

Reflections on Traceability for Visualization Research

J. Rogers¹  D. Akbaba²  J. S. Brown³  A. Lex^{4,5}  and M. Meyer⁶ 

¹Idaho National Lab, USA

²KTH Royal Institute of Technology, Sweden

³University of Edinburgh, UK

⁴Graz University of Technology, Austria

⁵University of Utah, USA

⁶Linköping University, Sweden

Abstract

Decades of advocacy for reproducibility and replication have advanced open, transparent practices in the sciences. However, traditional notions of reproducibility fit poorly with design-oriented visualization research, where insights emerge through subjective, situated, and iterative work. So how can we ensure rigor and transparency in processes that are inherently unreproducible? To introduce transparency in design-oriented research, we propose to focus on traceability: surfacing the origin and development of research contributions based on rich sets of artifacts documenting the design process. We investigated traceability through a collaborative autoethnographic reflection that builds on several years of work exploring ways to make design-oriented research transparent. This exploration includes an experiment to build a tool to support traceability, which we called tRRRacer. The tRRRacer tool provided a testbed for us to operationalize the three tenets of a traceable process: (1) Record abundant, annotated artifacts representative of research activities; (2) Report curated research threads that articulate rationale and evolution of the process, allowing others to (3) Read via interfaces that help retrace claims and assess plausibility. Reflecting on our experiences, we contribute a theorization of traceability and reflections on how we might support it.

CCS Concepts

• *Human-centered computing* → *Visualization design and evaluation methods*;

1. Introduction

Reproducibility and replication of empirical scientific research are prominent focuses and concerns in computer science [Nat19]. Reproducing and replicating results provides a research community with the means to validate and trust their truthfulness. The visualization research community has championed reproducibility and replication for decades, producing a shift towards open research practices, more access to data, preregistration of studies, and an increased valuing of replication studies [FF20, BBD*21, NEDM18].

However, the field of visualization — like computer science more broadly — is increasingly methodologically diverse with a plurality of epistemic foundations and approaches. Today, many visualization methods of inquiry include subjective and reflective practices, such as design methods and qualitative analysis, making them ill-suited to traditional reproducibility and replication requirements. In considering subjective and reflective research, scholars in the social sciences and design research communities instead advocate for transparency and scrutinizability so that others can assess the credibility, plausibility, and rigor of the research results [LG85, Tra10, ZF08]. But what does it mean for design-oriented visualization research to be transparent? How do

we record and report reflective processes? What is — or could be — the evidence of iterative, subjective, and reflective research methods? And how do we make this evidence legible to others?

In this work, we propose a theoretical framing called *traceability* to support the transparency and scrutinizability of non-reproducible visualization research. This framing is grounded in the concept of material traces from science and technology studies (STS), which theorizes that objects in the world are embedded with traces of their making. From this position, we articulate a set of concepts for visualization research — research artifacts and research threads — for making design and reflective work in research legible to others. This theoretical framing originates from a collaborative autoethnographic reflection [LCB*25] on a design experiment in which we created a tool for supporting traceability. Called tRRRacer, this tool was designed to support the core tasks of tracing: recording, reporting, and reading. Although our experiment did not result in a tool for public dissemination, it did provide us with new ways to think of how we might support traceability in the future.

The core contribution of this work is a theoretical conceptualization of traceability intended as a generative framing for reasoning about recording and reporting visualization research. We ground

this conceptualization in theories from STS as well as a design experiment in which we designed, developed, and tested a new tool for recording and reporting on visualization research. Reflecting on our successes and failures in the experiment, we propose speculative future directions to support the traceability of visualization research and question how much transparency is enough. We see this paper as part of a continuing discussion on visualization design, open practices, and rigor across an increasingly diverse set of visualization research methods.

2. Background

The ability to reproduce and replicate scientific results is a hallmark of good science, providing a means to validate the truthfulness of studies and, ultimately, to establish confidence and trust to build upon others' prior work. The inability to reproduce and replicate vast numbers of published scientific studies, however, has gained attention in both the scientific literature and popular media. In response to this *replication crisis*, the US government commissioned a study in 2017 to assess the reproducibility and replicability of scientific and engineering research. The study's final report defines reproducibility as the ability of a second researcher to recompute original results through the "availability of data, code, and methods that make that recomputation possible" [Nat19]. Replication is defined similarly, but using independently obtained data. The report acknowledges, however, that not all scientific research is replicable, recognizing that "reproducibility and replicability are not, in and of themselves, the end goals of science, nor are they the only way in which scientists gain confidence in new discoveries" [Nat19].

A study by Raghupathi et al. looked at reproducibility characteristics across a corpus of computer science studies and found that the multidisciplinary nature of the field, along with the development and validation of artifacts in socio-technical contexts, contributes to problems with reproducibility [RRR22]. A critical issue their study highlights is the subjective and largely undocumented nature of technology design and development: practices that are poorly understood and do not fit neatly into normative frameworks for verifying the validity of hypothesis-driven scientific research.

Visualization researchers have a long history of advocating for making visualization research reproducible [SFC07, FF20, BBD*21], as well as in developing provenance tools for supporting the reproducibility of visual data analysis [XOW*20]. Provenance tools focuses on addressing the questions *what happened* and *when* in an interactive visual analysis session [RESC16, CGL20, GLG*16a, SFC07, SGP*18, CZGR09]. This work is important in the context of this paper for understanding current practices and tools for tracing the iterative work of data analysts and creating readable traces for others [GLG*16a, WAF22]. This work, however, focuses on making visual analysis processes transparent during the *use* of a visualization tool, while the work we present in this paper highlights the unique considerations and challenges of tracing the subjective and undocumented practices of visualization *design*.

Visualization researchers have argued that some approaches to research — such as visualization design studies [SMM12] that focus on creating visualization tools for real-world contexts — are not only inherently unreproducible, but attempting to make them

so would undermine the rigor of the design-oriented research itself. Alternative requirements focus instead on making research *transparent* and *scrutinizable* such that others can make judgments about the appropriateness of methods, quality of evidence, and reasonableness of conclusions [MD20]. Other visualization researchers have proposed pragmatic approaches to enabling transparency of some design processes, such as *literate visualization* that calls for explicitly including rationale for decisions made while iteratively coding visualization tools [WKD18].

The arguments and theories behind calls for transparency in visualization research — as well as in HCI more broadly [WEHE20, SNGSBR*23] — build from long-established research norms in the social sciences, and for qualitative research in particular [LG85, Tra10]. Software systems for coding qualitative data help enable transparent reporting of analysis [LS08], and reflective memoing during the research process makes researchers' subjectivity and biases more visible [BCF08, Cha06]. Additionally, *audit trails* are a widely used method for supporting the evaluation of qualitative studies by giving auditors access to a researcher's methodology, decisions, and analytical processes to confirm a study's findings [Car09]. We observe that qualitative researchers have already established rigorous methods to trace their non-reproducible findings, unlike visualization design researchers, who are the focus of our efforts. An exception is the work by Rogers et al. which experimented in building audit trails for visualization research by providing access to study materials via web-based tools [RPH*21]. We found existing guidance on building audit trails to report on interviews and observations, but not for transparently reporting on the design and development of technology artifacts.

