of the project' (25.56%), 'Type of commits / commit messages' (24.44%) and 'Most recent commits' (21.11%). The final category was 'others' (4.44%).

The qualitative analysis of the survey draws a similar, yet somewhat different picture as is shown in the Sankey Diagram in **Fig. 2**. The left side lists the quality indicators: (i) (Active) Community, (ii) Documentation, (iii) Reproducibility, (iv) Comments, (v) Commits, (vi) Downloads, (vii) Popularity, (viii) Test protocols, (ix) Tutorials, (x) Availability of test runs and (xi) References. The wider a bar, the more frequently an indicator was mentioned.

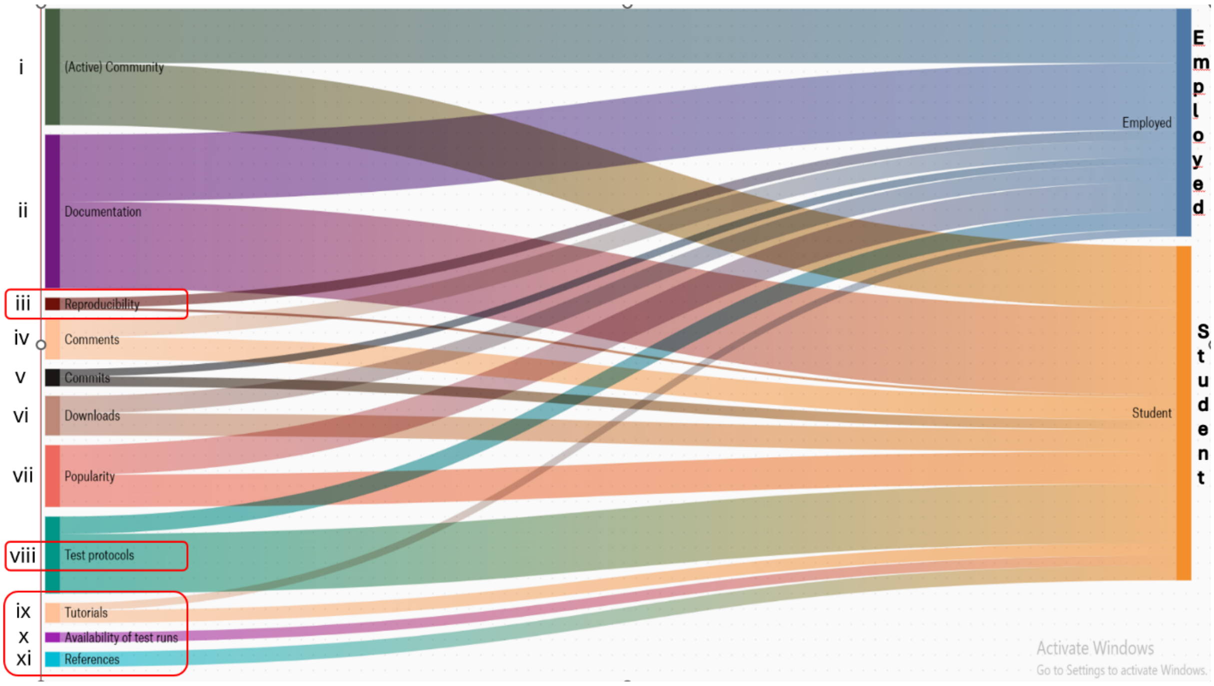

Figure 2: Sankey Diagram showing the frequency with which quality indicators ((Active) Community (i), Documentation (ii), Reproducibility (iii), Comments (iv), Commits (v), Downloads (vi), Popularity (vii), Test protocols (viii), Tutorials (ix), Availability of test runs (x) and References (xi)) are cited by two groups – namely those who are Employed and Students. Indicators (iii, viii, ix, x and xi) with differing ratings between the two groups are highlighted in red.