Instead, work within the design research community has focused on acknowledging and communicating acquired design knowledge embedded within artifacts [ZF08]. As carriers of knowledge, designed artifacts are imbued with what a designer comes to know about needs in the world and ways of shaping materials into a desired object [Cro99]. This knowledge is, however, opaque and requires explicit documentation about the underlying design rationale and decisions. *Annotated portfolios* are an approach for revealing design insights from the abstraction of artifacts, designed for specific situations, into generalized knowledge [GB12]. This is often accomplished by bringing together a collection of artifacts and making explicit the designer's particular knowledge embedded in each of them, and how that knowledge generalizes across the collection. While annotated portfolios provide insight into the importance, novelty, and innovativeness of artifacts, they do not reveal the underlying processes of how these artifacts came to be.

In summary, visualization researchers have acknowledged and embraced requirements for making their work reproducible, yet not all visualization research is inherently so. In particular, the design of visualization interfaces and tools requires some amount of subjectivity and reflection, making these design practices ill-suited to reproducibility and replication requirements. More recently, researchers have instead proposed that these studies be *transparent* and *scrutinizable*, building arguments from similar calls in the fields of social science and design research. These fields, however, do not account for the specific challenges, needs, and processes of designing visualization tools, leaving a gap in knowledge about

how to make visualization research transparent and scrutinizable for others. Our work, described in this paper, is an exploration of how to fill this gap.

3. Traceability

The goal of our work is to make visualization research **traceable**: that *the doing and learning from research practices are transparent, scrutinizable, and legible*. We consider a traceable research process to be one that supports others in understanding what was done and why, and how the findings, insights, contributions, and conclusions of the research came to be. Importantly, tracing a research study is something that is done by **the reader** of a research report as they seek to understand what was done and why, with support from **the researcher** through the evidence and descriptions made available.

Our conceptualization of traceability draws from scholarship on material traces from STS – the idea that objects are imbued with cultural and social knowledge that can be read through their material properties. Material traces capture this link between a phenomenon and the marks it leaves, for example, footprints in sand left from a walk on the beach [Off19]. More broadly, these marks can support scrutiny of what happened and why [Off19, Dou22]. History of science scholars describe science as a process of trace-making, where traces serve as the objective evidence left behind [DG21]. More recent theorizing emphasizes that traces are not simply found but become meaningful through interpretation when someone engages with those marks [Bar07].

Building from these foundational concepts, we treat traceability as an intentional research practice, in which researchers must deliberately record and richly describe evidence so that readers can reconstruct what was done and why. This perspective aligns with transferability in qualitative research, where readers assess how insights might carry to their own context, supported by *thick description* – abundant, context-rich accounts of the research setting, methods, observations, and reflections [Gee08, LG85, MD20]. Here, thick descriptions can take the form of contextual and interpretive notes akin to qualitative fieldnotes and analytic memos that capture what was happening, why an artifact matters, and how interpretations and decisions evolved, leaving an auditable trail beyond the artifact alone [BCF08].

We consider traces as something interpreted by readers as they engage with marks left behind from the research process — the evidence — that are made available by the researcher. This perspective has two important implications. First, like critiques of big data that argue data is meaningless without context [bc12, GMV*21], the evidence collected from a research process requires context, description, and annotation to be meaningfully interpreted by others. Second, a reader's sense-making of a trace is a type of knowledge- and meaning-making that is situated within a reader's own knowledge, experiences, biases, and history [Har13, Bar07, DK20, AKM25]. Taken together, these two implications mean that a researcher cannot control *exactly* how a reader will interpret the evidence they make available, but they *can* make their research process more legible through abundant evidence that is thickly annotated. Thus, for researchers, the goal of traceability is to leave abundant marks that are rich, vibrant, and authentic, such that the reader's interpreta-

tion of the traces reflects the nuance of the underlying processes as closely as possible.

But what, more specifically, is important to trace from visualization research? We propose two concepts — research artifacts and research threads — to address this question, as well as three design requirements of recording, reporting, and reading to support tracing in practice.

3.1. Research Artifacts

A visualization design process consists of a set of research activities. When designing visualizations, these activities can span a myriad of possibilities, from conducting an interview, collecting measurements during a user study, or iterating on a visual encoding, to ideating on a whiteboard, coding a feature in a software prototype, or running an analysis script. Supporting the traceability of these activities gives readers insight into *what* was done during the study.

To make activities traceable, we propose a specific conceptualization of **research artifacts**. Research artifacts are *the marks left behind from a design process; they are the evidence of what was done*. Different types of artifacts are evidence of different visualization design activities: visualization tools reflect how a researcher interpreted a data analysis need in the world; transcripts from interviews capture how a researcher engaged with a domain and the people in it; images taken during a participatory workshop show the activities that a researcher used to engage participants, and the resulting ideas that emerged; analysis code and scripts capture the statistical models a researcher used to interpret data; and much more. It is from an abundant and rich collection of artifacts that a reader can interact and interpret the trace of the activities a researcher engaged with to better understand the research process.

Constructing rich, vibrant, and authentic traces of these activities requires more than just a dump of numbers or a pile of screen-grabs. Although these artifacts are imbued with the details of the research activities that produced them, these details may not be legible to anyone other than the researcher who recorded them [Cro99, GB12]. Annotating research artifacts with rich details about the context of the activity can provide a more vivid picture of the processes that led to the artifacts [Bow12], literate visualization captures rationale during visualization development [WKD18], and computational notebooks make data analysis legible [RTH18].

3.2. Research Threads

The insights that a researcher gains through designing a visualization come from exploring, testing, and learning across *multiple* activities. For example, a researcher interviews collaborators about their domain, learning about where their data comes from and the kinds of analysis they hope to accomplish. The researcher then sketches ideas of possible visualizations, going back to the interview to motivate and justify different design ideas. Taking the sketches back to the collaborators for feedback, the researcher collects their thoughts on the feasibility and usefulness of the ideas. These three activities — an initial interview, an ideation session, and a feedback interview — are not isolated but are meaningfully connected by their involvement in a researcher's process of producing a new visualization technique. Supporting the traceability

of this learning process is important for communicating *how* the researcher's insights and conclusions came to be.

Traceable visualization research requires making research activities and their connection to research results transparent and scrutinizable. We propose **research threads** as a mechanism for supporting readers in tracing how a researcher came to conclusions, designed a new tool, or gained some insight over the course of a study. More specifically, a thread is a *curated collection of artifacts that are evidence of a researcher's learning process that trace a research result from its emergence*. Like artifacts for constructing traces of activities, threads for constructing traces of learning require annotation and descriptions to make the meaning of the relationships legible to readers: to explain why artifacts matter and how they fit together.

3.3. The Three Rs: Design Considerations for Traceability

Supporting traceability requires the consideration of three critical tasks — **recording**, **reporting**, and **reading** — that mediate the communication of research activities between researchers and readers. The first two of these tasks — recording and reporting — focus on supporting researchers as they document their processes, while the last — reading — supports consumers of a research paper (reviewers or readers) in the understanding of traces, i.e., how an insight or result emerged from the process. Crucially, these 3Rs support the transparency and scrutinizability of visualization research.

R1: Recording Traceability fundamentally builds from an abundant and diverse collection of artifacts, which researchers must thoughtfully produce, collect, store, and annotate. Recording of research activities produces artifacts, which can be considered a making activity that creates persistent evidence of an otherwise ephemeral process. Tools and processes that support recording need to accommodate a wide range of research artifacts, as well as provide mechanisms to annotate and curate them.

R2: Reporting Researchers can construct threads from a collection of artifacts through a process of *reflection-on-action* [Sch17]. These threads report on the researcher's actions and reasoning during the research process to facilitate the validation and auditing of findings by others. Reporting requires a researcher to access and review artifacts, to thread together a curated set of artifacts, and to describe why and how these artifacts fit together.

R3: Reading Finally, research activities and threads must be accessible, legible, and scrutinizable by external readers, such as students, colleagues, reviewers, collaborators, and others. Readers should be able to engage with threads and artifacts to understand and validate the research process, as well as assess the plausibility of the derived outcomes. Traceability rests on readers' understanding of how research findings came to be through access to threads and artifacts.

4. The tRRRacer Experiment

In this section, we outline the design experiment we conducted to refine our understanding of tracing visualization research. We conducted our experiment as a collaborative autoethnographic study,

grounded in years of previous work exploring transparency practices within design study research. This study, like other recent collaborative autoethnographies in visualization [LCB*25] and HCI research [CFFGA*25], provided us the opportunity to use our firsthand experiences as a starting point for new knowledges and understandings [HDF21].

We include a description of our experimental journey, details of the tool we developed called tRRRacer, and what we learned through it all. We supplement this description with deep links to artifacts [🔗], activities [🔗], and threads [🔗] to enhance tracing of our journey by you, dear reader.

4.1. Process

The origins of the design experiment were rooted in our collective struggles in reporting on design-oriented, reflective visualization research. We initially explored ways to report on design studies using audit trail-like supplemental materials where we reported on research activities in a timeline view [🔗, 🔗]. In these projects, we did not focus on recording a diverse collection of artifacts; instead, we predominantly documented them through meeting notes and screenshots. In a later design study, we focused on newly proposed rigor criteria [MD20] and worked to collect an abundance of evidence [🔗]. Here, our artifact collection included emails, annotated literature, sketches, photographs, transcripts, and more. Our reporting included deep links to individual artifacts directly in the paper [RPH*21] as well as a timeline view of the full collection.

These early experiments focused on how to visualize a collection of research artifacts, relying on timeline-based views. Each project increased our awareness of the need to systematically and consistently record research artifacts, as well as to consider the reading experience. Building on these early efforts, we wanted to extend these ideas toward practical and technical support for traceability. To better understand the needs, challenges, and opportunities for supporting tracing, we embarked on a design experiment to build and test a new tool called tRRRacer. We conceived tRRRacer as a tool to support the 3Rs of tracing: recording, reporting, and reading.

Early in the experiment, we identified several internal research projects that could serve as a testbed for both gathering ideas and eliciting feedback. The first was a retrospective look at the data from our published design study in which we had digitized an abundant collection of artifacts [RPH*21]. The second project was the development of tRRRacer itself, where we used tRRRacer as it was being developed to document our design process and experimentation. The third project was an ongoing interview study about visualization design study collaborations [ALC*23], and the fourth was an ongoing design study that remains unpublished. Through these four projects, we sought to identify commonalities between the recording and reporting needs of different researchers in different types of visualization research projects. We used the retrospective project as a starting point in the development of tRRRacer, and incorporated the other ongoing projects once the tool had sufficient features for use.

The tRRRacer experiment took place from the spring of 2021 until the fall of 2022. During this period, the research team met every week, discussing the status of the tool, receiving feedback on

its use, and brainstorming for future directions. Rogers and Brown led the development of tRRRacer, with Akbaba serving as an additional user of the tool. This period of active development and use lasted until the main researcher on the project, Rogers, completed her dissertation [10], and as a team, we drafted an initial manuscript of our findings. Our initial manuscript was submitted but not accepted for publication [10]. At this point, the team had moved on to other research projects; the lead author started a new position; and the tracing research was paused.

This pause provided us time and space to rethink our experiences with tRRRacer and our transparency needs across a variety of research projects. During this period, Meyer gave a number of invited talks that included discussions of tRRRacer and our initial theorizing about traceability. Positive feedback from these discussions encouraged us not to abandon the project, and in the spring of 2025, we began internal discussions to rethink our theoretical framing and recommendations for future tracing research. Over the course of several months, Rogers, Lex, and Meyer met to discuss the tRRRacer experiment and to reflect on what we would do differently next time. Along with Akbaba, we developed a digital whiteboard of reflections [10] that we grouped into a final set of lessons learned and future considerations we present in this paper.

4.2. The tRRRacer Tool

We built two versions of tRRRacer: a desktop version built with Electron for use by researchers during their work, and a read-only website [10] for readers to trace a research process. Both applications share a single codebase and connect to the same Google Drive repository of notes and artifacts; the Electron app exposes the full authoring interface, while the Reader web app reuses the same logic but removes all editing functionality so readers can only browse. You can try tRRRacer Reader here: <https://trrracer.netlify.app/>

We designed tRRRacer as an overview + details-on-demand interface, shown in Fig. 1 and Fig. 2. Activities (light bubbles) and their nested artifacts (dark bubbles) are laid out over time in the middle of the interface to reveal bursts, lulls, and clusters of research work. This overview view is flanked on the left by a view detailing the available threads. When a thread is selected, it is shown in the middle overview as a connected path through a subset of activities. In the desktop tRRRacer (Fig. 1), the right side view shows artifacts in detail in the context of a selected activity. In tRRRacer Reader (Fig. 2), artifact details are shown in a pop-up window, with the right side view showing a research report that contains embedded, deep links to artifacts, activities, and threads. All views are linked via user interactions and selections. In the following subsections, we describe our specific design decisions for supporting the 3Rs of tracing.

4.2.1. Record an Abundant Collection of Artifacts

Traceability relies on an abundant and structured collection of artifacts. To make captured evidence legible, tRRRacer requires the user to specify a type for every artifact recorded — such as *transcript* [10], *sketchbook page* [10], *memo* [10] — and a brief description. Artifact types add semantic context and enable filtering; annotations record why the artifact matters and what decision or insight

it supports. In tRRRacer, artifacts — notes, screenshots, recordings — are attached to a design activity — meetings [10], workshops [10], design sessions [10], interviews [10]. We integrated the Google Drive API to support the creation of files directly within the tool.

Activities and artifacts can also be tagged. Tags are a lightweight method for researchers to synthesize important information during recording, to leave breadcrumbs for later reflection, and to seed threads by linking related evidence across activities. Tags also support readers by enabling filtering and discovery during reading [10].

4.2.2. Report with Threads

The activity of reporting creates a bridge from raw records to legible evidence. We designed tRRRacer's reporting workflow to: curate artifacts into useful narratives, expose the rationale behind design decisions, preserve temporal and contextual links among activities, and prepare evidence so others can later retrace and scrutinize claims. All of these activities are captured in research threads.

We designed research threads to curate and annotate evidence that traces how a research idea emerges and evolves [10]. Researchers create a thread when a concept becomes salient as an important insight of the project. To emphasize which artifacts predate the explicit recognition of an insight, we visually distinguish post-hoc from later additions to the thread (dotted vs. solid links; Fig. 1D–E). We ask users to provide a brief rationale whenever evidence is added to a thread (Fig. 1A) to encourage articulation of why the evidence matters at the moment of curation and to leave an explicit record for transparency. Researchers can highlight areas of text to include in a research thread, which we refer to as artifact fragments (as seen in this example thread [10]). tRRRacer also supports editable and mergeable threads. Editing prevents premature lock-in; merging supports abstraction by combining related threads into higher-level ideas.