The right side separates the survey's participants into two groups - Students and those who were also employed in subject-related fields (Employed)[10]. The stream-shaped connections between the indicators on the left and the groups of respondents on the right provide an approximate visualization of which groups identified specific indicators and the frequency with which they did so.

The most frequently cited quality indicators are (i) (Active) Community, (ii) Documentation, (vii) Popularity, and (viii) Test Protocols. The analysis of the stream-like connections reveals that (i) (Active) Community, (ii) Documentation, and (viii) Popularity are considered important by both groups. However, (i) (Active) Community and (vii) Popularity appear to hold slightly greater significance for students. (viii) Test Protocols are predominantly valued by students, whereas (iii) Reproducibility is more relevant to Data Scientists. Two indicators – (x) Availability of Test Runs and (xi) References – were mentioned exclusively by students. In other words, certain quality indicators are exclusively, or at least predominantly, relevant to students ((viii)Test Protocols, (ix) Tutorials, (x) Availability of test runs, (xi) References), whereas others are more applicable to professionals ((iii) Reproducibility).

[10] These results as well as the graphic are based on interim findings after the analysis of around one third of the survey.

4.3 Documentation

Documentation (ii), emerges as the most important quality criterion for both groups, as confirmed by two independent methods of analysis. This finding is further supported by the results of the interview analysis. For example, one Computer Scientist from the semi-structured interview series put it that way:

> 'We collect provenance information, i.e., how the data is being defined, being generated, being collected and so on and so forth, to provide at least some basic information about the Quality and what can be expected out of them. Documenting the original goals of data generation helps to ensure that datasets are not taken out of context and used for things that simply do not align with them at all.' (Ekaputra and Czuray, 2024)

A Communication Scientist who participated in the same interview series emphasised the relevance of documentation when stating that 'Qualitative Research (...) is enormously context sensitive. In order to use them [data] again, it would probably require a great deal of contextual information and metadata, since, for example, it would have to be known who created, processed, analysed and interpreted the data, when, how and why and under what limitations.' (Schreiber and Flicker, 2023), while a linguist interviewed in the same context supported the need to contextualize research data via metadata:

> 'I think it is crucial to contextualise the data, which is very important in cultural studies. That usually happens through the metadata, but in the case of social media that might not be enough.' (Reichl and Blumesberger, 2024)

Closer examination during the interviews, and further analysis of the survey data, revealed that responses to the question of what should be documented, or what constitutes good documentation, were often superficial and/or highly variable. Among other aspects, documentation was associated with elements such as metadata, the provision of information on prior use and its outcomes, illustrative use cases and examples, as well as test protocols and test data. While this may partially reflect limitations in the data collection design, the variability underscores the need for further investigation given the importance of documentation as a quality indicator.

Furthermore, it is important to recognise that quality is not defined by universally applicable criteria; rather, it is context-dependent and should be evaluated in terms of fitness for purpose within specific research settings. The indicators discussed above serve as tools to assess whether a given resource can and should be reused for scientific purposes within research contexts.

Considering the implications for the design and development of research infrastructures is therefore important. For instance and according to findings of the semi-structured interview-series, peer-reviewed journal publications are often regarded as more trustworthy than pre-print versions:

> ‘I am not a fan of data that has appeared in the context of pre-prints – although there seems to be a trend towards this. Many of these pre-prints unfortunately get stuck at exactly this stage and are never published in a recognised journal. This already raises the question of what is wrong with these publications and how reliable are data from these articles that have not been accepted by reviewers.’ (Hofer and Flicker, 2023)

However, the peer review process has also been subject to criticism, particularly regarding the unpaid and time-constrained contributions of researchers, and the perceived subjectivity and lack of transparency in evaluations. These concerns highlight the need to explore practical improvements to the peer review system.

More broadly, it is important to consider how infrastructure-integrated processes can be designed and implemented to promote not only the quality, trust, and trustworthiness of research, but also of associated research artifacts such as data and code. As potential initial steps, researchers highlighted the importance of establishing rules and regulations governing infrastructure use – for example, mandatory standards for data submission to repositories – bringing up questions about how compliance with standards would be monitored and what consequences would follow in cases of non-compliance. While these issues fell outside the scope of the present study, they warrant further investigation.