4.2.3. Read and Retrace

To support readers we built an explicit reading view – tRRRacer Reader – to support understanding of what happened and when, and to trace salient ideas via research threads. Researchers can copy an auto-generated citation for any piece of evidence (Fig. 1C) and embed it in a report. tRRRacer Reader then embeds a report view (Fig. 2) that interacts with the rest of tRRRacer through the deep links specified in the report. tRRRacer Reader shows per-page deep-link previews (Fig. 2C) so readers can scan where evidence is concentrated and jump directly to cited artifacts, activities, and threads. Clicking a tRRRacer link in an external PDF viewer opens tRRRacer Reader [10].

tRRRacer Reader is read-only to protect evidence integrity and enable safe sharing beyond the authoring team. We set privacy defaults that hide activities marked as private and replace names with initials to reduce re-identification risk and align with requirements for double-blind reviews and potential ethical expectations (e.g., hiding full-text interviews) [10]. To support discovery beyond citations, we provide a time-based overview of activities with nested artifacts, answering “*what happened, when?*” at a glance. Navigation is thread-centered: the left sidebar lists research threads, and

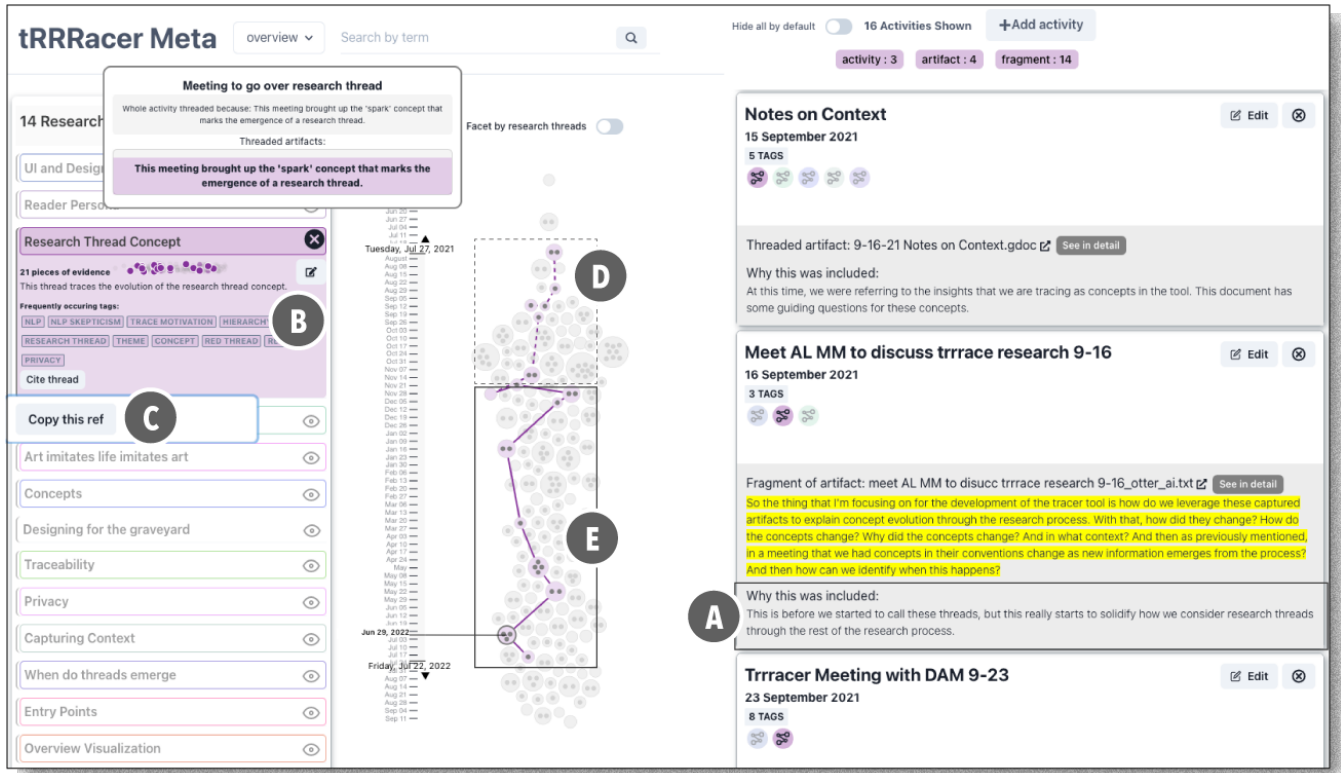


Figure 1: The tRRRacer desktop interface. (A) The detail view shows artifacts collected in the selected thread. When adding a piece of evidence to a thread, the researcher is encouraged to also provide a rationale for why it contributes to a given thread. (B) A list of all threads is shown in the sidebar, which can also be used to create new threads. (C) Researchers can copy an automatically generated thread citation to use in a paper. The citation looks like this: `\trrracer{project name}{type of view}{granularity of evidence}{id}` (D-E) Threaded activities are shown as connected in the overview bubble visualization. (D) The dotted lines linking activity bubbles represent activities threaded retroactively; (E) solid lines indicate that activities were added after the thread was created.

selecting one highlights it in the overview to foreground the curated rationale connecting evidence to claims. Readers can move fluidly between paper-linked entry points and process-wide exploration to retrace claims at both narrative and procedural levels.

4.3. Threads as Windows into the Research Process

To make the concept of research threads more concrete, we briefly explore their affordances in the context of three vignettes. These speculative vignettes, derived from our internal testbed projects, are meant to explore how threads might support reporting on (a) the merging and branching of ideas, (b) dead ends and abandoned directions, and (c) the gradual surfacing of concepts over time. We offer these as an experiment for you, the reader, to reflect on both the opportunities and the challenges of tracing visualization research.

4.3.1. Tracing Concept Development: Merging Threads

To inform the initial design of tRRRacer, we constructed a retrospective research thread for our prior design study in evolutionary biology [RPH*21]. In that project, the evolutionary concept of *convergence* emerged in initial meetings and sketches with our biology

collaborators as a tentative, standalone direction for our tool design. As the project progressed and our understanding of the domain and stakeholders solidified, *convergence* was gradually absorbed into a broader, more expansive concept we called *patterns of evolution*, which we ultimately used as a design goal and reported as a central focus of our design study. The associated thread [5] connects early notes, sketches, and design reflections centered on *convergence* to later design rationales and manuscript text, focused on *patterns of evolution*.

This example motivated our support in tRRRacer for *branching* and *merging* operations for threads. Branching captures how lines of inquiry naturally split into parallel probes and alternative framings, while merging captures how recurring themes later coalesce into higher-level concepts. We speculate that making these research paths explicit can document idea growth, reduce redundancy, and preserve provenance.

4.3.2. Tracing Abandoned Directions: Dead-End Threads

Threads can capture abandoned research directions. An example of this is our thread *NLP for Entry points* for this work, constructed early in the research process when we attempted to use off-the-

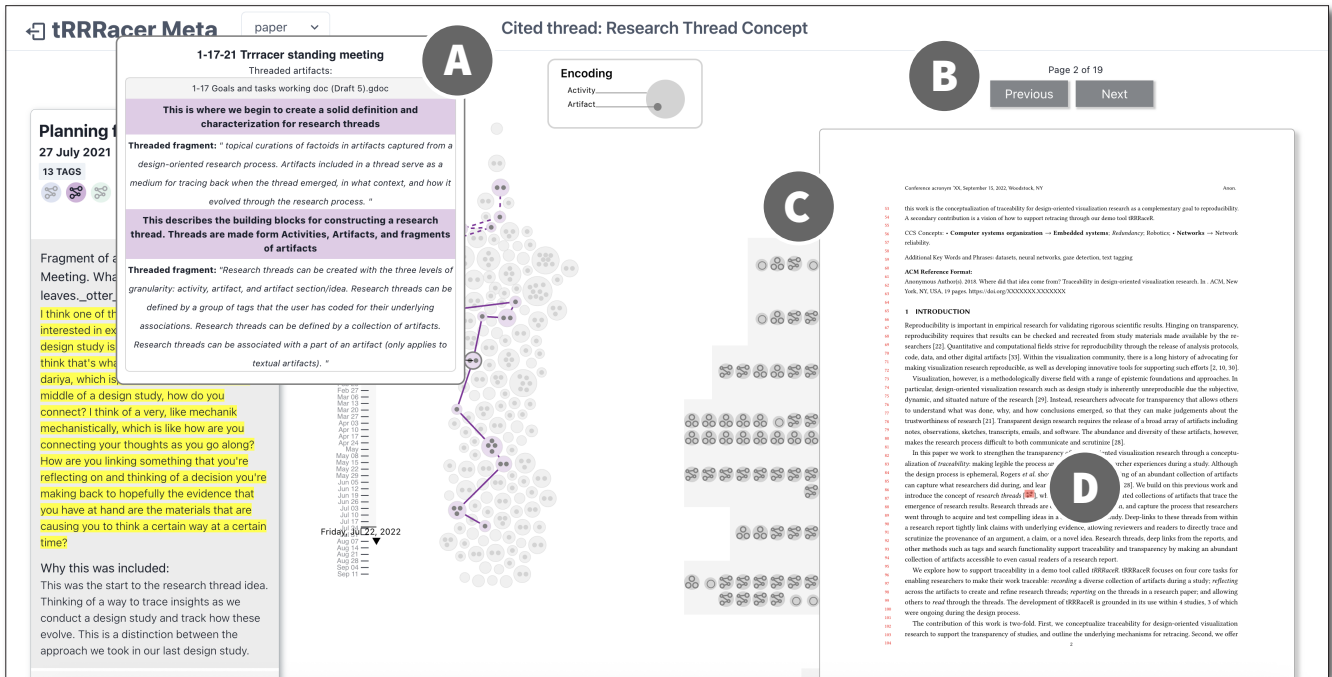


Figure 2: The read-only tRRRacer Reader interface with a paper view. (A) The panel on the left shows relevant activities and artifacts for a selected citation. This example shows a tooltip explaining the rationale for artifacts’ inclusion in the research thread for the ‘Research Thread Concept’. (B) A reader can navigate by page, reading the manuscript in the view, or (C) they can explore with the side annotations for links on the side. (D) An example of an inline deep link. Clicking on these navigates to the detail view for a given activity or artifact.

shelf NLP methods to surface promising starting points in a large corpus of project notes. Despite initial enthusiasm, we ultimately concluded that the recommended leads did not meaningfully guide our work and decided to abandon this direction. The associated thread connects initial brainstorming notes “could we auto-suggest leads?”, prototype screenshots, pilot evaluation notes, and later reflections where we articulated why the approach was not worth pursuing further [35].

Documenting failed avenues in threads has the potential to counter hindsight and survivorship bias, showing that final claims arose from a systematic inquiry rather than cherry-picking. Dead-ends can expose alternatives considered and the rationale for abandoning them, which can strengthen the credibility of the chosen approach. They also improve transferability: readers can judge boundary conditions and avoid repeating unpromising paths in different contexts. Finally, unsuccessful artifacts could become seeds for later ideas, and we speculate that recording them in a thread could preserve reusable knowledge and starting points for future explorations.

4.3.3. Tracing a Core Concept: Evolving Threads

Recording and visualizing threads can surface when and how ideas evolve over time. To make our process for this paper traceable, we created threads documenting the development of central concepts that we grappled with as a research team, including the concept of research threads [35]. This thread illustrates our evolution of

thought around research threads: from its first mention as an idea for creating “traces of a process as we go” [35], to the inclusion of tags as a helpful tool for “pre-threading” [35], and what information should be included within a thread.

Another thread reveals that we were plagued with the recurring question: “when should you begin threading?” [35]. This lingering question marked an interesting tension for us throughout the design experiment. We initially theorized that threads could support reflection-in-action, and yet we struggled to create threads during the research process. It was only during writing that we could more easily generate threads. We return to this tension in the following section.

4.4. Lessons Learned in Traceability

Developing the tRRRacer tool was a design experiment that gave us the opportunity to explore the challenges and opportunities of supporting tracing of visualization research. Through cycles of reflection, we distilled our learning about traceability into a set of lessons, drawn from both our experiences using the tool and the insights it enabled about our own research practices. We detail those lessons here.

We note that we would not build tRRRacer again and do not plan on maintaining or expanding it. However, our experience with developing a singular tool for recording and reporting was insightful

both in terms of how to support traceability, as well as in how we might build research tools that last.

4.4.1. Successes

One clear success of the tRRRacer experiment was our attention to artifact curation. Across the four experimental projects, we focused first and foremost on recording abundant artifacts — sketches, meeting notes, emails, screen shots, reflections, and more. We began to approach all of our research activities with an eye towards documentation, and we made time to systematically and regularly reflect on why those activities mattered. The tRRRacer tool provided curatorial features beyond a storage space for digital files; instead, it oriented our recording processes toward organization, annotation, and externalizing how artifacts related to other concepts within the research project. Both student authors on the paper, Rogers and Akbaba, took these recording practices into their later research projects, where they used commercial tools to continue recording abundant collections of research artifacts. These practices of attending to context through tags and annotations of abundant artifacts have continued to support a nuanced and rich view of how research ideas are entangled and develop over time, serving as important reminders where our memory fails.

As part of artifact curation, using tagging was, and remains, important for discoverability. We frequently used the tags in tRRRacer for reviewing artifacts as well as for creating threads, and are continuing the practice of tagging artifacts in our current use of commercial tools. Additionally, the tags enabled a quick form of reflection-in-action: they are a best guess in the moment of why an artifact matters. While we struggled with creating threads in tRRRacer in the middle of projects — our more intentional feature for reflection-in-action — tags were lightweight and familiar. We discuss our difficulties creating threads in Section 4.4.2.

Beyond simple tags, tRRRacer also enabled reflection-in-action through explicit foregrounding of annotation practices. In particular, annotating research artifacts instilled a process of articulating why an artifact, and thus activity, was important within the research process. Annotating artifacts became akin to reflective journaling.

We found that good annotation of artifacts and threads was critical to making threads readable. Our review of threads we created in the four projects required annotations to make sense of the artifacts by team members not involved in the projects. Annotations of threads also left us with a feeling of reading through a research journal. In short, a dump of artifacts is not enough. Evidence without context is insufficient for tracing a research process.

Our final point of success within the tRRRacer experiment was the use of deep links within a report. Inspired by positive feedback on our use of deep links in one of our grounding design studies [RPH*21], we have continued to use this mechanism for reporting in more recent work [CWS*26, AST*26]. In the tRRRacer experiment we imagined taking this even further, extending the idea to include links to curated threads. Placing access to research evidence directly at the point it matters in a report provides a new way for readers to engage with supplemental materials. It makes the scrutiny of materials easy for a reader and adds a layer of richness to otherwise static and flat PDF paper formats.

4.4.2. Challenges

For all our successes with tRRRacer, there were also a number of challenges that we confronted. First and foremost were the technical ones. Our decision to build a tool that supported recording, reporting, and reading in a single interface led us down a path of significant engineering efforts that mirrored efforts being taken by companies of commercial documentation tools like Notion. Furthermore, our desire to support a broad range of research artifacts — from JPEGs to Google Docs — added additional engineering efforts in order to accommodate the various file formats and document APIs. And once we started using the tool across four projects, backward compatibility became a challenge for any significant changes we wanted to make to the tool. In the end, we found ourselves in a quagmire of software engineering needs and an increasingly complex tool that we would need to maintain. These technical challenges made it difficult for us to focus on the interactive and visual features of tRRRacer, and also difficult to change course when our feedback indicated the need for different approaches.