5 Conclusions and Future Work

The results presented so far form a solid basis for informing and shaping the design of research infrastructures as well as for follow-up studies to elicit concrete guidelines on mechanisms, processes, and human factors increasing both trust and trustworthiness when developing, deploying and operating data services and research infrastructures. Follow-up studies, however, are in need of focus in terms of both disciplines and topics. In addition to the necessity of concentrating on a limited number of disciplines, the survey and interview series have identified several potential avenues for future research. These include quality management and quality checks, the relevance of how reputation facilitates trust and questions about how to not only design but also implement infrastructure-integrated processes promoting quality, trust and trustworthiness regarding research and its artifacts such as data and code. Regarding the latter, the importance of transparent rules and regulations such as mandatory standards for data uploading was emphasized, highlighting key governance-related concerns. Further research in this context would also necessitate an investigation into the mechanisms for monitoring compliance and consequences in cases of non-compliance.

Researchers noted that reputation becomes a critical factor when the processes of checking and verifying data, services, and tools for reuse become exceedingly difficult, time-consuming, or even unfeasible. In such cases, the trustworthiness of the source – such as research organizations, repositories, or journals – gains importance. One factor contributing to perceived trustworthiness is the extent to which these sources adhere to established scientific principles and ethical codes. Further research is needed to identify additional indicators of trustworthiness and to explore how reputation can be effectively communicated in dynamic research environments, particularly when stakeholders originate from outside specific communities of practice or from different geographic regions.

In the context of quality management and quality assurance, it is important to recognize that, given the diversity of research practices and objectives, the concept of ‘fitness for purpose’ – that is, suitability for a specific research question or context – may be more appropriate than the notion of general quality.

Documentation is key for assessing fitness for purpose. However, how documentation is perceived and evaluated varies across social settings warranting further investigation. Additional fitness for purpose criteria should be identified, and research should explore the extent to which their development and implementation can be automated or integrated into existing research infrastructures. Such efforts must be mindful of community acceptance to avoid proposing standards that are impractical or unlikely to be adopted. Furthermore, the robustness of these criteria against intentional manipulation should also be critically examined.

A discrepancy between research ideals and practical realities was also identified – the expectation that data, services, and tools intended for reuse should be evaluated for their fitness for purpose prior to reuse, and the practical challenges (including time constraints) preventing consistent implementation of this requirement. Thus, further questions emerge regarding the identification of indicators that can support the assessment of fitness for purpose and concerning mechanisms enabling researchers to maintain confidence in their work and assume liability, even when full verification is not feasible.

Acknowledgements

This research was partially funded by the EOSC Focus project (Horizon Europe, Grant No. 101058432). The views expressed are those of the authors and do not reflect those of the European Union.

References

Barber, Bernard. (1983) *The Logic and Limits of Trust*. New Brunswick, NJ: Rutgers Univ. Press.

Begley, C. G. and Ioannidis, J. P. A. (2015) 'Reproducibility in Science: Improving the Standard for Basic and Preclinical Research', *Circulation Research*, 116(1). doi: https://doi.org/10.1161/CIRCRESAHA.114.303819

Boeckhout, M. et al. (2018) 'The FAIR Guiding Principles for Data Stewardship: Fair Enough?', European Journal of Human Genetics, 26(7): pp. 931–36. doi: https://doi.org/10.1038/s41431-018-0160-0

Borgman, C. L. and Groth, P. (2025) 'From Data Creator to Data Reuser: Distance Matters', Harvard Data Science Review, 7(2). doi: https://doi.org/10.1162/99608f92.35d32cfc

Busch, L. (2011) *Standards: Recipes for Reality*. Cambridge, Massachusetts: The MIT Press. doi: https://doi.org/10.7551/mitpress/8962.001.0001

Collins, H. (2007) *Rethinking Expertise*. Chicago, Ill.: University of Chicago Press.