We were unable to fully realize anonymization features due to both technical and conceptual challenges. Manual anonymization of artifacts is time-consuming and also results in artifacts that are more difficult to trace by the researcher. We wondered whether we should maintain two versions of the artifact collection, one anonymized and one not, and if so, how should we implement this? Or, should we try to codify an anonymization process in the tool that anonymizes automatically in reading mode? We found the desire to protect the privacy of our team and collaborators in tension with our desire to encourage rich and transparent research documentation.

We also struggled technically and conceptually with how to maintain long-term persistence of research evidence. We were using tRRRacer to curate, manage, and store the artifacts and threads from several projects, but due to the increasing complexity of the tool, we could not imagine a long-term maintenance plan for keeping deep links in reports available in perpetuity. Here, too, we struggled to decide whether to focus on making the evidence scrutinizable through interactive visualizations in tRRRacer or persistent and accessible through a simple dump of data.

Despite our awareness of the need to accommodate diverse research and documentation practices, our commitment to a single tool enforced the use of certain formats, tools, and concepts. For example, long text documents from meeting transcripts or interviews could contain multiple ideas involved with multiple threads. How to subset these text documents for threading in tRRRacer required specific documentation practices, which differed depending on whether the researcher documented in plain text files or Google Docs. Here again, we were stymied by the engineering required to realize our ideas.

More conceptually, we found that creating threads as a project is developing — a form of reflection-in-action — was extremely difficult and rarely productive. We had imagined researchers creating threads as new ideas emerged, ideas that might merge together or branch apart, or even die altogether. We imagined a visualization that could show how learning happens within a project through a visualization of all the threads. In practice, however, we found threads difficult to create during projects; it turns out that knowing

what will be most meaningful or what could develop into higher level ideas is incredibly difficult in the moment. Threads we did create during projects tended to be superficial and unstable upon later reflection. It was much easier to thread together artifacts and annotate their meaning in our retrospective project, where we were looking to tell the stories of our findings.

As a final challenge, we found across all of the projects in our experiment that recording and reporting were incredibly time-intensive. Although curating and annotating research artifacts is a practice we are maintaining, we use these curated collections largely for internal use and not for transparent reporting. During our tRRRacer experiment, we experienced collecting artifacts and creating threads as an additional, lengthy step without clear benefits or demands from the research community.

5. Discussion and Future Directions

In this section, we put forward several ideas of how to move traceability forward in the visualization research community based on learning through the tRRRacer experiment. These ideas ask us to rethink the role of large, complex tools; to imagine threads as stories about our research; to embrace reflection throughout our research practices; and to reconsider whether more is always better.

5.1. Process and Persistence

Our findings from the tRRRacer experiment, both the successes and challenges, point to the need for processes over tools for traceability. Instead of trying to build yet another tool to support documenting research projects, we envision a set of guidelines and standards for the community that detail what kinds of artifacts to collect for different kinds of activities, what metadata and annotations to contextualize the artifacts, and when and how best to record. This approach would leave open the opportunity for researchers to use the tools that are most suited to their needs while providing them with actionable goals for what and how to record. These guidelines and standards, however, will require further experiments by visualization researchers in a range of settings who are committed to exploring traceability of their own work.

Additionally, we advocate for the use of simple, time-tested file formats for storing research artifacts. Formats like JPEG, MPEG, PDF, and plain text are all likely to be supported for many years to come. Restricting our documentation to these types of formats will ensure evidence is accessible to others using a range of tools without dependencies on singular companies. While interactive visual interfaces have been great for supporting the needs of many, we argue that maintenance and persistence of tools is no longer something we should ignore. Simple formats, along with guidelines and standards for recording practices, provide a common basis that will stand the test of time.

Preservation of evidence has long been a priority for advocates of open science [Rie12]. With the rise of open data in the sciences, researchers began looking for robust ways to maintain data and findings, and repositories like arXiv and OSF came online. While we think the practices in the open science community of using online repositories are reasonable, current geopolitics and the restructuring of academic research give us pause. It is clear that open science

repositories that rely on governmental funding, or any sort of funding, are vulnerable to change. As such, we advocate for redundancy when storing artifacts and threads. Researchers may decide to use established preprint repositories, while also including the collection of evidence with publishers and on their personally maintained websites, or commercial ones like GitHub. If using deep links to research collections in a paper, we could imagine a persistent version of the paper that could be downloaded along with the collection that has links to locally stored files, or that links to an attached appendix section. Nothing is forever, but we may be able to hedge our bets through redundancy.

5.2. Threads as Stories

Although we found thread creation to be difficult *during* a research project, it was a useful mechanism for reflection-on-action in later stages of projects. The process of creating threads is an opportunity to revisit evidence about how important and salient ideas came to be. We recommend that threads be a focus of reporting and that they be created during the process of writing up research results. In this way, threads serve as an alternative view into a research project, complementary to more traditional text-based reports.

Placing thread creation in the reporting phase of research opens up the idea of threading as storytelling. This perspective acknowledges threading as an act of curation rather than completeness. Threads-as-stories also emphasizes the interpretive, subjective, and selective nature of what aspects of the research process are made visible, and how. Thus, it can be argued that threads are biased reconstructions, but we would counter that *all* reporting on research inherently has this quality. Positioning threads as stories acknowledges the subjectivity of reporting and gives space for thoughtful recording and reporting practices to reduce biases, such as the use of thick descriptions [Gee08, BCF08] and abundant evidence [MD20].

We see interesting future work in how we might design authoring tools for thread creation. Similar to work in provenance for crafting stories about insights from visual analysis sessions [GLG*16b], these tools would provide researchers opportunities to explore through past research activities via a collection of artifacts, and to link them together in ways that reveal a thinking process. Designing these tools would be an opportunity to explore anonymization strategies, such as outputting a curated collection of artifacts associated with a thread that have been sufficiently anonymized. An interesting design question is how to best visualize threads and their associated contextualizing annotations and metadata, and how to do so in a way that is persistent.

5.3. Other Benefits of Traceability

This paper emphasizes how traceability strengthens rigor in the research process. Yet when we first initiated this project, we also envisioned that the act of recording and reporting could actively improve research outcomes through reflection-in-action. We asked whether routine engagement with our prior notes, ideas, and decisions might spark new insights, reveal blind spots, or accelerate refinement. We likewise anticipated that maintaining rich artifacts and traces would simplify reporting — making paper writing

easier because arguments and ideas would already exist in a well-documented form.

In practice, achieving this reflection proved challenging, as previously discussed. We were, in effect, “building the plane while flying it”: our technical infrastructure had to co-evolve with the project’s ideas, yet that same infrastructure was also required to collect artifacts and develop traces. This dynamic limited our ability to fully exploit the benefits of ongoing reflection during the project itself.

Nevertheless, we see our experience as an important first step toward realizing these benefits. Traceable research practices may ultimately enhance research outcomes, much as early evidence suggests that open science practices positively influence research quality [SES*21]. By developing and sharing tools, workflows, and lessons learned, we aim to help others more easily test and extend these ideas. Demonstrating such benefits conclusively, however, will require coordinated and systematic investigation by the broader research community.

5.4. The Burdens of Tracing

Over the past decade, the procedural demands on researchers have grown substantially. Increasingly explicit and implicit rules are intended to enhance fairness, equity, inclusion, and rigor. Examples include pre-registration, detailed supplementary materials with documented analysis scripts, supplementary videos, and accessibility-focused text descriptions. While these practices serve worthy goals, they also create tangible downsides, including the privileging of well-resourced teams able to meet requirements, increasing the workload of junior researchers, and potentially slowing the pace of research.

Traceability exemplifies this tension. Documenting, reflecting on, and curating research processes *can* improve rigor, transparency, and opportunities for others to learn. Deep linking to supplemental material *can* enrich the reading experience. Yet these benefits come at a cost. Recording, digitizing, annotating, and organizing research artifacts requires substantial time and expertise. Even with supportive tools—such as automatic transcription and summarization, these tasks are labor-intensive and potentially disruptive to research flow. Not to mention the increased work of reading and reviewing papers that come with significantly more information to parse.

Community buy-in could help, and we see two complementary paths forward. First, the community could jointly invest in infrastructure for tRRRacer-like traceability, following precedents from the open science movement (e.g., OSF.io provides preregistration, supplemental material hosting, and anonymization features). Alternatively, we could develop low-tech standards for traceability, such as aggregated artifacts in supplementary PDFs, including thickly annotated threads, that are then included in reports via mechanisms such as deep links.

However, open questions remain about the feasibility of tracing that we argue warrant further research. Is it possible to capture the tacit knowledge of researchers within artifacts? What kinds of artifacts should we be collecting? How much evidence and descrip-

tion is enough? What are appropriate incentives to encourage researchers to be more transparent? And ultimately, what is a reasonable balance between the benefits of rigor and transparency with the practical realities of research practice?

5.5. Limitations

Our reflections on traceability are grounded in a design experiment conducted within our own research group, and therefore inherit the scope and biases of that setting. Furthermore, our experiment opens more questions than answers around how to support researchers and readers in tracing. Our work did not thoroughly explore all the kinds of artifacts that could be recorded, the ways artifacts may be annotated, or the myriad of opportunities for visualizing and communicating threads.

We did not evaluate the researcher or reader experience of tRRRacer with audiences beyond our team. As a result, we do not yet claim that threads and artifact collections are appropriately scoped or beneficial for broader audiences in realistic reading contexts. Our tRRRacer tool was an experiment, conducted in the designerly approach of *research through design*; as such, we do not claim the tool as a final endpoint, or as an artifact worthy of extensive user testing. The development of tRRRacer was a sandbox for us, as a research team, to explore ideas and learn along the way. We see the potential for future studies to explore how threads and deep links, for example, can beneficially supplement the reading experience, as well as the testing of new tools and workflows for supporting bespoke and light-weight artifact collection and curation.

6. Conclusions

We introduced traceability as a framework for making design-oriented visualization research more transparent and scrutinizable. Through our tRRRacer design experiment, we explored how recording, reporting, and reading can work together to surface the origins and development of research contributions. Our reflections revealed both the promise and the burden of implementing traceability: it enriches rigor, transparency, and learning, yet requires time, infrastructure, and thoughtful curation.

We offer our theoretical framing of traceability as both a generative tool for thoughtful recording and reporting of visualization research, as well as a goal for the community to consider from a socio-technical perspective. We hope this paper seeds discussions towards concrete next steps for lightweight practices and community standards for traceability that can support our increasingly diverse and pluralistic research community.

7. Acknowledgments

This work is funded in part by the Swedish Research Council (Grant No. 2024-05726); by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation; and the U.S. National Science Foundation (NSF OAC 1835904).

References

- [AKM25] AKBABA D., KLEIN L., MEYER M.: Entanglements for visualization: Changing research outcomes through feminist theory. *IEEE Transactions on Visualization and Computer Graphics* 31, 1 (2025), 1279–1289. doi:10.1109/TVCG.2024.3456171. 3
- [ALC*23] AKBABA D., LANGE D., CORRELL M., LEX A., MEYER M.: Troubling collaboration: Matters of care for visualization design study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023), pp. 1–15. 4
- [AST*26] AKBABA D., SVENSSON C., TORELLI C., CALLMERYD M., MEYER M.: Thinking inside the box: Considerations for putting data physicalization workshops in a box. *to appear* (2026). 8
- [Bar07] BARAD K.: *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press, 2007. 3
- [BBD*21] BESANÇON L., BEZERIANOS A., DRAGICEVIC P., ISENBERG P., JANSEN Y.: Publishing Visualization Studies as Registered Reports: Expected Benefits and Researchers' Attitudes. In *Proceedings of the alt.VIS workshop* (2021), IEEE VIS. 1, 2
- [bc12] BOYD D., CRAWFORD K.: Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15, 5 (2012), 662–679. doi:10.1080/1369118X.2012.678878. 3
- [BCF08] BIRKS M., CHAPMAN Y., FRANCIS K.: Memoing in qualitative research: Probing data and processes. *Journal of Research in Nursing* 13, 1 (Jan. 2008), 68–75. doi:10.1177/1744987107081254. 2, 3, 9
- [Bow12] BOWERS J.: The logic of annotated portfolios: Communicating the value of 'research through design'. In *Proceedings of the Designing Interactive Systems Conference* (San Francisco, CA, US, 2012), DIS '12, Association for Computing Machinery, pp. 68–77. URL: <https://doi.org/10.1145/2317956.2317968>, doi:10.1145/2317956.2317968. 3
- [Car09] CARCARY M.: The research audit trail: enhancing trustworthiness in qualitative inquiry. *Electronic Journal of Business Research Methods* 7, 1 (2009), 11–24. URL: <https://academic-publishing.org/index.php/ejbrm/article/view/1239>, doi:https://doi.org/10.34190/JBRM.18.2.008. 2
- [CFFGA*25] CIOLFI FELICE M., FELDFEBER I., GLASSERMAN APICELLA C., QUIROGA Y. B., ANSALDO J., LAPENNA L., BEZCHINSKY S., BARRIGA RUBIO R., GARCÍA M.: Doing the feminist work in ai: Reflections from an ai project in latin america. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (2025), pp. 1–18. 4
- [CGL20] CUTLER Z., GADHAVE K., LEX A.: Ttrack: A library for provenance-tracking in web-based visualizations. In *Proceedings of IEEE VIS* (2020), IEEE VIS, pp. 116–120. doi:10.1109/VIS47514.2020.00030. 2
- [Cha06] CHARMAZ K.: *Constructing grounded theory: A practical guide through qualitative analysis*. Sage, 2006. 2
- [Cro99] CROSS N.: Design research: A disciplined conversation. *Design Issues* 15, 2 (1999), 5–10. doi:10.2307/1511837. 2, 3
- [CWS*26] CUTLER Z., WILBURN J., SHRESTHA H., DING Y., BOLLEN B., NADIB K. A., HE T., MCNUTT A., HARRISON L., LEX A.: Revisit 2: A full experiment life cycle user study framework. *IEEE Transactions on Visualization and Computer Graphics (VIS)* 32 (jan 2026). 8
- [CZGR09] CHANG R., ZIEMKIEWICZ C., GREEN T. M., RIBARSKY W.: Defining insight for visual analytics. *IEEE Computer Graphics and Applications* 29, 2 (2009), 14–17. 2
- [DG21] DASTON L., GALISON P.: *Objectivity*. Princeton University Press, 2021. 3
- [DK20] D'IGNAZIO C., KLEIN L. F.: *Data feminism*. MIT Press, 2020. 3
- [Dou22] DOURISH P.: *The stuff of bits: An essay on the materialities of information*. MIT press, 2022. 3
- [FF20] FEKETE J.-D., FREIRE J.: Exploring reproducibility in visualization. *IEEE Computer Graphics and Applications* 40, 5 (2020), 108–119. doi:10.1109/MCG.2020.3006412. 1, 2
- [GB12] GAVER B., BOWERS J.: Annotated portfolios. *Interactions* 19, 4 (2012), 40–49. doi:10.1145/2212877.2212889. 2, 3
- [Gee08] GEERTZ C.: Thick description: Toward an interpretive theory of culture. In *The cultural geography reader*. Routledge, 2008, pp. 41–51. 3, 9
- [GLG*16a] GRATZL S., LEX A., GEHLENBORG N., COSGROVE N., STREIT M.: From visual exploration to storytelling and back again. *Computer Graphics Forum* 35, 3 (2016), 491–500. doi:10.1111/cgf.12925. 2
- [GLG*16b] GRATZL S., LEX A., GEHLENBORG N., COSGROVE N., STREIT M.: From Visual Exploration to Storytelling and Back Again. *Computer Graphics Forum* 35, 3 (2016), 491–500. doi:10.1111/cgf.12925. 9
- [GMV*21] GEBRU T., MORGENSTERN J., VECCHIONE B., VAUGHAN J. W., WALLACH H., III H. D., CRAWFORD K.: Datasheets for datasets. *Commun. ACM* 64, 12 (Nov. 2021), 86–92. URL: <https://doi.org/10.1145/3458723>, doi:10.1145/3458723. 3
- [Har13] HARAWAY D.: Situated knowledges: The science question in feminism and the privilege of partial perspective 1. In *Women, science, and technology*. Routledge, 2013, pp. 455–472. 3
- [HDF21] HOWELL N., DESJARDINS A., FOX S.: Cracks in the success narrative: Rethinking failure in design research through a retrospective trioethnography. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 6 (2021), 1–31. 4
- [LCB*25] LILY W. G., CABOUAT A.-F., BONILLA K., CUI Y., DING Y., RAKOTONDRAVONY N., CREAMER M. M., OTTO J. T., HEDAYATI M., KWON B. C., ET AL.: An autoethnography on visualization literacy: A wicked measurement problem. *IEEE Transactions on Visualization and Computer Graphics* (2025). 1, 4
- [LG85] LINCOLN Y. S., GUBA E. G.: Establishing trustworthiness. *Naturalistic Inquiry* 289 (1985), 289–327. 1, 2, 3
- [LS08] LU C.-J., SHULMAN S. W.: Rigor and flexibility in computer-based qualitative research: Introducing the coding analysis toolkit. *International Journal of Multiple Research Approaches* 2, 1 (2008), 105–117. doi:10.5172/mra.455.2.1.105. 2
- [MD20] MEYER M., DYKES J.: Criteria for rigor in visualization design study. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 87–97. doi:10.1109/TVCG.2019.2934539. 2, 3, 4, 9
- [Nat19] NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE: Reproducibility and replicability in science. *National Academies of Sciences, Engineering, and Medicine* (2019). doi: <https://doi.org/10.17226/25303>. 1, 2
- [NEDM18] NOSEK B. A., EBERSOLE C. R., DEHAVEN A. C., MEL-LOR D. T.: The preregistration revolution. *Proceedings of the National Academy of Sciences* 115, 11 (2018), 2600–2606. doi:10.1073/pnas.1708274114. 1
- [Off19] OFFENHUBER D.: Data by proxy—material traces as auto-graphic visualizations. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 98–108. doi:10.1109/TVCG.2019.2934788. 3
- [RESC16] RAGAN E. D., ENDERT A., SANYAL J., CHEN J.: Characterizing provenance in visualization and data analysis: An organizational framework of provenance types and purposes. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 31–40. doi:10.1109/TVCG.2015.2467551. 2