Daniel, H. D. and Mittag, S. and Bornmann, L. (2007) 'The Potential and Problems of Peer Evaluation in Higher Education and Research', *Quality Assessment for Higher Education in Europe*, pp. 71–82.

David, P. (1990) 'The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox', *American Economic Review*, 80(2), pp. 355–61.

DiPrete, T. A. and Eirich, G. M. (2006) 'Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments', *Annual Review of Sociology*, 32(1), pp. 271–97. doi: https://doi.org/10.1146/annurev.soc.32.061604.123127

Edwards, P. et al. (2013) *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Available at: https://knowledgeinfrastructures.org/2012-workshop-report/ (Accessed: 3 June 2025)

Edwards, P. N. (2010) *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming*. London, Cambrige, MA: The MIT Press.

Ekaputra, F. J. and Czuray, M. (2024) 'EOSC Support Office Austria: Visions, needs and requirements for research data and practices. An interview with Fajar J. Ekaputra'. Available at: https://zenodo.org/records/14332331

Flicker, K. *et al*. (2024) 'Factors influencing Perceptions of Trust in Data Infrastructures, *International Journal of Digital Curation*, 18(1), pp. 1-12. doi: https://doi.org/10.2218/ijdc.v18i1.921

França, T. F. A. and Monserrat, J. M. (2019) 'Reproducibility Crisis, the Scientific Method, and the Quality of Published Studies: Untangling the Knot', Learned Publishing, 32(4), pp. 406–8. doi: https://doi.org/10.1002/leap.1250

Gambetta, D. (1988) 'Can We Trust Trust?', In Gambetta, D. (Ed.) *Trust: Making and Breaking Cooperative Relations*, New York u.a.: Blackwell, pp. 213–38.

Giddens, A. (1996) *The Consequences of Modernity*. Cambridge: Polity Press.

Hardin, R. (2001) 'Conceptions and Explanations of Trust'. In Cook, Karen S. (Ed.) *Trust in Society*. New York: Russell Sage Foundation, pp. 3–39.

Hardin, R. (2006) *Trust*. Cambridge, UK: Polity.

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

Hofer, T. and Flicker, K. (2023) 'EOSC Support Office Austria: Visions, needs and requirements for research data and practices. An interview with Thomas Hofer'. Available at: https://zenodo.org/records/7855137

Hopf, C. (2013) 'Qualitative Interviews - ein Überblick' in Flick, U., von Kardorff, E. and Steinke, I. (eds.) *Qualitative Forschung. Ein Handbuch*. Leipzig: Rowohlt Taschenbuch Verlag, pp. 349-360.

Horbach, S. P. J. M. and Halffman, W. (2018) 'The Changing Forms and Expectations of Peer Review', *Research Integrity and Peer Review*, 3(1), pp. 8. doi: https://doi.org/10.1186/s41073-018-0051-5

Hosseini, M. *et al.* (2024) 'Messing with Merton: The Intersection between Open Science Practices and Mertonian Values', *Accountability in Research*, 31(5), pp. 428–55. doi: https://doi.org/10.1080/08989621.2022.2141625

Ioannidis, J. P. A. (2005) 'Why Most Published Research Findings Are False', PLOS Medicine, 2(8), e124. doi: https://doi.org/10.1371/journal.pmed.0020124

Jacovi, A. et al. (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*, pp. 624–635. doi: https://doi.org/10.1145/3442188.3445923

Kelle, U. (2013) 'Computergestützte Analyse qualitativer Daten' in Flick, U., von Kardorff, E. and Steinke, I. (eds.) *Qualitative Forschung. Ein Handbuch*. Leipzig: Rowohlt Taschenbuch Verlag, pp. 485-502.

Knorr Cetina, K. (1981) *The Manufacture of Knowledge. An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press.

Kowal, S. and O'Connell, D. C. (2013) 'Zur Transkription von Gesprächen' in Flick, U., von Kardorff, E. and Steinke, I. (eds.) *Qualitative Forschung. Ein Handbuch*. Leipzig: Rowohlt Taschenbuch Verlag, pp. 437-447.

Latour, B. (1987) *Science in Action. How to Follow Scientists and Engineers through Society.* Cambridge, MA: Harvard University Press.

Latour, B. (1993) *The Pasteurization of France.* Cambridge, MA: Harvard University Press.

Latour, B. and Woolgar, S. (1986) *Laboratory Life. The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

Leonelli, S. (2016) *Data-Centric Biology. A Philosophical Study*. Chicago, London: The University of Chicago Press.

Leonelli, S. (2018) 'Re-Thinking Reproducibility as a Criterion for Research Quality', *Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise (Research in the History of Economic Thought and Methodology)*, 40A.

Leonelli, S. (2020) 'Learning from Data Journeys', in Leonelli, S. and Tempini, N. (eds.) *Data Journeys in the Sciences*. Cham: Springer International Publishing, pp. 1-12. doi: https://doi.org/10.1007/978-3-030-37177-7_1

Leonelli, S. and Tempini, N. (2020) Data Journeys in the Sciences. Cham: Springer International Publishing. doi: https://doi.org/10.1007/978-3-030-37177-7

Lewis, J. D. and Weigert, A. J. (1985) 'Trust as a Social Reality', Social Forces, 63(4), pp. 967–85. doi: https://doi.org/10.1093/sf/63.4.967

Lewis, J. D., and Weigert, A. J. (2012) 'The Social Dynamics of Trust: Theoretical and Empirical Research, 1985-20121', Social Forces, 91(1), pp. 25–31. doi: https://doi.org/10.1093/sf/sos116

Luhmann, N. (1979) *Trust and Power. Two Works by Niklas Luhmann. With Introduction by Gianfranco Poggi.* Chichester: Wiley.

Mayring, P. (2013) 'Qualitative Inhaltsanalyse' in Flick, U., von Kardorff, E. and Steinke, I. (ed(s).) *Qualitative Forschung. Ein Handbuch*. Leipzig: Rowohlt Taschenbuch Verlag, pp. 468-475.

McLeod, C. (2021) 'Trust'. Available at: https://plato.stanford.edu/entries/trust/ (Accessed: 27.05.2025)

McKnight, D. H. et al. (2011) 'Trust in a specific technology: An investigation of its components and measures', *ACM Transactions on management information systems (TMIS)*, 2(2), pp. 1-25. doi: https://doi.org/10.1145/1985347.1985353

Merton, R. K. (1936) 'The Unanticipated Consequences of Purposive Social Action', *American Sociological Review*, 1(6), pp. 894-904. doi: https://doi.org/10.2307/2084615

Merton, R. K. (1968) ‘The Matthew Effect in Science: The Reward and Communication Systems of Science Are Considered’, Science, 159(3810), pp. 56–63. doi: https://doi.org/10.1126/science.159.3810.56

Merton, R. K. (1973) ‘The Normative Structure of Science’, in Storer, N. M. (eds.) *The Sociology of Science*. Chicago: The University of Chicago Press, pp. 267–80.

Nickel, P. J., Franssen, M. and Kroes, P. (2010) ‘Can We Make Sense of the Notion of Trustworthy Technology?’, Knowledge, Technology & Policy, 23(3), pp. 429–44. doi: https://doi.org/10.1007/s12130-010-9124-6

Oreskes, N. (2019) *Why Trust Science?* Princeton, NJ: Princeton University Press.

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

Reicher, D. (2013) *Nationensport und Mediennation. Zur Transformation von Nation und Nationalismus im Zeitalter elektronischer Massenmedien*. Göttingen, Deutschland: Vandenhoek & Ruprecht.