- [Rie12] RIEGER O. Y.: Sustainability: Scholarly repository as an enterprise. *Bulletin of the American Society for Information Science and Technology* 39, 1 (2012), 27–31. doi:10.1002/bult.2012.1720390110. 9
- [RPH*21] ROGERS J., PATTON A. H., HARMON L., LEX A., MEYER M.: Insights from experiments with rigor in an evobio design study. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1106–1116. doi:10.1109/TVCG.2020.3030405. 2, 4, 6, 8
- [RRR22] RAGHUPATHI W., RAGHUPATHI V., REN J.: Reproducibility in computing research: An empirical study. *IEEE Access* 10 (2022), 29207–29223. doi:10.1109/ACCESS.2022.3158675. 2
- [RTH18] RULE A., TABARD A., HOLLAN J. D.: Exploration and explanation in computational notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), CHI '18, Association for Computing Machinery, p. 1–12. URL: <https://doi.org/10.1145/3173574.3173606>, doi:10.1145/3173574.3173606. 3
- [Sch17] SCHÖN D. A.: *The reflective practitioner: How professionals think in action*. Routledge, 2017. 4
- [SES*21] SODERBERG C. K., ERRINGTON T. M., SCHIAVONE S. R., BOTTESINI J., THORN F. S., VAZIRE S., ESTERLING K. M., NOSEK B. A.: Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour* 5, 8 (Aug. 2021), 990–997. doi:10.1038/s41562-021-01142-4. 10
- [SFC07] SILVA C. T., FREIRE J., CALLAHAN S. P.: Provenance for visualizations: Reproducibility and beyond. *Computing in Science & Engineering* 9, 5 (2007), 82–89. doi:10.1109/MCSE.2007.106. 2
- [SGP*18] STITZ H., GRATZL S., PIRINGER H., ZICHNER T., STREIT M.: KnowledgePearls: Provenance-based visualization retrieval. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 120–130. 2
- [SMM12] SEDLMAIR M., MEYER M., MUNZNER T.: Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2431–2440. doi:10.1109/TVCG.2012.213. 2
- [SNGSBR*23] SALEHZADEH NIKSIRAT K., GOSWAMI L., S. B. RAO P., TYLER J., SILACCI A., ALIYU S., AEBLI A., WACHARAMANOTHAM C., CHERUBINI M.: Changes in research ethics, openness, and transparency in empirical studies between chi 2017 and chi 2022. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2023), CHI '23, Association for Computing Machinery. URL: <https://doi.org/10.1145/3544548.3580848>, doi:10.1145/3544548.3580848. 2
- [Tra10] TRACY S. J.: Qualitative quality: Eight “big-tent” criteria for excellent qualitative research. *Qualitative Inquiry* 16, 10 (2010), 837–851. doi:10.1177/1077800410383121. 1, 2
- [WAF22] WANG B., ADACHI N., FUJISHIRO I.: Forseti: A visual analysis environment enabling provenance awareness for the accountability of e-autopsy reports. *Visual Informatics* 6, 3 (2022), 69–80. 2
- [WEHE20] WACHARAMANOTHAM C., EISENRING L., HAROZ S., ECHTLER F.: Transparency of chi research artifacts: Results of a self-reported survey. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–14. doi:10.1145/3313831.3376448. 2
- [WKD18] WOOD J., KACHKAEV A., DYKES J.: Design exposition with literate visualization. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 759–768. doi:10.1109/TVCG.2018.2864836. 2, 3
- [XOW*20] XU K., OTTLEY A., WALCHSHOFER C., STREIT M., CHANG R., WENSKOVITCH J.: Survey on the analysis of user interactions and visualization provenance. In *Computer Graphics Forum* (2020), vol. 39, Wiley Online Library, pp. 757–783. doi:10.1111/cgf.14035. 2
- [ZF08] ZIMMERMAN J., FORLIZZI J.: The Role of Design Artifacts in Design Theory Construction. *Artifact* 2, 1 (Nov. 2008), 41–45. doi:10.1080/17493460802276893. 1, 2