Reichl, S. and Blumesberger, S. (2024) ‘EOSC Support Office Austria: Visions, needs and requirements for research data and practices. An interview with Susanne Reichl’. Available at: https://zenodo.org/records/11357635

Reichmann, S. *et al.* (2021) ‘Between Administration and Research: Understanding Data Management Practices in an Institutional Context’, Journal of the Association for Information Science and Technology, 11(72), pp. 1415–31. doi: https://doi.org/10.1002/asi.24492

Ribes, D. (2017) ‘Notes on the Concept of Data Interoperability: Cases from an Ecology of AIDS Research Infrastructures’, in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. Portland, Oregon, USA: Association for Computing Machinery, pp. 1514–26. doi: https://doi.org/10.1145/2998181.2998344

Ribes, D. *et al.* (2019) ‘The Logic of Domains’, *Social Studies of Science*, 49(3). doi: https://doi.org/10.1177/0306312719849709

Ross-Hellauer, T. *et al.* (2022) ‘Dynamics of Cumulative Advantage and Threats to Equity in Open Science: A Scoping Review’, *Royal Society Open Science*, 9(1), 211032. doi: https://doi.org/10.1098/rsos.211032

Schmutzhard, E. and Flicker, K. (2023) ‘EOSC Support Office Austria: Visions, needs and requirements for research data and practices. An interview with Erich Schmutzhard’. Available at: https://zenodo.org/records/7868201

Schreiber, M. and Flicker, K. (2023) ‘EOSC Support Office Austria: Visions, needs and requirements for research data and practices. An interview with Maria Schreiber’. Available at: https://zenodo.org/records/7758011

Slota, S. C. and Bowker, G. C. (2017) 'How Infrastructures Matter', in Felt, U. et al. (eds.), *The Handbook of Science and Technology Studies*. Cambridge, Massachusetts: MIT Press, pp. 529-554.

Smith, R. (2006) ‘Peer Review: A Flawed Process at the Heart of Science and Journals’, *Journal of the Royal Society of Medicine*, 99(4), pp. 178–82.

Spiekermann, S. (2016) *Ethical IT Innovation. A Value-Based System Design Approach*. London, New York: CRC Press.

Star, S. L. and Ruhleder, K. (1994) ‘Steps towards an Ecology of Infrastructure: Complex Problems in Design and Access for Large-Scale Collaborative Systems’, in *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*. New York, NY, USA: Association for Computing Machinery, pp. 253–64. doi: https://doi.org/10.1145/192844.193021

Star, S. L. and Ruhleder, K. (1996) 'Steps Towards an Ecology of Infrastructure Complex Problems in Design and Access for Large-Scale Collaborative Systems', *Information Systems Research*, 7(1), pp. 111-134. doi: https://doi.org/10.1287/isre.7.1.111

Star, S. L. (1999) 'The Ethnography of Infrastructure', *American Behavioral Scientist*, 43(9), pp. 377-391. doi: https://doi.org/10.1177/00027649921955326

Sztompka, P. (1999) *Trust: A Sociological Theory.* Cambridge u.a.: Cambridge Univ. Press.

Taddeo, M. (2010) ‘Modelling Trust in Artificial Agents, A First Step Toward the Analysis of e-Trust’, Minds and Machines, 20(2), pp. 243–57. doi: https://doi.org/10.1007/s11023-010-9201-3

Taddeo, M. and Floridi, L. (2011) ‘The Case for E-Trust’, Ethics and Information Technology, 13(1). pp 1–3. doi: https://doi.org/10.1007/s10676-010-9263-1

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

Walker, M. U. (2006) *Moral Repair: Reconstructing Moral Relations after Wrongdoing*. Cambridge: University Press. doi: https://doi.org/10.1017/CBO9780511618024

Wilkinson, M. D. et al. (2016) ‘The FAIR Guiding Principles for Scientific Data Management and Stewardship’, Scientific Data 3, 160018. doi: https://doi.org/10.1038/sdata.2016.18

Wilkinson, M. et al. (2018). ‘A Design Framework and Exemplar Metrics for FAIRness’, *Scientific Data 5*: 180118. doi: https://doi.org/10.1038/sdata.2018.118